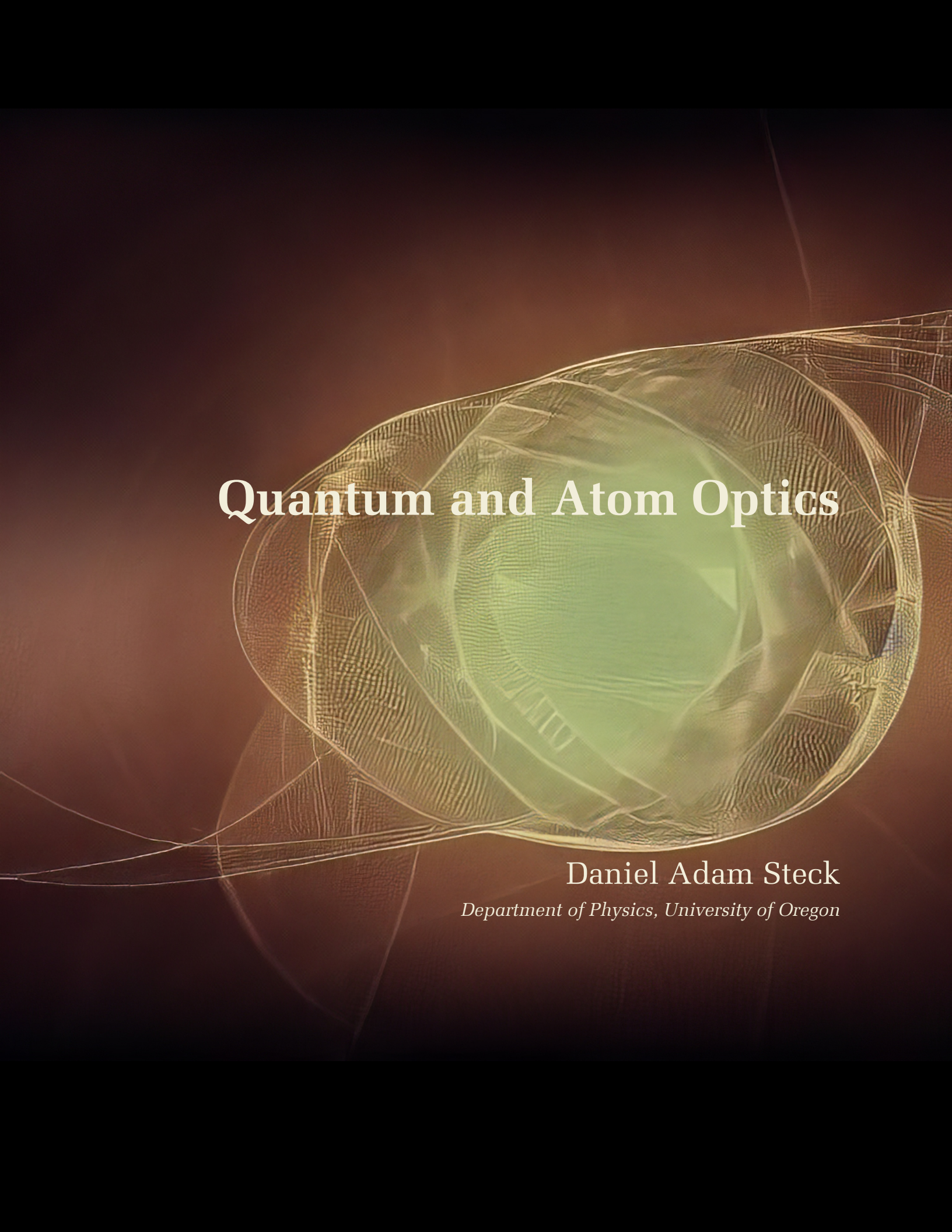


Quantum and Atom Optics

Daniel Adam Steck

Department of Physics, University of Oregon

Quantum and Atom Optics

Daniel Adam Steck
Department of Physics, University of Oregon

Copyright © 2007, by Daniel Adam Steck. All rights reserved.

This material may be distributed only subject to the terms and conditions set forth in the Open Publication License, v1.0 or later (the latest version is available at `http://www.opencontent.org/openpub/`). Distribution of substantively modified versions of this document is prohibited without the explicit permission of the copyright holder. Distribution of the work or derivative of the work in any standard (paper) book form is prohibited unless prior permission is obtained from the copyright holder.

Original revision posted June 2007.

This is revision 0.16.7, 13 March 2026.

Cite this document as:
Daniel A. Steck, *Quantum and Atom Optics*, available online at `http://steck.us/teaching` (revision 0.16.7, 13 March 2026).

Author contact information:
Daniel Steck
Department of Physics
1274 University of Oregon
Eugene, Oregon 97403-1274
dsteck@uoregon.edu

Acknowledgements

A number of people have improved this work by graciously providing corrections, comments, and suggestions.

Special thanks to:

- Jeremy Thorn (U. Oregon) has pointed out many corrections and helped to clarify some of the concepts.
- Jonathan Mackrory (U. Oregon) has also pointed out a vast number of corrections.
- Steven van Enk (U. Oregon) has provided a number of corrections.
- Anders Larsen (Aarhus U.) identified a sloppy argument in Section 18.1.1, and the derivation there follows his suggested proof. Anders also pointed out a sloppy and overzealous use of the term ‘‘Poisson process.’’

Also, thanks to:

- Gregor Anich (U. Innsbruck)
- Boris Albrecht (Niels Bohr Inst.)
- Paul Andreini (U. Oregon)
- Bernd Baumann (HAW Hamburg)
- Jean-Baptiste Béguin (Niels Bohr Inst.)
- Xiao Chen 陈宵 (Wuhan University)
- Dave Grych (U. Oregon)
- Bruce Klappauf (U. Texas Austin)
- Wil Kao (Stanford U.)
- Mark Kuzyk (Washington State U.)
- Paul Lauria (UCSD)
- Arscire Li (USTC)
- Didi Liebfried (NIST)
- Kyle Lynch–Klarup (U. Oregon)
- Andrew McClung (Caltech)
- Erin Mondloch (U. Oregon)
- Asaf Paris Mandoki (U. Nacional Autónoma de México)

- Krishna Myneni (U. Alabama Huntsville)
- Craig Price (Penn State)
- Milan Radonjić (U. Hamburg)
- Kevin Randles (U. Oregon)
- Lijun Ren 任丽君 (Sun Yat-Sen U.)
- Jack Roth (U.C. Berkeley)
- Heather Shannon (U. Maryland)
- Fabian Schmid (Max-Planck-Institut für Quantenoptik)
- Stephen Schoun (Ophir Corp.)
- Eric Smoll (Northwestern U.)
- Richard Wagner (U. Oregon)
- Shuhao Wu (U. Oregon)
- Jun Yin (U. Oregon)
- Dylan Young (U. Colorado, Boulder)
- Pu Zhang (Huazhong University of Science and Technology)
- Peter Zheng (U. Oregon)
- Zhiliang Yuan 袁之良 (Beijing Academy of Quantum Information Sciences)

Contents

Contents 7

I Classical and Semiclassical Light–Matter Interactions 29

1 Classical Atom–Field Interactions 31
1.1 Polarizability . . . 31
1.1.1 Connection to Dielectric Media . . . 33
1.2 Damping: Lorentz Model . . . 33
1.2.1 Oscillator Strength . . . 35
1.3 Dipole Radiation . . . 36
1.3.1 Damping Coefficient . . . 38
1.4 Atom Optics: Mechanical Effects of Light on Atoms . . . 39
1.4.1 Dipole Force . . . 40
1.4.1.1 Dipole Potential: Standard Form . . . 41
1.4.1.2 Photon Scattering Rate . . . 43
1.4.1.3 Optical Theorem . . . 44
1.4.1.4 Scaling . . . 45
1.4.2 Radiation Pressure . . . 45
1.4.3 Laser Cooling: Optical Molasses . . . 46
1.4.3.1 Doppler Cooling Limit . . . 47
1.4.3.2 Magneto-Optical Trap . . . 49
1.5 Cooperative Radiation . . . 50
1.5.1 Atom–Mirror Interaction . . . 50
1.5.2 Two-Atom Radiation . . . 53
1.6 Exercises . . . 57

2 Classical Coherence 63
2.1 Wiener–Khinchin Theorem . . . 64
2.2 Optical Wiener–Khinchin Theorem . . . 65
2.2.1 Michelson Interferometer . . . 66
2.2.2 Example: Monochromatic Light . . . 67
2.2.3 Spectrum of Atomic Radiation . . . 67
2.2.4 Normalized One- and Two-Sided Spectra . . . 68
2.3 Visibility . . . 70
2.4 Coherence Time, Coherence Length, and Uncertainty Measures . . . 70
2.5 Interference Between Two Partially Coherent Sources . . . 72
2.6 Second-Order Coherence . . . 73
2.6.1 Thermal Light . . . 74
2.6.2 Experiment of Hanbury Brown and Twiss . . . 75
2.7 Phase Noise . . . 76

2.7.1 Spectra of Phase and Frequency Fluctuations . . . 76
2.7.2 Variance of Phase Fluctuations . . . 77
2.7.3 Spectrum of the Signal . . . 77
2.7.3.1 Example: White Noise . . . 78
2.8 Optical Linewidth Measurements . . . 78
2.8.1 Photodetection Spectrum . . . 78
2.8.2 Heterodyne Spectroscopy . . . 79
2.8.2.1 Example: White Noise in Heterodyne Spectroscopy . . . 81
2.8.3 Self-Heterodyne Spectroscopy . . . 81
2.8.3.1 Example: White Noise in Self-Heterodyne Spectroscopy . . . 83
2.8.3.2 Calculation of General Self-Heterodyne Spectra . . . 85
2.8.3.3 Self-Heterodyne Spectrum of $1/f$ Noise . . . 86
2.8.3.4 Observation Time and Linewidth of $1/f$ Noise . . . 87
2.9 Exercises . . . 94

3 Rate-Equation Model **97**
3.1 Quantization . . . 97
3.2 Fundamental Light–Atom Interactions . . . 98
3.3 Einstein Rate Equations . . . 98
3.4 Relations Between the Einstein Coefficients . . . 99
3.5 Line Shape and Spectral Distributions . . . 100
3.5.1 Broadband Light . . . 101
3.5.2 Nearly Monochromatic Light . . . 101
3.6 Absorption Coefficient and Saturation . . . 102
3.7 Exercises . . . 104

4 The Quantum State **107**
4.1 Density Operator . . . 107
4.1.1 Example . . . 107
4.1.2 Evolution . . . 108
4.1.3 Expectation Values . . . 109
4.1.4 The Density Matrix . . . 109
4.1.5 Purity . . . 110
4.2 Pictures . . . 111
4.2.1 Unitary Time-Evolution Operator . . . 111
4.2.1.1 Infinitesimal Form . . . 111
4.2.1.2 Differential Equation for the Evolution Operator . . . 112
4.2.1.3 General Form . . . 112
4.2.2 Schrödinger vs. Heisenberg Picture . . . 113
4.2.2.1 Heisenberg Equation of Motion . . . 113
4.2.3 Interaction Picture . . . 114
4.3 Wigner Distribution . . . 115
4.3.1 Marginal Distributions . . . 115
4.3.2 Overlap . . . 116
4.3.3 Area . . . 117
4.3.4 Sample Wigner Distributions . . . 118
4.3.4.1 Gaussian State . . . 118
4.3.4.2 Coherent Superpositions . . . 119
4.3.4.3 Harmonic Oscillator States . . . 121
4.3.5 Weyl Correspondence and Operator Ordering . . . 122
4.3.5.1 Weyl's Rule . . . 123
4.3.5.2 Expectation Values . . . 123
4.3.5.3 Weyl Correspondence: Inverse Form . . . 125

4.3.5.4 Weyl Ordering . . . 125
4.3.6 Operator Products and Commutators . . . 126
4.3.7 Moyal Bracket . . . 127
4.3.8 Summary: Defining Properties . . . 128
4.3.9 Other Representations . . . 129
4.3.9.1 Husimi or Q Function . . . 129
4.3.9.2 P Function . . . 130
4.3.9.3 Standard-Ordered Distribution . . . 130
4.3.9.4 Antistandard-Ordered Distribution . . . 130
4.4 Multiple Degrees of Freedom . . . 130
4.4.1 Merging Hilbert Spaces . . . 130
4.4.2 Entanglement . . . 131
4.4.2.1 Cloning . . . 132
4.4.3 Peres–Horodecki Criterion . . . 132
4.4.3.1 Wigner Representation . . . 134
4.4.3.2 Generalized Uncertainty Relation . . . 135
4.4.3.3 Sufficiency for Gaussian States . . . 138
4.4.4 Indistinguishability . . . 141
4.4.4.1 Exchange ‘‘Force’’ . . . 141
4.4.5 Open Systems: Church of the Larger Hilbert Space . . . 143
4.5 Master Equation . . . 144
4.5.1 Interaction Representation . . . 144
4.5.2 Born–Markov Approximation . . . 145
4.5.3 Interaction . . . 146
4.6 Exercises . . . 149

5 Two-Level Atom Interacting with a Classical Field 151
5.1 Atom–Field Interaction . . . 151
5.1.1 Parity and the Dipole Operator . . . 152
5.1.2 Rotating-Wave Approximation . . . 153
5.1.3 Rabi Frequency . . . 154
5.1.4 Schrödinger Equation . . . 154
5.1.5 Rotating Frame . . . 154
5.1.5.1 Unitary Transformations . . . 155
5.1.5.2 Digression: Field Operators . . . 156
5.2 Rabi Flopping . . . 157
5.2.1 Resonant Interaction . . . 157
5.2.1.1 Example: Initially Unexcited Atom . . . 158
5.2.2 Nearly Resonant Interaction . . . 159
5.2.2.1 Example: Initially Unexcited Atom . . . 160
5.3 Dressed States . . . 161
5.3.1 Rabi Oscillations in the Dressed-State Picture . . . 162
5.3.2 Adiabatic Passage and Landau–Zener Crossings . . . 163
5.4 Bloch Sphere . . . 166
5.4.1 Atomic Timekeeping and Ramsey Fringes . . . 171
5.4.2 Spin Echoes and Photon Echoes . . . 174
5.4.3 Adiabatic Following . . . 177
5.5 Optical Bloch Equations . . . 177
5.5.1 Steady State . . . 179
5.5.2 Damped Rabi Oscillations . . . 180
5.5.2.1 Laplace Transform . . . 180
5.5.2.2 Torrey’s Solutions . . . 181
5.5.2.3 Exact Resonance . . . 183

5.5.3 Operator Form . . . 185
5.5.4 Orders of Magnitude . . . 185
5.6 Consistency with Other Models . . . 186
5.6.1 Classical Limit . . . 186
5.6.1.1 Review: Harmonic Oscillator in Quantum Mechanics . . . 186
5.6.1.2 Evolution of the Means: Damped Quantum Harmonic Oscillator . . . 187
5.6.1.3 Evolution of the Variances . . . 189
5.6.2 Rate-Equation Limit . . . 191
5.6.2.1 Saturation Intensity . . . 193
5.6.2.2 Validity of the Rate-Equation Limit . . . 194
5.7 Spectrum of Resonance Fluorescence . . . 194
5.7.1 Scattering Cross Section, Line Shape, and Power Broadening . . . 196
5.7.2 Coherent and Incoherent Scattering . . . 197
5.7.3 Quantum Regression Theorem . . . 199
5.7.3.1 Alternate Form . . . 201
5.7.4 Mollow Triplet . . . 202
5.7.4.1 Off Resonance . . . 205
5.7.4.2 Interpretations . . . 206
5.7.4.3 Energy Conservation . . . 207
5.7.4.4 Nonclassical Correlations . . . 208
5.7.5 Antibunching of Resonance Fluorescence . . . 209
5.7.6 Probe Absorption . . . 211
5.7.6.1 Autler–Townes Doublet . . . 217
5.7.6.2 Lamb Dip . . . 221
5.8 Mechanical Effects of Light on Two-Level Atoms . . . 228
5.8.1 Atom-Field Interaction . . . 228
5.8.2 Schrödinger Equation . . . 229
5.8.3 Adiabatic Approximation . . . 230
5.8.3.1 Master-Equation Approach . . . 230
5.8.3.2 Bragg Scattering in an Optical Standing Wave . . . 232
5.8.4 Nonperturbative Analysis . . . 235
5.8.5 Dressed-State Interpretation . . . 238
5.8.6 Fluctuations of the Optical Force . . . 238
5.8.6.1 Fokker–Planck Equation . . . 239
5.8.6.2 Diffusion Coefficient . . . 240
5.8.6.3 Quantum Regression Theorem . . . 241
5.8.6.4 Interpretation of the Diffusion Rate . . . 243
5.8.6.5 Dressed-State Model . . . 244
5.8.6.6 Examples: Plane and Standing Waves . . . 245
5.8.6.7 Spontaneous Emission . . . 246
5.8.7 Velocity Dependence . . . 247
5.8.8 Doppler Cooling Limit . . . 250
5.9 Bloch–Siegert Shift . . . 251
5.9.1 Magic Wavelength . . . 252
5.10 Exercises . . . 254

6 Three-Level Atom Interacting with a Classical Field 269
6.1 Stimulated Raman Transitions . . . 269
6.1.1 Effective Two-Level Dynamics . . . 271
6.1.1.1 Cross-Couplings . . . 273
6.1.2 Spontaneous Emission . . . 273
6.1.3 Multiple Excited States . . . 273
6.1.4 Velocity Selectivity . . . 274

6.1.5 Pulse-Shape Considerations . . . 275
6.1.5.1 Square Pulse . . . 275
6.1.5.2 Blackman Pulse . . . 277
6.1.6 Stimulated Raman Cooling . . . 278
6.1.6.1 Free Space . . . 278
6.1.6.2 Resolved-Sideband Raman Cooling . . . 280
6.1.7 Atom Interferometry . . . 281
6.2 Coherent Population Trapping . . . 283
6.2.1 VSCPT . . . 287
6.2.2 Electromagnetically Induced Transparency . . . 287
6.2.3 Stimulated Raman Adiabatic Passage . . . 290
6.2.4 Quantum Beats . . . 291
6.2.4.1 Master Equations and Quantum Beats . . . 292
6.2.4.2 Steady-State Quantum Beats . . . 293
6.3 Exercises . . . 296

7 Atomic Angular-Momentum Structure 301
7.1 Angular Momentum . . . 301
7.1.1 Operators and Eigenstates . . . 301
7.1.2 Ladder Operators and Eigenvalues . . . 302
7.1.3 Addition of Two Angular Momenta: Clebsch–Gordan Coefficients . . . 304
7.1.3.1 Basis States . . . 304
7.1.3.2 Transformation between Bases and Clebsch–Gordan Coefficients . . . 305
7.1.3.3 Calculation of Clebsch–Gordan Coefficients . . . 306
7.1.3.4 Explicit Formula . . . 307
7.1.3.5 Symmetry Relations and Wigner 3-j Symbols . . . 307
7.1.4 Addition of Three Angular Momenta: Racah Coefficients and Wigner 6-j Symbols . . . 309
7.1.4.1 Explicit Forms . . . 311
7.1.4.2 Symmetry Relations . . . 313
7.1.4.3 Addition of Four Angular Momenta: Wigner 9-j Symbols . . . 313
7.2 Static Angular-Momentum Structure of Atoms . . . 315
7.2.1 Fine Structure . . . 316
7.2.1.1 Spectroscopic Notation for Simple Atoms . . . 318
7.2.2 Hyperfine Structure . . . 318
7.3 Rotations and Irreducible Tensor Operators . . . 321
7.3.1 Rotation Operator . . . 321
7.3.1.1 Rotation Matrix . . . 322
7.3.1.2 Euler Angles . . . 322
7.3.1.3 Clebsch–Gordan Series . . . 324
7.3.2 Spherical Harmonics . . . 325
7.3.2.1 Sum Rule and Addition Theorem . . . 326
7.3.2.2 Relation to the Rotation Matrix . . . 327
7.3.3 Irreducible Tensor Operators . . . 328
7.3.3.1 Spherical Basis . . . 328
7.3.3.2 General Definition . . . 329
7.3.3.3 Cartesian Tensors . . . 329
7.3.3.4 Products of Tensors . . . 331
7.3.3.5 Commutation Rules . . . 333
7.3.4 Wigner–Eckart Theorem . . . 334
7.3.4.1 Dipole Operator . . . 335
7.3.4.2 Projection Theorem . . . 336
7.3.5 Hermitian Conjugates of Tensor Operators . . . 337
7.3.5.1 Conjugates of Reduced Matrix Elements . . . 337

7.3.6 Relations Between Reduced Matrix Elements of Tensor Operators . . . 338
7.3.6.1 Tensor Operator Acting on One Component . . . 338
7.3.6.2 Scalar Products of Tensor Operators . . . 339
7.3.6.3 Matrix Elements of Tensor Products Operating on the Same System . . . 340
7.3.7 Application to Atomic Transitions . . . 341
7.3.7.1 Decomposition and Calculation of Reduced Matrix Elements . . . 341
7.3.7.2 Fine-Structure Selection Rules . . . 343
7.3.7.3 Hyperfine Selection Rules . . . 344
7.3.7.4 Decay Rate and the Reduced Matrix Element . . . 345
7.4 Interaction with Static Fields . . . 346
7.4.1 Static Magnetic Fields: Zeeman Effect . . . 346
7.4.1.1 Anomalous Zeeman Effect: Weak Fields . . . 346
7.4.1.2 Paschen–Back Effect: Strong Fields . . . 348
7.4.1.3 Incomplete Paschen–Back Effect: Intermediate Fields . . . 349
7.4.2 Static Electric Fields: Stark Effect . . . 352
7.4.2.1 Effective, First-Order Interaction . . . 353
7.4.2.2 Scalar Shift: Fine Structure . . . 353
7.4.2.3 Tensor Shift: Fine Structure . . . 354
7.4.2.4 Hyperfine Structure: Weak Fields . . . 356
7.4.2.5 Hyperfine Structure: Stronger Fields . . . 358
7.5 Interactions with Optical Fields . . . 360
7.5.1 Atomic Fine-Structure Hamiltonian . . . 360
7.5.2 Dipole and Atomic Lowering Operators . . . 360
7.5.3 Dipole Interaction . . . 362
7.5.3.1 Magnetic-Sublevel Transitions: Notation . . . 363
7.5.4 Dipole Interaction: Hyperfine Structure . . . 364
7.5.4.1 Atomic Hyperfine Hamiltonian . . . 364
7.5.4.2 Atom–Field Interaction . . . 364
7.6 Angular Distribution of Dipolar Resonance Fluorescence . . . 366
7.6.1 Angular-Distribution Tensor . . . 367
7.6.2 Spectral Tensor and Total Scattered Power . . . 368
7.6.2.1 Hyperfine Structure and Interference . . . 370
7.7 Optical Stark Shifts . . . 370
7.7.1 Irreducible Parts . . . 371
7.7.1.1 Scalar Part . . . 371
7.7.1.2 Vector Part . . . 372
7.7.1.3 Tensor Part . . . 373
7.7.2 Total Shift . . . 373
7.7.2.1 Excited States . . . 374
7.7.3 Example: Stark Shifts of the $F = 1 \longrightarrow F' = 0$ Transition . . . 374
7.7.4 Polarizability Tensor Revisited . . . 376
7.7.5 Large Detuning . . . 376
7.7.5.1 Effective Dipole Moment . . . 377
7.7.5.2 Alkali Ground States . . . 379
7.8 Atomic Master Equation . . . 380
7.8.1 Fine Structure . . . 380
7.8.2 Hyperfine Structure . . . 381
7.8.3 Rate-Equation Limit . . . 382
7.8.3.1 Single Field Polarization . . . 384
7.8.3.2 Multiple Fields . . . 384
7.8.3.3 Hyperfine Structure . . . 385
7.9 Whither has Wandered the Two-Level Atom? . . . 388
7.9.1 Optical Pumping to Stretched States . . . 389

7.9.2 Optical Pumping with Linearly Polarized Light . . . 391
7.10 Exercises . . . 395

II Quantum Light–Matter Interactions 399

8 Quantization of the Electromagnetic Field 401
8.1 Classical Electromagnetic Field . . . 401
8.2 Hamiltonian Structure of the Classical Electromagnetic Field . . . 402
8.2.1 Variational Calculus . . . 402
8.2.2 Action Principles . . . 403
8.2.2.1 Lagrangian . . . 403
8.2.2.2 Hamiltonian . . . 404
8.2.3 Electromagnetic Lagrangian and Hamiltonian . . . 404
8.2.3.1 Electromagnetic Functional Derivatives . . . 406
8.3 Quantization of a Single Field Mode . . . 406
8.4 Quantization of Many Modes . . . 408
8.4.1 Example: Quantization in a Perfectly Conducting Box . . . 409
8.4.2 Example: Quantization in Free Space . . . 409
8.4.3 Example: Quantization in Half Space . . . 410
8.4.4 Example: Quantization in a Spherical Cavity . . . 411
8.4.4.1 Scalar Field . . . 411
8.4.4.2 Vector Field . . . 414
8.4.4.3 Asymptotics . . . 418
8.4.4.4 Vector Multipole Modes . . . 419
8.5 Transverse and Longitudinal Fields . . . 421
8.5.1 Helmholtz Theorem . . . 421
8.5.1.1 Coulomb Gauge . . . 423
8.5.2 Transverse and Longitudinal Delta Functions . . . 423
8.5.2.1 Momentum Representation . . . 423
8.5.2.2 Position Representation . . . 424
8.6 Field Commutators . . . 425
8.6.1 Free-Space Commutators . . . 425
8.6.1.1 Direction Cosines . . . 425
8.6.1.2 Evaluation . . . 426
8.6.1.3 Equal-Time Commutators in Free Space . . . 428
8.6.2 Half-Space Commutators . . . 429
8.7 Unconfined Mode Functions . . . 430
8.8 Hamiltonian Viewpoint of Electromagnetic Gauge Freedom . . . 431
8.8.1 Hamiltonian . . . 432
8.8.2 Hamilton Equations and Gauge Freedom . . . 433
8.8.3 Continuity Constraint . . . 434
8.9 Quantization with Dielectric Media . . . 434
8.9.1 Fresnel Coefficients: Modes of a Planar, Dielectric Interface . . . 434
8.9.1.1 Boundary Conditions . . . 435
8.9.1.2 TE Modes . . . 437
8.9.1.3 TM Modes . . . 442
8.9.2 Quantum Fields at a Planar Dielectric Interface . . . 443
8.9.2.1 Hamiltonian Structure . . . 443
8.9.2.2 Mode Functions . . . 444
8.9.2.3 Quantized Fields . . . 445
8.9.3 Transverse and Longitudinal Fields: Dielectric . . . 446
8.9.3.1 Scalar Green Function . . . 446

8.9.3.2 Transverse and Longitudinal Fields: Free Space . . . 447
8.9.3.3 Scalar Green Function: Dielectric Case . . . 448
8.9.3.4 Dielectric Transverse and Longitudinal Fields: Green-Function Form . . . 449
8.9.3.5 Completeness . . . 450
8.10 Exercises . . . 451

9 Atomic Interaction with the Quantized Field 457
9.1 Lorentz Force . . . 457
9.1.1 Lagrangian . . . 458
9.1.2 Hamiltonian . . . 458
9.2 Quantization and Minimal Coupling . . . 458
9.3 Dipole Interaction . . . 460
9.3.1 Power–Zienau Transformation . . . 460
9.3.1.1 Electric Displacement . . . 462
9.3.1.2 Active and Passive Viewpoints . . . 462
9.3.1.3 Göppert-Mayer Transformation . . . 463
9.3.2 $\mathbf{p}\cdot\mathbf{A}$ vs. $\mathbf{r}\cdot\mathbf{E}$ . . . 463
9.4 Why the Vector Potential? . . . 465
9.5 Multipole Interactions . . . 467
9.5.1 Atomic Polarization Field . . . 467
9.5.2 Atomic Magnetization Field . . . 468
9.5.3 Power–Zienau Transformation . . . 469
9.5.3.1 Electric Field . . . 469
9.5.3.2 Canonical Electron Momentum . . . 470
9.5.3.3 Hamiltonian . . . 471
9.5.4 Electric Multipole Expansion . . . 473
9.5.5 Magnetic Multipole Expansion . . . 474
9.6 Center-of-Mass Röntgen Interaction . . . 475
9.6.1 Polarization . . . 475
9.6.2 Center-of-Mass Coordinates . . . 476
9.6.3 Transformation: Electric Dipole Approximation . . . 477
9.6.4 Full Transformation . . . 479
9.6.4.1 Effecting the General Transformation . . . 479
9.6.4.2 Final Result . . . 481
9.7 Exercises . . . 483

10 Cavity QED and the Jaynes–Cummings Model 485
10.1 Single Cavity Mode . . . 485
10.2 Dynamics . . . 486
10.3 Dressed States and the Vacuum Rabi Doublet . . . 489
10.3.1 Photon Blockade . . . 490
10.3.2 Atom-Photon "Molecule" . . . 490
10.4 Refinements of the Model . . . 491
10.5 Exercises . . . 492

11 Spontaneous Emission 495
11.1 Atom–Field Coupling . . . 495
11.2 Evolution . . . 496
11.3 Large-Box Limit . . . 497
11.4 Decay Rate . . . 498
11.5 Master Equation for Spontaneous Emission . . . 500
11.6 Fermi's golden rule . . . 501
11.6.1 Free-Space Decay Rate . . . 502

11.7 Corrections to Exponential Decay . . . 503
11.7.1 Short Times . . . 503
11.7.1.1 Quantum Zeno Effect . . . 505
11.7.2 Long Times . . . 506
11.8 Exercises . . . 507

12 Coupled-Mode Theory **511**
12.1 Cavity QED . . . 511
12.1.1 Classical Field in a Single Cavity . . . 511
12.1.2 Classical Coupled Modes of Two Cavities . . . 512
12.1.3 Quantization of the Coupled Modes . . . 514
12.1.4 Cavity Driven by a Classical Field . . . 515
12.1.4.1 Cavity Decay Rate . . . 516
12.1.5 Cavity Decay . . . 517
12.1.5.1 Master Equation . . . 519
12.2 Input–Output Formalism . . . 519
12.2.1 Quantum Langevin Equation . . . 520
12.2.1.1 Evolution of the Mean . . . 522
12.2.2 Output Field . . . 522
12.2.3 Input–Output Relation . . . 523
12.2.4 General Heisenberg Equations . . . 523
12.2.5 Causality . . . 524
12.2.6 Example: Reflections from a Cavity . . . 524
12.2.7 Example: Cavity Transmission . . . 525
12.2.8 Example: Driven Cavity . . . 526
12.2.9 Example: Atomic Motion in an Optical Cavity . . . 527
12.2.9.1 Adiabatic Approximation . . . 528
12.3 Exercises . . . 531

13 Mechanical Effects of the Quantum Vacuum **533**
13.1 Setup . . . 533
13.2 Atom–Vacuum Interaction . . . 534
13.3 Renormalization . . . 536
13.4 Large-Box Limit . . . 536
13.5 Spherical Coordinates . . . 537
13.5.1 Thomas–Reiche–Kuhn Sum Rule . . . 539
13.5.2 Simplification . . . 540
13.5.3 Spherical Symmetry . . . 541
13.6 Asymptotic Behavior . . . 541
13.7 Excited-Level Shift . . . 542
13.7.1 Classical Antenna Behavior . . . 544
13.8 Power–Zienau Transformation in Half-Space . . . 545
13.9 Calculation in the Coulomb Gauge . . . 548
13.10 Evaluation . . . 551
13.11 Numerical Evaluation: ^{87}Rb . . . 552
13.11.1 Tabulated Data . . . 553
13.11.2 Results . . . 554
13.12 Lamb Shift . . . 557
13.12.1 Coulomb Gauge . . . 557
13.12.1.1 Evaluation . . . 560
13.12.2 Electric Dipole Interaction . . . 561
13.12.2.1 Dipole Self-Energy . . . 561
13.12.2.2 Mass Renormalization . . . 562

13.13 Casimir–Polder Potential for a Rarefied Dielectric Surface . . . 563
13.13.1 TE Energy Density . . . 565
13.13.1.1 Digression: Back to the Perfect Conductor . . . 566
13.13.1.2 Integral for the Rarefied Dielectric . . . 566
13.13.2 TM Energy Density . . . 567
13.13.2.1 Digression: TM Energy Density for the Perfect Conductor . . . 569
13.13.3 Total Casimir–Polder Potential for a Polarizable Atom . . . 569
13.13.4 Magnetic-Field Energies . . . 570
13.13.4.1 Magnetic TE Energy Densities . . . 571
13.13.4.2 Magnetic TM Energy Densities . . . 572
13.13.4.3 Total Magnetic and Electromagnetic Energy Densities . . . 572
13.14 Exercises . . . 574

14 QED with Dielectric Media **579**
14.1 Classical Electrodynamics in Dielectric Media . . . 579
14.1.1 Effective Sources . . . 579
14.1.2 Linear, Dispersive Media . . . 580
14.1.2.1 Frequency Domain . . . 580
14.1.2.2 Time Domain . . . 581
14.1.3 Classical Green Tensor . . . 581
14.1.3.1 Example: Green Tensor in Free Space . . . 583
14.1.3.2 Green Tensor in Free Space: Alternate Forms . . . 585
14.1.3.3 Derivation of the Formula for the Dipole Radiation Field . . . 585
14.1.4 Permittivity Properties . . . 586
14.1.4.1 Energy Loss and Poynting's Theorem . . . 587
14.1.4.2 Kramers–Kronig Relations . . . 589
14.1.4.3 Imaginary Frequencies . . . 591
14.2 Generalized Susceptibility and Linear-Response Theory . . . 592
14.2.1 Proof . . . 593
14.2.2 Atom and Field Susceptibilities . . . 594
14.3 Atom–Surface Potentials Near Dielectric Media . . . 596
14.3.1 Kramers–Heisenberg Formula . . . 596
14.3.2 Green Tensor . . . 599
14.3.2.1 Mode Expansion of the Green Tensor . . . 600
14.3.3 Interaction Energy . . . 600
14.3.4 Renormalization . . . 601
14.3.5 Planar Interface . . . 601
14.3.5.1 Reflection Coefficients . . . 602
14.3.5.2 Scattering Green Tensor Due to the Planar Interface . . . 603
14.3.5.3 Explicit Expressions for the Atom–Surface Potential . . . 604
14.3.5.4 Perfect-Conductor Limit . . . 605
14.3.5.5 Near-Field Limit . . . 606
14.3.5.6 Far-Field Limit . . . 608
14.3.5.7 Asymmetric Atoms and Molecules . . . 610
14.3.5.8 Dielectric Thin Films . . . 612
14.3.5.9 Metallic Thin Films . . . 613
14.3.6 Perfectly Conducting, Spherical Cavity . . . 615
14.3.7 Ground-State Atom–Atom Potentials . . . 616
14.3.7.1 Near-Field van der Waals–London Potential . . . 617
14.3.7.2 Far-Field Potential . . . 618
14.3.7.3 General Form for Scalar Polarizabilities . . . 619
14.3.7.4 Three-Atom Potential . . . 620
14.3.8 Temperature Dependence . . . 621

14.3.8.1 Fluctuation–Dissipation Relation . . . 621
14.3.8.2 Fluctuation–Dissipation Example: Johnson Noise . . . 624
14.3.8.3 Temperature-Dependent Shifts . . . 626
14.3.8.4 Imaginary Time and the Low-Temperature Limit . . . 629
14.3.8.5 High-Temperature Limit . . . 629
14.3.8.6 Planar Boundaries at Nonzero Temperature . . . 630
14.3.9 Excited-Level Shifts . . . 632
14.3.9.1 Example: Spherically Symmetric Atom, Perfectly Conducting Plane . . . 633
14.3.10 Lifetime Shifts . . . 635
14.3.10.1 Decay Rate Near a Macroscopic Body . . . 635
14.3.10.2 Free Space: Green-Tensor Example . . . 637
14.3.10.3 Planar Reflector . . . 637
14.4 Casimir Energy . . . 638
14.4.1 Casimir Energy of Dispersive Dielectric Bodies . . . 639
14.4.1.1 Induced Interaction Energies . . . 640
14.4.1.2 Assembling a Dielectric Medium, Atom by Atom . . . 641
14.4.1.3 Conversion to Permittivity . . . 643
14.4.1.4 Introduction of the Dielectric Green Tensor . . . 643
14.4.1.5 Explicit Permittivity Dependence . . . 644
14.4.1.6 Free-Space Energy . . . 645
14.4.1.7 Reduction to the Casimir–Polder Potential . . . 646
14.4.1.8 Temperature Dependence . . . 647
14.5 Exercises . . . 648

15 Resolvent Operator 653
15.1 Definition . . . 653
15.2 Green Functions for the Schrödinger Equation . . . 653
15.2.1 Energy-Space Green Functions . . . 653
15.2.2 Time-Dependent Green Functions and Propagators . . . 654
15.2.3 Relation to Laplace Transform . . . 656
15.3 Transitions Between Discrete States . . . 656
15.3.1 Example: Rabi Oscillations . . . 657
15.4 Level-Shift Operator . . . 659
15.4.1 Decomposition of the Level-Shift Operator . . . 660
15.4.2 Perturbation Expansion . . . 661
15.5 Spontaneous Decay . . . 662
15.5.1 Pole Approximation . . . 662
15.5.2 Line Shape of Spontaneous Decay . . . 663
15.5.3 Branches of the Resolvent . . . 665
15.5.4 Nonexponential Decay . . . 666
15.5.5 Frequency-Dependent Decay Rate . . . 668
15.5.6 Branch Contribution . . . 670
15.5.7 Pole Contribution . . . 670
15.5.8 Short Times . . . 671
15.5.8.1 Hard Cutoff . . . 672
15.5.8.2 Soft Cutoff . . . 673
15.5.9 Intermediate Times . . . 673
15.5.10 Interpretation of Nonexponential Decay . . . 675
15.6 Spontaneous Raman Scattering . . . 675
15.6.1 Weak Pumping . . . 677
15.6.2 Strong Pumping . . . 678
15.6.3 General Case . . . 679
15.7 Exercises . . . 681

16 Photodetection **685**
16.1 Counting Photons . . . 685
16.2 Beam Splitters . . . 687
16.3 Collision of One Photon and a Beam Splitter . . . 687
16.4 Two-Photon Interference . . . 689
16.4.1 Simple Theory . . . 689
16.4.2 Coherence Effects . . . 690
16.4.2.1 Quantum Beam . . . 690
16.4.2.2 Pulse-Annihilation Operators . . . 691
16.4.2.3 Detection . . . 692
16.4.2.4 Interference of Coherence . . . 692

III Stochastic Trajectories in Quantum Mechanics **695**

17 Stochastic Processes **697**
17.1 Finite Random Walks, Diffusion, and the Central Limit Theorem . . . 697
17.1.1 Two-Step Distribution . . . 698
17.1.1.1 Example 1: Convolution with a Delta Function . . . 698
17.1.1.2 Example 2: Convolution of Box Functions . . . 699
17.1.2 Convolution Theorem . . . 699
17.1.2.1 Example: Convolution of Two Gaussians . . . 700
17.1.3 Proof of the Central Limit Theorem . . . 701
17.1.3.1 Example: Square Distribution . . . 702
17.1.3.2 Application: Standard Deviation of the Mean . . . 703
17.1.4 Variances Add in Quadrature . . . 704
17.1.5 A Walk on the Cauchy Side . . . 704
17.1.6 Arbitrary Combinations of Random Variables . . . 706
17.1.6.1 Divergence Theorem . . . 706
17.1.6.2 Transformation of Surface Delta Functions . . . 706
17.1.6.3 Direct Derivation . . . 708
17.1.6.4 Chain Rule for Coordinate Transformations . . . 709
17.1.6.5 Probability Density for Combinations of Random Variables . . . 709
17.1.6.6 Example: Convolution . . . 710
17.1.6.7 Example: Quotient of Normal Deviates . . . 710
17.2 Continuous Random Walks: Wiener Process . . . 710
17.3 Itō Calculus . . . 712
17.3.1 Usage . . . 713
17.3.2 Itō Rule: Justification . . . 713
17.3.3 Ensemble Averages . . . 714
17.3.4 Correlation Function . . . 715
17.3.5 Diffusion . . . 716
17.3.5.1 Fokker–Planck Equation . . . 716
17.3.5.2 Multidimensional Fokker–Planck Equation . . . 718
17.3.6 Ornstein–Uhlenbeck Process . . . 719
17.3.6.1 Brownian Motion . . . 720
17.4 Stratonovich Calculus . . . 721
17.4.1 Example: Stochastic Integration . . . 722
17.4.2 Itō–Stratonovich Conversion . . . 723
17.4.3 Stratonovich Calculus and the Chain Rule . . . 724
17.4.4 Comparison . . . 725
17.5 Poisson Process . . . 726
17.5.1 The Poisson Process Implies the Poisson Distribution . . . 727

17.5.2 Inhomogeneous Poisson Process and State Dependence . . . 728
17.5.3 White-Noise Limit . . . 728
17.5.3.1 Shot Noise . . . 729
17.6 Particle Subject to a Stochastic Force . . . 731
17.6.1 Free-Particle Limit . . . 731
17.6.2 Thermal Equilibrium . . . 733
17.6.3 Strong-Damping Limit: Brownian Motion Revisited . . . 734
17.6.3.1 Momentum Distribution . . . 736
17.7 Stochastic Boundary-Value Problems: Brownian Bridges . . . 737
17.7.1 Finite Bridge Generation: Homogeneous Case . . . 740
17.7.2 Finite Bridge Generation: Inhomogeneous Case . . . 742
17.7.3 SDE and Integral Representations of the Brownian Bridge . . . 744
17.7.4 State-Dependent Diffusion in Brownian Bridges . . . 745
17.7.4.1 Drift . . . 746
17.7.4.2 Lamperti Transform . . . 747
17.7.4.3 Temporal Rescaling . . . 748
17.8 Boundary Crossings . . . 749
17.8.1 Wiener Process . . . 749
17.8.2 Standard Brownian Bridge . . . 750
17.8.3 Brownian Bridge . . . 752
17.8.4 First-Passage Time of the Brownian Bridge . . . 752
17.9 Escape Probability . . . 754
17.9.1 Wiener Process . . . 754
17.9.2 Standard Brownian Bridge . . . 756
17.9.3 Brownian Bridge . . . 758
17.10 Dirichlet Problem and Connection to Electrostatics . . . 758
17.10.1 Laplace Equation . . . 758
17.10.2 Laplacian as Spherical Average . . . 759
17.10.3 Poisson Equation . . . 760
17.11 Feynman–Kac Formula . . . 761
17.11.1 Proof: Simple Diffusion . . . 762
17.11.2 Proof: Diffusion with Damping . . . 763
17.11.3 Proof: Diffusion with Source . . . 764
17.11.4 Proof: Diffusion with Damping and Source . . . 764
17.11.5 Other Forms . . . 765
17.11.5.1 General form for State-Dependent Diffusion . . . 766
17.11.5.2 Time-Dependent Drift and Diffusion: First Form . . . 767
17.11.5.3 Time-Dependent Drift and Diffusion: Second Form . . . 769
17.12 Sojourn Times . . . 770
17.12.1 Wiener Process . . . 770
17.12.1.1 Arcsine Laws . . . 775
17.12.1.2 Alternate Arcsine Law: Last Crossing time . . . 775
17.12.1.3 Alternate Arcsine Law: Maximum-Value Time . . . 776
17.12.2 Standard Brownian Bridge . . . 777
17.12.3 Brownian Bridge . . . 783
17.12.4 Path-Pinning Normalization as a Constrained Integration . . . 791
17.13 Local Time . . . 793
17.13.1 Wiener Process . . . 795
17.13.2 Standard Brownian Bridge . . . 797
17.13.3 Brownian Bridge . . . 797
17.13.3.1 Moments . . . 800
17.13.3.2 Moment-Generating Function . . . 800
17.13.3.3 Application to the Sojourn Time . . . 801

17.13.4 Local Time and Discontinuities in Stochastic Processes . . . 802
17.13.4.1 Reflected Brownian Motion . . . 802
17.13.4.2 Discontinuous Diffusion Rate . . . 802
17.13.4.3 Skew Brownian Motion . . . 803
17.13.4.4 Reflections, Images, and Transition Densities for Skew Brownian Motion . 804
17.13.4.5 Stratonovich Discontinuous Diffusion . . . 806
17.14 Bessel Processes . . . 807
17.14.1 Generator . . . 808
17.14.2 Brownian Recurrence . . . 809
17.14.2.1 One Dimension . . . 809
17.14.2.2 Two Dimensions . . . 809
17.14.2.3 Three and More Dimensions . . . 810
17.15 Differentiation of Stochastic Quantities . . . 812
17.15.1 Parameter Differentiation and Likelihood Estimation . . . 812
17.15.1.1 Likelihood . . . 813
17.15.1.2 Application: Differentiation of Brownian-Bridge Path Averages . . . 813
17.15.1.3 Variance Reduction . . . 815
17.15.2 Stochastic Differentiation . . . 819
17.15.2.1 Example: Stochastic Derivative of the Wiener Process . . . 819
17.15.2.2 Example: Stochastic Derivative of Additive Diffusion . . . 819
17.15.2.3 Variation Process . . . 820
17.15.3 Integration by Parts . . . 821
17.15.3.1 Digression: Faddeev–Popov Approach to Conditional Averages . . . 823
17.16 Exercises . . . 826

18 Quantum Trajectories for Photodetection 833
18.1 Quantum Jumps . . . 833
18.1.1 Ensemble Average . . . 835
18.1.2 Quantum Trajectories and the Stochastic Schrödinger Equation . . . 836
18.1.3 Information Gain . . . 836
18.1.4 Monte Carlo Trajectories . . . 837
18.1.5 Detector Efficiency . . . 839
18.2 Homodyne Detection . . . 840
18.2.1 State Collapse . . . 842
18.2.2 Quantum-State Diffusion . . . 842
18.2.3 Measurement Record . . . 845
18.2.4 Information Gain from the Measurement Record . . . 846
18.2.5 Diffusion Form of the Stochastic Schrödinger Equation . . . 847
18.2.6 Balanced Homodyne Detection . . . 847
18.2.6.1 Master Equation . . . 849
18.2.6.2 Measurement Record . . . 851
18.2.7 Heterodyne Detection . . . 852
18.2.8 Detector Efficiency and Multiple Observers . . . 853
18.3 Conditioned Dynamics and Squeezing . . . 855
18.3.1 Moment Equations . . . 856
18.3.2 Quadrature Moments . . . 857
18.3.3 Interpretation . . . 858
18.3.4 Squeezing (or Lack Thereof) . . . 858
18.3.5 Homodyne Detection . . . 859
18.3.6 Heterodyne Detection . . . 859
18.3.7 Explicit Solutions for the Uncertainty Dynamics . . . 861
18.3.8 Phase Estimation . . . 861
18.3.8.1 Adaptive Measurements . . . 863

18.4 Exercises . . . 864

19 Position Measurement **865**
19.1 Prelude: General Form for the Master Equation . . . 865
19.1.1 Positive Maps . . . 865
19.1.2 Lindblad Form . . . 866
19.1.3 Stochastic Terms . . . 867
19.1.4 Generalization . . . 867
19.2 A Second Prelude: Positive-Operator-Valued Measures . . . 868
19.2.1 Discrete, Finite Spaces . . . 868
19.2.2 Measure . . . 869
19.2.3 General Definition . . . 871
19.2.4 Realization . . . 872
19.2.5 Example: Spontaneous Emission . . . 873
19.2.6 Example: Gaussian Projectors . . . 873
19.3 Continuous Position Measurement . . . 875
19.3.1 State Collapse and the Stochastic Schrödinger Equation . . . 876
19.3.1.1 Gaussian Noise . . . 877
19.3.2 Stochastic Master Equation . . . 877
19.3.3 Inefficient Detection and Multiple Observers . . . 878
19.3.4 Interpretation . . . 879
19.3.5 Linear Stochastic Evolution Equations . . . 881
19.3.5.1 Norm of the Linear Solution . . . 882
19.3.5.2 Interpretation of the Solution . . . 882
19.3.5.3 Explicit Solutions of Measurement Dynamics . . . 883
19.4 Imaged Resonance Fluorescence as a Position Measurement . . . 885
19.4.1 Center-of-Mass Dynamics . . . 885
19.4.2 Imaging . . . 886
19.4.2.1 Example: 4π Detection . . . 890
19.4.2.2 Example: Small Gaussian Aperture . . . 890
19.4.3 Adiabatic Elimination of the Internal Atomic State . . . 891
19.4.3.1 Internal Quasi-Equilibrium . . . 891
19.4.3.2 External Master Equation . . . 892
19.4.4 White-Noise Limit: Gaussian Aperture . . . 893
19.4.4.1 Spatial Continuum Approximation . . . 893
19.4.4.2 Quantum-State Diffusion . . . 894
19.4.4.3 Diffusion Rates . . . 895
19.5 Position Measurement via Excitation by a Local Probe Field . . . 896
19.5.1 Localized Probe Field . . . 896
19.5.2 Scanning Probe Field . . . 898
19.5.3 Example: Gaussian Probe . . . 898
19.6 Continuous Momentum Measurement by EIT . . . 899
19.6.1 General Remarks . . . 899
19.6.2 Homodyne Detection of the EIT Probe . . . 900
19.6.3 Adiabatic Approximation . . . 901
19.6.4 Spontaneous Scattering . . . 901
19.6.5 Phase . . . 902
19.6.6 Detection Efficiency . . . 903
19.7 Exercises . . . 905

20 Path Integration **907**
20.1 Feynman Path Integral . . . 907
20.1.1 Dividing the Evolution . . . 908

20.1.2 Splitting the Evolution Operator . . . 908
20.1.3 Functional Integral . . . 910
20.1.4 Phase-Space Path Integral . . . 910
20.1.5 Example: Free-Particle Propagator . . . 911
20.2 Classical Limit . . . 913
20.2.1 Semiclassical Propagator . . . 914
20.2.1.1 Gel'fand–Yaglom Method . . . 917
20.2.1.2 Van Vleck–Morette Determinant . . . 918
20.2.2 Example: Semiclassical Propagator for the Harmonic Oscillator . . . 921
20.2.2.1 Classical Action . . . 921
20.2.2.2 Determinant . . . 922
20.3 Monte-Carlo Methods in Quantum Statistical Mechanics . . . 922
20.3.1 Path Integral as Ensemble Average . . . 923
20.3.1.1 Alternate Normalization of the Periodic Path . . . 925
20.3.1.2 YANPP (Yet Another Normalization of the Periodic Path) . . . 926
20.3.2 Feynman–Kac Formula . . . 927
20.3.3 Thermal Density Matrix . . . 928
20.3.3.1 Example: Free-Particle Thermal State . . . 930
20.3.3.2 Example: Partition Function of an Infinite Square Well . . . 930
20.3.4 Propagator Paths . . . 932
20.4 Ordering Issues . . . 933
20.4.1 Model Problem . . . 933
20.4.1.1 Basic Structure . . . 933
20.4.1.2 Coordinate Transformation and Curvature . . . 934
20.4.1.3 "Natural" Ordering . . . 937
20.4.2 Point Transformations of the Path Integral . . . 938
20.4.2.1 Midpoint (Stratonovich) Expansion of the Path Integral . . . 939
20.4.2.2 Pre-Point (Itō) Expansion of the Path Integral . . . 942
20.4.2.3 "Moment" Relations . . . 944
20.4.2.4 Midpoint–Pre-point Conversion . . . 946
20.4.2.5 Discussion: Integration Under Different Path Measures . . . 948
20.4.2.6 Cameron–Martin–Girsanov Transformation and Drift Potentials . . . 949
20.4.3 Operator Orderings . . . 950
20.4.3.1 Itō Calculus and **p**-**x** Ordering . . . 951
20.4.3.2 Anticipating Calculus and **x**-**p** Ordering . . . 954
20.4.3.3 Stratonovich Calculus and Weyl Ordering . . . 954
20.4.3.4 Symmetrized Ordering . . . 957
20.4.3.5 Product Ordering . . . 959
20.4.3.6 Equivalence Classes of Orderings . . . 960
20.4.4 Normalizing the Weyl-Ordered Path Integral . . . 960
20.4.4.1 A Similar Path Integral . . . 962
20.5 Exercises . . . 963

21 Path-Integral Calculation of Casimir Energies **965**
21.1 Scalar Theory . . . 965
21.1.1 Quantum Scalar Field . . . 965
21.1.2 Partition Function . . . 966
21.1.2.1 Free Energy . . . 967
21.1.2.2 Partition-Function Renormalization . . . 968
21.1.3 Path Integral of the Quantum Field . . . 968
21.1.3.1 Momentum Projectors and Imaginary Time . . . 969
21.1.4 Reduction of the Hilbert Space . . . 970
21.1.4.1 Evaluation of the Gaussian Functional Integral . . . 970

21.1.4.2 Digression: Second Quantization and Mode Summation . . . 972
21.1.4.3 Digression Part II: Mode Summation of the Functional Determinant . . . 974
21.1.4.4 Integral Representation of the Logarithm . . . 975
21.1.4.5 Particle Path Integral . . . 977
21.1.4.6 Monte-Carlo Integration and Stochastic ‘‘Loops’’ . . . 979
21.1.5 Analytic Evaluation of Scalar Casimir Energies . . . 980
21.1.5.1 Strong-Coupling Limit: Atom–Plane Interaction . . . 980
21.1.5.2 Strong-Coupling Limit: Atomic Interaction with Two Planes . . . 983
21.1.5.3 Strong-Coupling Limit: Plane–Plane Interaction . . . 984
21.1.6 World Lines at Finite Temperature . . . 987
21.1.6.1 Example: Temperature-Dependent Atom–Planar-Conductor Potential . . . 989
21.2 Worldlines and the Relativistic Scalar Particle . . . 989
21.2.1 Action . . . 989
21.2.2 Reparameterization Independence . . . 991
21.2.3 Quadratic Action . . . 992
21.2.3.1 Massless-Particle World Lines . . . 994
21.2.3.2 Variable-Mass Potential . . . 994
21.2.4 Constrained Action . . . 995
21.2.4.1 Gauge Fixing . . . 996
21.2.4.2 Terminology . . . 997
21.2.4.3 Variable-Mass Potential . . . 997
21.2.5 Klein–Gordon Equation . . . 998
21.2.6 Path Integral . . . 998
21.2.6.1 Proper-Time Path Integral . . . 999
21.2.6.2 Imaginary Time . . . 1001
21.3 Exercises . . . 1002

22 Electromagnetic Casimir Energies as Path Integrals 1005
22.1 Scalar Representation of Electromagnetism . . . 1005
22.1.1 Planar Interface: TE polarization . . . 1005
22.1.2 Planar Interface: TM polarization . . . 1007
22.2 Interaction Energy . . . 1008
22.2.1 Maxwell Equations in Magnetodielectric Materials . . . 1009
22.2.2 Interaction Energies for Magnetodielectric Materials: Static Fields . . . 1011
22.2.2.1 Static-Electric-Field Energies . . . 1011
22.2.2.2 Static-Magnetic-Field Energies . . . 1012
22.2.2.3 Variation of Dielectric Materials . . . 1014
22.2.2.4 Variation of Magnetic Materials . . . 1015
22.2.2.5 Interpretation of Minus Signs . . . 1016
22.2.2.6 Small Material Perturbations . . . 1018
22.2.2.7 Application to Casimir–Polder Potentials . . . 1019
22.2.2.8 Functional Derivatives and Expectation Values: A Second Look . . . 1020
22.2.3 Hamiltonian Structure of Electromagnetism and Linear Materials . . . 1021
22.2.3.1 Material Perturbations . . . 1022
22.3 Development of the Path Integrals . . . 1024
22.3.1 Gaussian Field Integral . . . 1024
22.3.2 Imaginary-Time Green Operator . . . 1025
22.3.3 Rescaling the Green Operator: Coupling to the Potential . . . 1025
22.3.4 Worldline Form of the Path Integral . . . 1027
22.3.4.1 Recap: Normalization of the Particle Path Integral . . . 1027
22.3.4.2 Normalization of the Worldline Path Integral . . . 1028
22.3.4.3 Alternate Derivation of the Normalized Path Integral: Space-Dependent Mass . . . 1030

22.3.5 Worldline Path Integral for Casimir–Polder Potentials . . . 1031
22.3.5.1 Interaction with a Point Particle: Path-Integral View (TE Polarization) . 1034
22.3.5.2 Interaction with a Point Particle: Path-Integral View (TM Polarization) . 1036
22.4 Casimir–Polder Potential Near a Dielectric Half-Space . . . 1037
22.4.1 TE Polarization: Strong Coupling . . . 1037
22.4.2 TE Polarization: Weak Coupling . . . 1038
22.4.3 TE Polarization: General Coupling . . . 1039
22.4.4 TM Polarization: Strong Coupling . . . 1040
22.4.4.1 Fluctuations and the Interface Potential . . . 1042
22.4.5 TM Polarization: Weak Coupling . . . 1043
22.5 Fresnel Reflection Coefficients . . . 1044
22.5.1 TE Reflection Coefficient: Vacuum–Dielectric Interface . . . 1044
22.5.1.1 Vacuum-Side Reflection (TE) . . . 1045
22.5.1.2 Dielectric-Side Reflection (TE) . . . 1045
22.5.2 TM Reflection Coefficient: Vacuum–Dielectric Interface . . . 1046
22.5.2.1 Vacuum-Side Reflection (TM) . . . 1046
22.5.2.2 Dielectric-Side Reflection (TM) . . . 1046
22.6 Laplace–Mellin Method for Evaluating Path Integrals . . . 1047
22.6.1 Laplace and Mellin Transforms . . . 1047
22.6.2 Laplace Transform and Inverse Moments . . . 1048
22.6.3 Feynman–Kac Formula: Review . . . 1049
22.6.4 Application to Worldline Path Integrals . . . 1049
22.7 TE Casimir–Polder Path Integral: Dielectric Interface . . . 1050
22.7.1 Feynman–Kac Formula . . . 1050
22.7.2 Mellin Transform . . . 1051
22.7.3 Inverse Moments . . . 1052
22.7.4 Result: 3D Electromagnetism . . . 1053
22.7.5 Result: 1D Electromagnetism . . . 1054
22.8 Evaluation of the TM Interface Potential . . . 1055
22.8.1 Regularized TM Potential . . . 1056
22.8.2 Model Path Integral . . . 1057
22.8.3 Feynman–Kac Formula . . . 1057
22.8.3.1 Setup . . . 1057
22.8.3.2 Boundary Matching . . . 1058
22.8.3.3 Overall Scaling of the Potential . . . 1059
22.8.3.4 Sharp-Interface Limit . . . 1059
22.8.4 Weak-Coupling Expansion . . . 1060
22.8.5 Transfer-Layer Formalism . . . 1061
22.8.6 Mellin Transform . . . 1062
22.8.6.1 Comparison to Strong-Coupling . . . 1063
22.8.7 Generalization for Open Bridges . . . 1064
22.8.7.1 Feynman–Kac Formula . . . 1064
22.8.7.2 Inversion of the Laplace Transform . . . 1065
22.8.8 Assembling Closed Bridges . . . 1067
22.9 TM Casimir–Polder Path Integral: Dielectric Interface . . . 1069
22.9.1 Feynman–Kac Formula . . . 1069
22.9.2 Mellin Transform . . . 1070
22.9.3 Inverse Moments . . . 1070
22.9.4 Result: 3D Electromagnetism . . . 1071
22.9.4.1 Digression: Derivative-Free Path Integrals . . . 1072
22.9.5 Result: 1D Electromagnetism . . . 1074
22.9.6 Casimir–Polder Potential: Full Electromagnetism . . . 1075
22.10 Casimir–Polder Potentials Within Media . . . 1076

22.10.1 TE Polarization . . . 1076
22.10.1.1 Feynman–Kac Formula . . . 1076
22.10.1.2 Mellin Transform . . . 1076
22.10.1.3 Inverse Moments . . . 1077
22.10.1.4 Result: 3D Electromagnetism . . . 1078
22.10.1.5 Result: 1D Electromagnetism . . . 1079
22.10.2 TM Polarization . . . 1079
22.10.2.1 Feynman–Kac Formula . . . 1079
22.10.2.2 Mellin Transform . . . 1079
22.10.2.3 Inverse Moments . . . 1080
22.10.2.4 Result: 3D Electromagnetism . . . 1081
22.10.2.5 Casimir–Polder Efficiencies . . . 1082
22.11 Electric vs. Magnetic Casimir–Polder Interactions . . . 1084
22.11.1 Dielectric Particle, Magnetic Surface . . . 1084
22.11.1.1 Weak-Coupling Limit . . . 1084
22.11.2 Magnetic Particle, Magnetic Surface . . . 1085
22.11.3 Magnetic Particle, Dielectric Surface . . . 1085
22.12 Casimir Potential . . . 1086
22.12.1 TE Path Integral . . . 1087
22.12.1.1 Feynman–Kac Formula . . . 1087
22.12.1.2 Mellin Transform . . . 1090
22.12.1.3 Inverse Moments . . . 1090
22.12.1.4 Spatial Integration and Renormalization . . . 1091
22.12.1.5 Casimir Energy . . . 1093
22.12.1.6 Small-χ Limit . . . 1095
22.12.1.7 Casimir–Polder Potential: Two Parallel Planes . . . 1096
22.12.2 TM Path Integral . . . 1098
22.12.2.1 Small-χ Expansion . . . 1098
22.12.2.2 Feynman–Kac Formula . . . 1100
22.13 Exercises . . . 1101

IV Numerical Methods in Quantum Optics 1107

23 Welcome to the Machine 1109
23.1 Finite Representations of Real Numbers . . . 1109
23.2 Machine Structure and Optimization . . . 1110
23.2.1 Memory Hierarchy . . . 1110
23.2.2 Pipeline . . . 1113
23.2.2.1 Out-of-Order Execution . . . 1114
23.2.2.2 Loop Unrolling . . . 1114
23.2.2.3 Branch Prediction . . . 1115
23.2.2.4 Addition, Multiplication, and Division . . . 1116
23.2.3 Avoiding Overhead . . . 1117
23.2.3.1 Procedure Inlining . . . 1117
23.2.3.2 Compiler Issues . . . 1117
23.2.4 Parallel Programming . . . 1118
23.2.5 Tuning Your Code . . . 1118

24 Ordinary Differential Equations 1121
24.1 Convergence . . . 1121
24.1.1 Sequences . . . 1121
24.1.2 O and o . . . 1121

24.1.2.1 Example . . . 1122
24.1.3 Convergence and Scaling of Functions . . . 1122
24.1.3.1 Truncation Error . . . 1122
24.1.3.2 Example . . . 1123
24.2 Euler Methods . . . 1123
24.2.1 Local and Global Truncation Error . . . 1123
24.2.2 Implicit Euler Method and Stiff Equations . . . 1124
24.3 Runge–Kutta Methods . . . 1125
24.3.1 Second-Order Methods . . . 1125
24.3.1.1 Variations . . . 1126
24.3.2 Fourth-Order and General Schemes . . . 1127
24.3.3 Implicit Runge–Kutta Methods . . . 1129
24.3.4 Adaptive Stepping . . . 1130
24.4 Multistep and Predictor–Corrector Methods . . . 1131
24.5 Exercises . . . 1132

25 Fourier Transforms 1133
25.1 Sampling Theorem . . . 1133
25.1.1 Critical Sampling . . . 1134
25.1.2 Reconstruction . . . 1135
25.2 Discrete Fourier Transform . . . 1136
25.2.1 Periodicity and Transform Ordering . . . 1138
25.3 Aliasing . . . 1139
25.4 Fast Fourier Transform . . . 1140
25.5 Conventions . . . 1141
25.5.1 Temporal Signals . . . 1141
25.5.2 Temporal Correlation Functions . . . 1142
25.5.3 Standard Frequency . . . 1143
25.5.4 Wave Functions . . . 1143
25.6 Discrete Wigner Transform . . . 1145

26 Split-Operator Methods 1149
26.1 Splitting the Unitary Evolution Operator . . . 1149
26.2 Time-Dependent Potentials . . . 1151
26.3 Richardson Extrapolation . . . 1152
26.3.1 Numerical Test . . . 1153
26.4 Unitary Evolvers . . . 1154
26.4.1 Hierarchical Construction . . . 1154
26.4.1.1 High-Order Methods with Reduced Substep Intervals . . . 1156
26.4.1.2 High-Order Minimal-Product Methods . . . 1157
26.4.1.3 High-Order Treatment of Time-Dependent Potentials . . . 1158
26.4.1.4 Numerical Test: Fourth Order . . . 1158
26.4.1.5 Numerical Test: Sixth Order . . . 1159
26.4.2 Nonlinear Schrödinger Equations . . . 1160
26.4.3 Symplectic Integration . . . 1161
26.4.3.1 Euler–Cromer Method . . . 1162
26.4.3.2 Verlet Method . . . 1163
26.4.3.3 Higher Order Methods . . . 1164
26.4.3.4 Time-Dependent Potentials . . . 1165
26.5 Exercises . . . 1166

27 Stochastic Differential Equations 1167
27.1 Stochastic Euler Method . . . 1167

27.1.1 Truncation Error . . . 1168
27.2 Milstein Method . . . 1169
27.2.1 Multiplicative vs. Additive Noise . . . 1169
27.3 Stochastic Taylor Expansion . . . 1169
27.3.1 Single and Double Integrals . . . 1170
27.3.2 Multiple Itō Integrals . . . 1171
27.3.3 Expression for the Taylor Expansion . . . 1172
27.3.4 Multiple Wiener Processes . . . 1172
27.4 Stochastic Runge–Kutta Methods . . . 1172
27.5 Implicit Schemes . . . 1174
27.6 Strong and Weak Convergence . . . 1176
27.7 Consistent Brownian Paths . . . 1177
27.7.1 Consistent Multiple Itō Integrals . . . 1178
27.7.1.1 Lévy Areas . . . 1179
27.7.1.2 Direct Refinement . . . 1180
27.8 Random Numbers . . . 1181
27.8.1 Uniform Distribution . . . 1181
27.8.1.1 L'Ecuyer's Multiple Recursive Generator . . . 1182
27.8.1.2 Knuth's Lagged-Fibonacci Generator . . . 1184
27.8.1.3 Mersenne Twister . . . 1184
27.8.1.4 Randomizing Random Numbers . . . 1184
27.8.2 Gaussian Distribution . . . 1185
27.8.3 Other Distributions . . . 1186
27.8.3.1 Angular Distributions . . . 1186
27.8.3.2 Exponential Distributions . . . 1188
27.8.3.3 Power-Law Distributions . . . 1188
27.8.3.4 Gamma Distribution . . . 1189
27.9 Exercises . . . 1191

Index **1193**

Part I

Classical and Semiclassical Light–Matter Interactions

Chapter 1

Classical Atom–Field Interactions

We will now model the interaction between light and atoms, using a classical model of the atom. This will allow us to treat a variety of phenomena from the refractive index of atomic vapors to the conductivity of metals to laser cooling and trapping of atoms.

We will model the atom as a classical harmonic oscillator, an electron bound to the nucleus by a harmonic force (linear spring):

$$m\ddot{\mathbf{x}} + m\omega_0^2\mathbf{x} = 0. \tag{1.1}$$

Here, $\mathbf{x}$ represents the *average* position of the electron, since quantum-mechanically, the electron is not localized, and ω_0 is the resonant frequency of the harmonic potential. The above equation is also in center-of-mass coordinates, so that we can ignore the motion of the nucleus. Thus, m is the *reduced mass* of the electron, given by

$$m = \frac{m_\mathrm{e} m_\mathrm{n}}{m_\mathrm{e} + m_\mathrm{n}}, \tag{1.2}$$

where m_e is the electron mass, and m_n is the nuclear mass. Generally $m_\mathrm{e} \ll m_\mathrm{n}$, so

$$m \approx m_\mathrm{e}\left(1 - \frac{m_\mathrm{e}}{m_\mathrm{n}}\right), \tag{1.3}$$

and generally, it is a good approximation to use $m \approx m_\mathrm{e}$.

Why use a classical calculation, when an atom is a manifestly quantum-mechanical object? It turns out that the classical calculation gets many phenomena correct, and these results can be justified by quantum calculations. Essentially, the classical calculation is good for weak atomic excitations, when the harmonic potential, the lowest-order approximation to an arbitrary potential, is an accurate model. (It is even a good approximation to treat the quantum electromagnetic field classically as long as many photons are present, since the field turns out to be a set of harmonic oscillators, which are ‘‘not very quantum-mechanical.’’ Then our requirement of weak excitation of the atom implies an atom–field coupling that is in some sense very weak; we will see that this is true when discussing the atomic cross section in Section 1.2.1.) In particular, the classical model does not predict any saturation effects, and as we will see, it requires a bit of patching to make it quantitatively correct, even in the limit of small intensity.

1.1 Polarizability

We will now consider the interaction of the atom with a monochromatic field of the form

$$\mathbf{E}^{(+)}(t) = \hat{\varepsilon}E_0^{(+)}e^{-i\omega t}, \tag{1.4}$$

where $\hat{\varepsilon}$ is the unit polarization vector. Here, we are using the complex notation for the field, where we separate according to the positive- and negative-frequency components:

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) &= \mathbf{E}(\mathbf{r})\cos(\omega t+\phi)\\ &= \mathbf{E}(\mathbf{r})\frac{e^{-i\phi}}{2}e^{-i\omega t}+\mathbf{E}(\mathbf{r})\frac{e^{i\phi}}{2}e^{i\omega t}\\ &=: \mathbf{E}^{(+)}(\mathbf{r})e^{-i\omega t}+\mathbf{E}^{(-)}(\mathbf{r})e^{i\omega t}.\end{aligned} \tag{1.5}$$

Recall that we are defining $\mathbf{E}^{(\pm)}$ to go with $e^{\mp i\omega t}$, since by convention $e^{-i\omega t}$ corresponds to the *positive* frequency ω and $e^{i\omega t}=e^{-i(-\omega)t}$ corresponds to the *negative* frequency $(-\omega)$. The physical field is just the sum of the positive- and negative-frequency parts. But notice that these parts are complex conjugates, as is required to get a real (physical) field. Thus, we can always write the physical field as $E^{(+)}$ with its conjugate:

$$\mathbf{E}(\mathbf{r},t)=\mathbf{E}^{(+)}(\mathbf{r})e^{-i\omega t}+\text{c.c.}=2\text{Re}\left\{\mathbf{E}^{(+)}(\mathbf{r})e^{-i\omega t}\right\}. \tag{1.6}$$

Of course, we apply this notation to all other quantities driven by the field, such as the displacement of the electron that we consider below. Mathematically, it is simpler to keep only one part of the solution, but to obtain the physical result, you always need to add the complex conjugate (assuming that all the calculations are linear). Note that *classically*, this decomposition arises as a mathematical convenience. As we will see much later, in the quantum treatment of the field this decomposition is more fundamental and significant, since the two components will play the roles of photon creation and annihilation operators.

In writing down the expression (1.4), we are making the **dipole approximation**: we are assuming that the size of the atom is much smaller than the optical wavelength, so that the electron only sees the field at the nuclear position. Thus, we need not consider the spatial dependence or propagation direction of the field. The force on the electron due to the field is

$$\mathbf{F}^{(+)}=-e\mathbf{E}^{(+)}, \tag{1.7}$$

where e is the **fundamental charge**, the magnitude of the electron charge (so that the electron charge is $-e$).

Then the equation of motion for the electron becomes

$$m\ddot{\mathbf{x}}^{(+)}+m\omega_0^2\mathbf{x}^{(+)}=-\hat{\varepsilon}eE_0^{(+)}e^{-i\omega t}. \tag{1.8}$$

We need only worry about the electron motion in the direction of the electric field; we will ignore any motion except that induced by the field, as we will justify when considering the damped version of the harmonic oscillator.

We will now make the *ansatz* that the solution has the same time dependence as the field:

$$\mathbf{x}^{(+)}(t)=\hat{\varepsilon}x_0^{(+)}e^{-i\omega t}. \tag{1.9}$$

With this solution, Eq. (1.8) becomes

$$-m\omega^2x_0^{(+)}+m\omega_0^2x_0^{(+)}=-eE_0^{(+)}, \tag{1.10}$$

which we can solve for $x_0^{(+)}$ to obtain the solution

$$x_0^{(+)}=\frac{eE_0^{(+)}/m}{\omega^2-\omega_0^2}. \tag{1.11}$$

Again, we are breaking the electron displacement into its positive and negative components $x(t)=x^{(+)}(t)+x^{(-)}(t)$.

The dipole moment of the atom is

$$\mathbf{d}^{(+)}=-e\mathbf{x}^{(+)}, \tag{1.12}$$

where $\mathbf{x} = \hat{\varepsilon}x$. Since the dipole moment is induced by the field (the electron displacement is zero in equilibrium), we can define the **polarizability** α to describe how easily the field induces the dipole moment by

$$\mathbf{d}^{(+)} = \alpha(\omega)\mathbf{E}^{(+)}. \tag{1.13}$$
(polarizability definition)

From Eqs. (1.11) and (1.12), we can write the polarizability as

$$\alpha(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2}. \tag{1.14}$$
(classical polarizability)

The polarizability completely characterizes the response of the atom to the applied field. Of course, this is the frequency-space response function, which we have obtained via an implicit Fourier transform of the applied field.

1.1.1 Connection to Dielectric Media

Recall that the polarization density $\mathbf{P}$ is the dipole moment per unit volume. Thus, for an atomic vapor of number density N,

$$\mathbf{P}^{(+)} = N\mathbf{d}^{(+)} = N\alpha(\omega)\mathbf{E}^{(+)} = \hat{\varepsilon}\frac{Ne^2/m}{\omega_0^2 - \omega^2}E_0^{(+)}e^{-i\omega t}. \tag{1.15}$$

This expression is valid for a *rarefied* medium, where the interactions between the atoms are negligible. In dense media, correlations between dipoles cause deviations from these results. We can thus write the susceptibility for the vapor as

$$\chi(\omega) = \frac{Ne^2/m\epsilon_0}{\omega_0^2 - \omega^2}, \tag{1.16}$$
(classical susceptibility)

in view of the defining relation $\mathbf{P} = \epsilon_0\chi\mathbf{E}$. Keeping the polarizability as the fundamental microscopic quantity, we can of course also write

$$\chi(\omega) = \frac{N}{\epsilon_0}\alpha(\omega) \tag{1.17}$$
(susceptibility–polarizability relation)

for the susceptibility of a vapor of number density N in terms of the polarizability.

1.2 Damping: Lorentz Model

A better model of the atom is a *damped* harmonic oscillator. This improved model is known as the **Lorentz model** of the atom, and the equation of motion is

$$m\ddot{\mathbf{x}}^{(+)} + m\gamma\dot{\mathbf{x}}^{(+)} + m\omega_0^2\mathbf{x}^{(+)} = -\hat{\varepsilon}eE_0^{(+)}e^{-i\omega t}. \tag{1.18}$$
(Lorentz model)

The damping ("friction") term models radiation reaction due to the charge acceleration (the classical analogue of spontaneous emission) and collisions with other atoms. A quantum-mechanical calculation shows that for an isolated atom, the damping rate is the same as the Einstein A coefficient (spontaneous emission rate): $\gamma = A_{21}$.

Again, we assume a solution of the form $\mathbf{x}^{(+)}(t) = \hat{\varepsilon}x_0^{(+)}e^{-i\omega t}$. Following the method above, the solution is

$$x_0^{(+)} = \frac{eE_0^{(+)}/m}{\omega^2 - \omega_0^2 + i\gamma\omega}. \tag{1.19}$$

Now the displacement is complex, reflecting a *phase lag* of the displacement behind the field, with phase angle

$$\delta = \tan^{-1}\left(\frac{\gamma\omega}{\omega_0^2 - \omega^2}\right). \tag{1.20}$$

The phase lag approaches zero for $\omega \ll \omega_0$ and π for $\omega \gg \omega_0$ ($\delta = \pi/2$ exactly on resonance). Then for this case, the polarizability becomes

$$\alpha(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{1.21}$$
(polarizability with damping)

The susceptibility likewise becomes

$$\chi(\omega) = \frac{Ne^2/m\epsilon_0}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{1.22}$$
(susceptibility with damping)

It is worth reiterating here that α and χ are complex quantities defined for the positive-rotating fields via $\mathbf{d}^{(+)} = \alpha(\omega)\mathbf{E}^{(+)}$ and $\mathbf{P}^{(+)} = \epsilon_0\chi\mathbf{E}^{(+)}$, and therefore must be treated appropriately.

If χ is small (as for a dilute vapor), the complex refractive index is

$$\tilde{n}(\omega) = \sqrt{1+\chi(\omega)} \approx 1 + \frac{\chi(\omega)}{2} = 1 + \frac{Ne^2}{2m\epsilon_0}\frac{(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2} + i\frac{Ne^2}{2m\epsilon_0}\frac{\gamma\omega}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}. \tag{1.23}$$

The real and imaginary parts of the complex index have distinct interpretations. Recall that a plane wave propagates with a phase factor according to

$$E(z) = E_0\exp(ikz) = E_0\exp(i\tilde{n}k_0z) = E_0\exp(i\mathrm{Re}[\tilde{n}]k_0z)\exp(-\mathrm{Im}[\tilde{n}]k_0z), \tag{1.24}$$

and so we can define the **phase index** and **absorption coefficient** respectively as

$$\begin{aligned} n(\omega) &:= \mathrm{Re}[\tilde{n}(\omega)] \\ a(\omega) &:= 2k_0\mathrm{Im}[\tilde{n}(\omega)], \end{aligned} \tag{1.25}$$
(phase index and absorption coefficient)

where k is the wave number in the medium and k_0 is the vacuum wave number, so that $n(\omega)$ represents the phase shift of the propagating wave, while $a(\omega)$ represents attenuation of the field due to absorption. Note that the absorption coefficient is defined such that the intensity I of a wave propagating in the z direction decays according to

$$\frac{dI}{dz} = -aI \quad \Longrightarrow \quad I(z) = I_0e^{-az}, \tag{1.26}$$

which explains the factors of 2 and k_0. We can thus read off the phase index as the real part,

$$n(\omega) \approx 1 + \frac{Ne^2}{2m\epsilon_0}\frac{(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}, \tag{1.27}$$
(classical refractive index)

while the (intensity) absorption coefficient becomes

$$a(\omega) \approx \frac{Ne^2\omega^2}{m\epsilon_0 c}\frac{\gamma}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}.$$
(classical absorption coefficient) (1.28)

Significant absorption occurs in the region of small detunings of the field from the atomic resonance, $|\omega - \omega_0| \ll \omega_0$. Then

$$\omega_0^2 - \omega^2 = (\omega_0-\omega)(\omega_0+\omega) \approx 2\omega(\omega_0-\omega). \tag{1.29}$$

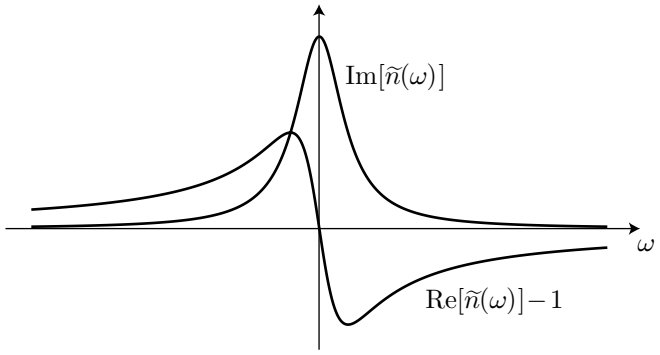
This is effectively equivalent to the **rotating-wave approximation** that we discuss later (Section 5.1.2). With this approximation, the phase index and absorption become

$$\begin{aligned} n(\omega) &\approx 1 + \frac{Ne^2}{2m\epsilon_0}\frac{(\omega_0-\omega)/2\omega}{(\omega_0-\omega)^2+(\gamma/2)^2} \\ a(\omega) &\approx \frac{Ne^2}{m\epsilon_0 c\gamma}\frac{(\gamma/2)^2}{(\omega_0-\omega)^2+(\gamma/2)^2}. \end{aligned} \tag{1.30}$$
(Lorentzian absorption profile)

Thus, we recover the Lorentzian absorption profile (hence the name) with full width at half maximum γ and resonant absorption coefficient $a_0 = a(\omega_0) = Ne^2/m\epsilon_0 c\gamma$. Also, we see that in the same regime,

$$n - 1 = \frac{2(\omega_0 - \omega)}{\gamma}\left[\frac{Ne^2}{2m\epsilon_0\gamma\omega}\frac{(\gamma/2)^2}{(\omega_0-\omega)^2 + (\gamma/2)^2}\right] = \frac{2(\omega_0-\omega)}{\gamma}\mathrm{Im}[\tilde{n}(\omega)], \tag{1.31}$$

as required by the Kramers–Kronig relations (Section 14.1.4.2).



This gives the dispersive form for the phase index, as shown here.

In general, we can have atoms with multiple electrons that we need to sum over. Then the polarizability and susceptibility become

$$\begin{aligned}\alpha(\omega) &= \sum_j \frac{e^2}{m}\frac{f_{0j}}{\left(\omega_{j0}^2 - \omega^2 - i\gamma_j\omega\right)} \\ \chi(\omega) &= \sum_j \frac{Ne^2}{m\epsilon_0}\frac{f_{0j}}{\left(\omega_{j0}^2 - \omega^2 - i\gamma_j\omega\right)}.\end{aligned} \tag{1.32}$$

(corrected response functions)

Here, f_{0j} is the *absorption oscillator strength*, which acts as a weighting factor for each electron, or possibly something like a probability for an electron to behave as different possible harmonic oscillators. The quantum-mechanical (and correct) interpretation of these expressions is that each term in the sum represents a transition from the ground level 0 to excited level j. The oscillator strength can only be obtained from a quantum calculation, and is necessary to make the classical calculation quantitatively correct. Because of the quantitative importance of this factor, we will explore it in more detail.

1.2.1 Oscillator Strength

Since the absorption coefficient scales as the susceptibility and thus the oscillator strength (for a dilute gas), the oscillator strength also scales with the cross section. On resonance, the cross section for absorption is defined by

$$a(\omega_0) = \sigma(\omega_0)N = \sigma_0 N. \tag{1.33}$$

Thus, using Eq. (1.28), we can write the classical absorption cross section as

$$\sigma_{\text{classical}}(\omega_0) = \frac{e^2\omega^2}{m\epsilon_0 c}\frac{\gamma}{(\omega_0^2-\omega^2)^2 + \gamma^2\omega^2}\bigg|_{\omega=\omega_0} = \frac{e^2}{m\epsilon_0 c\gamma}. \tag{1.34}$$

(classical cross section)

This cross section is not quantitatively correct, as the correct quantum-mechanical expression for the cross section for the transition to level j [see Eq. (3.21)] is

$$\sigma_{0j} = \sigma_j(\omega_0) = \frac{\lambda_{j0}^2}{2\pi} = \frac{2\pi c^2}{\omega_{j0}^2}. \tag{1.35}$$

(quantum cross section)

Note that this cross section assumes an orientational average (and thus a factor of 1/3) that is generally appropriate for our purposes. We can then define the absorption oscillator strength to be the "fudge factor" to fix the classical cross section:

$$f_{0j} := \frac{\sigma_{0j}}{\sigma_{0,\,\text{classical}}} = \frac{2\pi\epsilon_0 mc^3\gamma_j}{e^2\omega_{j0}^2}. \tag{1.36}$$
(oscillator strength)

We can also write the cross section as

$$\sigma_{0j} = f_{0j}\frac{e^2}{m\epsilon_0 c\gamma_j}, \tag{1.37}$$

which will be useful later. More commonly, the absorption oscillator strength is defined to include the degeneracy of the level structure,[1]

$$f_{0j} = \frac{2\pi\epsilon_0 mc^3\gamma_j}{e^2\omega_{j0}^2}\frac{g_j}{g_0}, \tag{1.38}$$

where g_α is the degeneracy of level α (i.e., the number of ways to have energy E_α), with a separate expression defined for the *emission* oscillator strength f_{0j} (which just flips the degeneracy ratio).

Also, in the limit of large frequency, the susceptibility of Eqs. (1.32) becomes

$$\chi(\omega) \longrightarrow -\frac{Ne^2}{m\epsilon_0\omega^2}\sum_j f_{0j}. \tag{1.39}$$

In this limit, the induced electron displacements are small, and thus the damping and harmonic-potential forces are not important. We thus expect to recover the behavior of the free-electron plasma in the high-frequency limit (i.e., the conductor without damping). This corresponds to Eq. (1.16) in the limit $\omega_0 \longrightarrow 0$, since the electrons are not bound. Thus, we can write

$$\chi(\omega) = -\frac{Ne^2}{m\epsilon_0\omega^2}. \tag{1.40}$$

Comparing these two expressions for the susceptibility, we find the **Thomas–Reiche–Kuhn sum rule** for the oscillator strength:

$$\sum_j f_{0j} = 1. \tag{1.41}$$
(Thomas–Reiche–Kuhn sum rule)

Since $f_{0j} > 0$, the sum rule tells us that $f_{0j} < 1$. The interpretation is that the classical cross section represents the *maximum possible* cross section, which turns out to be distributed over all the possible transitions from the ground level. Note that transitions to *unbound* (ionized) states are also included in this sum, making it difficult to verify this with atomic transition data.[2]

1.3 Dipole Radiation

The electric and magnetic fields for an oscillating dipole are[3]

$$\mathbf{E}^{(+)}(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0}\left[3(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\left[\frac{d^{(+)}(t_r)}{r^3}+\frac{\dot{d}^{(+)}(t_r)}{cr^2}\right]+\frac{1}{4\pi\epsilon_0}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\frac{\ddot{d}^{(+)}(t_r)}{c^2 r}$$
$$\mathbf{H}^{(+)}(\mathbf{r},t) = \frac{c}{4\pi}(\hat{\varepsilon}\times\hat{r})\left[\frac{\dot{d}^{(+)}(t_r)}{cr^2}+\frac{\ddot{d}^{(+)}(t_r)}{c^2 r}\right], \tag{1.42}$$
(dipole radiation fields)

[1] Alan Corney, *Atomic and Laser Spectroscopy* (Oxford, 1987).

[2] See Peter W. Milonni and Joseph H. Eberly, *Lasers* (Wiley, 1988), p. 239.

[3] See John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), p. 411 or Peter W. Milonni and Joseph H. Eberly, *Lasers* (Wiley, 1988), p. 44.

where $t_\mathrm{r} = t - r/c$ is the retarded time, and $\hat{\varepsilon}$ is the polarization unit vector of the applied field (and thus the dipole orientation vector). Only the $1/r$ terms actually transport energy to infinity (i.e., they correspond to radiation), so we can drop the rest to obtain

$$\begin{aligned}\mathbf{E}^{(+)}(\mathbf{r},t) &\approx \frac{1}{4\pi\epsilon_0 c^2}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\frac{\ddot{d}^{(+)}(t_\mathrm{r})}{r} \\ \mathbf{H}^{(+)}(\mathbf{r},t) &\approx \frac{1}{4\pi c}(\hat{\varepsilon}\times\hat{r})\frac{\ddot{d}^{(+)}(t_\mathrm{r})}{r}.\end{aligned} \tag{1.43}$$

The energy transport is governed by the Poynting vector, which we can write as

$$\begin{aligned}\langle\mathbf{S}\rangle &= \mathbf{E}^{(+)}\times\mathbf{H}^{(-)} + \text{c.c.} \\ &= \frac{1}{16\pi^2\epsilon_0 c^3}\frac{|\ddot{d}^{(+)}|^2}{r^2}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\times(\hat{\varepsilon}^*\times\hat{r}) + \text{c.c.} \\ &= \frac{\hat{r}}{16\pi^2\epsilon_0 c^3}\frac{|\ddot{d}^{(+)}|^2}{r^2}\left(1-|\hat{r}\cdot\hat{\varepsilon}|^2\right) + \text{c.c.},\end{aligned} \tag{1.44}$$

where we have used

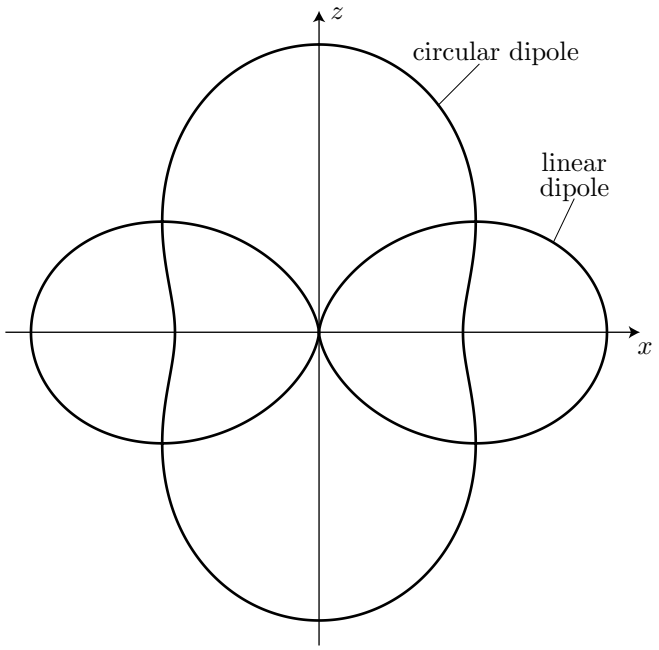
$$\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\times(\hat{\varepsilon}^*\times\hat{r}) = \left(1-|\hat{r}\cdot\hat{\varepsilon}|^2\right)\hat{r} \tag{1.45}$$

for the angular dependence.

There are two main possibilities for the polarization vector: the incident light can be linearly or circularly polarized.

1. **Linear polarization** ($\hat{\varepsilon} = \hat{z}$): $1-|\hat{r}\cdot\hat{\varepsilon}|^2 = \sin^2\theta$. This is the usual ''doughnut-shaped'' radiation pattern for an oscillating dipole.
2. **Circular polarization** ($\hat{\varepsilon} = \hat{\varepsilon}_\pm := \mp(\hat{x}\pm i\hat{y})/\sqrt{2}$): $1-|\hat{r}\cdot\hat{\varepsilon}|^2 = (1+\cos^2\theta)/2$. This is a ''peanut-shaped'' radiation pattern for a rotating dipole.

Here, θ is the angle from the z-axis, while ϕ is the angle around the azimuth. Note that any arbitrary polarization can be represented as a superposition of these three basis vectors. The (intensity/power) radiation patterns for the linear and circular dipole cases are shown here.



The three-dimensional distributions are generated by sweeping these patterns around the z-axis.

The corresponding electric fields for the dipole radiation are polarized. From Eq. (1.43), we can see that the polarization vector is proportional to $(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}$. For linear polarization ($\hat{\varepsilon}=\hat{z}$), this factor turns out to be $\sin\theta\,\hat{\theta}$, while for circular polarization ($\hat{\varepsilon}=\hat{\varepsilon}_{\pm}=\mp(\hat{x}\pm i\hat{y})/\sqrt{2}$), the polarization vector is proportional to $(\cos\theta\,\hat{\theta}\mp i\hat{\phi})e^{\mp i\phi}/\sqrt{2}$.

Now let's define the angular-distribution function via

$$f_{\hat{\varepsilon}}(\theta,\phi) := \frac{3}{8\pi}\left(1-|\hat{r}\cdot\hat{\varepsilon}|^2\right). \tag{1.46}$$
(radiative angular distribution)

For linear and circular polarization, this takes the form

$$\begin{aligned} f_{\hat{z}}(\theta,\phi) &= \frac{3}{8\pi}\sin^2(\theta) \\ f_{\pm}(\theta,\phi) &= \frac{3}{16\pi}\left[1+\cos^2(\theta)\right]. \end{aligned} \tag{1.47}$$

This function has the nice property that it is normalized, and thus represents a probability distribution for photon emission in quantum mechanics:

$$\int f_{\hat{\varepsilon}}(\theta,\phi)\,d\Omega = 1. \tag{1.48}$$

Here, $d\Omega=\sin\theta\,d\theta\,d\phi$ is the usual solid-angle element.

Now we can write the Poynting vector in terms of the angular-distribution function as

$$\langle\mathbf{S}\rangle = \frac{\hat{r}}{3\pi\epsilon_0 c^3}\frac{|\ddot{d}^{(+)}|^2}{r^2}f_{\hat{\varepsilon}}(\theta,\phi). \tag{1.49}$$

The power radiated per unit solid angle is then

$$\frac{dP_{\rm rad}}{d\Omega} = r^2\langle\mathbf{S}\rangle\cdot\hat{r} = \frac{|\ddot{d}^{(+)}|^2}{3\pi\epsilon_0 c^3}f_{\hat{\varepsilon}}(\theta,\phi), \tag{1.50}$$

and the total radiated power is

$$P_{\rm rad} = \int d\Omega\,\frac{dP_{\rm rad}}{d\Omega} = \frac{|\ddot{d}^{(+)}|^2}{3\pi\epsilon_0 c^3} = \frac{e^2|\ddot{x}^{(+)}|^2}{3\pi\epsilon_0 c^3}. \tag{1.51}$$

Of course, the incident intensity is contained implicitly in the electron acceleration $\ddot{x}$.

1.3.1 Damping Coefficient

Now we can connect the radiated power to the damping term in the Lorentz model,[4] Eq. (1.8). Note that the radiated power in Eq. (1.51) is the *time-averaged* power, since we used the complex representation. In terms of the real displacement, we can make the replacement

$$|\ddot{x}^{(+)}|^2 \longrightarrow \frac{\langle\ddot{x}^2\rangle}{2}, \tag{1.52}$$

where the angle brackets denote the time average. Then the average work done by radiation reaction must balance the energy emitted into the field:

$$\begin{aligned} \int_{x_0}^{x}\mathbf{F}_{\rm rr}\cdot d\mathbf{x}' = \int_{t_0}^{t}F_{\rm rr}\,\dot{x}(t')\,dt' &= -\frac{e^2}{6\pi\epsilon_0 c^3}\int_{t_0}^{t}(\ddot{x})^2\,dt' \\ &= -\frac{e^2}{6\pi\epsilon_0 c^3}\left[\dot{x}\ddot{x}\Big|_{t_0}^{t} - \int_{t_0}^{t}\dot{x}\dddot{x}\,dt'\right]. \end{aligned} \tag{1.53}$$

[4] This argument follows Alan Corney, *Atomic and Laser Spectroscopy* (Oxford, 1987), p. 230.

Here $\mathbf{F}_{\rm rr}$ refers to the radiation-reaction force. If we pick $t-t_0$ to be an integer multiple of the optical period, the boundary term vanishes (it is also negligible for large $t-t_0$). Then the radiation-reaction force is

$$\mathbf{F}_{\rm rr} = \frac{e^2}{6\pi\epsilon_0 c^3}\dddot{\mathbf{x}}. \tag{1.54}$$

This is, in fact, the **Abraham–Lorentz model** of radiation reaction, which has problems involving unphysical runaway solutions.[5] We can avoid such problems by going back to the complex representation and noting that the displacement is a harmonic function, so we can make the approximation

$$\mathbf{F}_{\rm rr}^{(+)} = \frac{e^2}{6\pi\epsilon_0 c^3}\dddot{\mathbf{x}}^{(+)} \approx -\frac{e^2\omega_0^2}{6\pi\epsilon_0 c^3}\dot{\mathbf{x}}^{(+)}, \tag{1.55}$$

which assumes the atom is driven close to resonance. If we define

$$\gamma = \frac{e^2\omega_0^2}{6\pi m\epsilon_0 c^3}, \tag{1.56}$$
(classical damping rate)

then we recover the damping term in the harmonic oscillator:

$$\mathbf{F}_{\rm rr}^{(+)} = -m\gamma\dot{\mathbf{x}}^{(+)}. \tag{1.57}$$

Note that the oscillator is highly underdamped here, since Eq. (1.56) can be written as $\gamma/\omega_0 = (4\pi/3)r_{\rm e}/\lambda$, where $r_{\rm e} \approx 2.8\times10^{-15}$ m is the classical electron radius and λ is the optical wavelength. For example, 122 nm is the lowest-lying hydrogen line, which gives $\gamma/\omega_0 \sim 10^{-7}$. For real quantum transitions, this ratio is slightly smaller (due to the addition of the oscillator strength as we mention below), on the order of 10^{-8}.

We now have the classical result for the spontaneous emission rate, which isn't quite correct. Again, we can patch this with the substitution $e^2/m \longrightarrow (e^2/m)f_{0j}$, with the result

$$\gamma_j = \frac{e^2\omega_{0j}^2 f_{0j}}{6\pi m\epsilon_0 c^3}. \tag{1.58}$$
(quantum damping rate)

This is consistent with Eq. (1.36) if we take $\sigma_{0j} = 3\lambda_{j0}^2/2\pi$ (i.e., no orientational average for the dipole). Again, there are some subtleties here regarding the cross sections and orientational averages that are better handled by angular-momentum algebra.

1.4 Atom Optics: Mechanical Effects of Light on Atoms

Now we will have a brief look at the field of **atom optics**, or optics with matter (de Broglie) waves. We will only be looking here at how to trap and cool atoms with laser light using the classical Lorentz model of the atom, so in a sense we will be doing ''geometrical atom optics.''

Broadly speaking, there are two types of mechanical forces that light can have on atoms.[6] The first, the *dipole force*, is related to the potential energy of the induced dipole in the electric field, and is thus related to the real part of $\alpha(\omega)$ [see Eq. (1.66)]. The second is *radiation pressure* due to absorption and rescattering of the incident light, which is thus related to the imaginary part of $\alpha(\omega)$ [see Eq. (1.85)].

[5] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice-Hall, 2013), p. 489, and John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Chapter 16, p. 748.

[6] Strictly speaking, this decomposition into two forces is only true for *scalar* atoms—atoms with no orientation. This is appropriate for the form of the polarizability we have assumed here, but in the case of atoms with nonvanishing vector or tensor polarizabilities, as in Eq. (7.472), other forces associated with polarization gradients arise. See, for example, G. Nienhuis, P. van der Straten, and S-Q. Shang, *Physical Review A* **44**, 462 (1991) (doi: 10.1103/PhysRevA.44.462).

1.4.1 Dipole Force

The dipole moment of the atom is induced by the external field, so we can write the potential energy of the induced dipole (with $\mathbf{d} = \alpha\mathbf{E}$) as

$$V_{\text{dipole}} = -\frac{\mathbf{d}\cdot\mathbf{E}}{2} = -\frac{d\,E}{2}. \tag{1.59}$$

The extra factor of 1/2 compared to the usual dipole energy is because the dipole is induced, and thus

$$V_{\text{dipole}} = -\int_0^E (d)\,dE = -\int_0^E \alpha E\,dE = -\frac{1}{2}\alpha E^2. \tag{1.60}$$

Since we found the solution for the positive-frequency component of the field, we should write out the potential in terms of the same components:

$$V_{\text{dipole}} = -\frac{1}{2}\left(\mathbf{d}^{(+)} + \mathbf{d}^{(-)}\right)\cdot\left(\mathbf{E}^{(+)} + \mathbf{E}^{(-)}\right). \tag{1.61}$$

Noting that

$$\mathbf{d}^{(\pm)} \sim e^{\mp i\omega t}, \qquad \mathbf{E}^{(\pm)} \sim e^{\mp i\omega t}, \tag{1.62}$$

we can see that the terms of the form

$$\mathbf{d}^{(\pm)}\cdot\mathbf{E}^{(\pm)} \sim e^{\mp i2\omega t} \tag{1.63}$$

rotate at twice the optical frequency, which is too fast for the atoms to respond mechanically. So we will drop these terms in the time average (the same average that leads to the intensity). The terms of the form

$$\mathbf{d}^{(\pm)}\cdot\mathbf{E}^{(\mp)} \sim 1 \tag{1.64}$$

are dc, so we can keep these. Thus,

$$\begin{aligned} V_{\text{dipole}} &= -\frac{1}{2}\mathbf{d}^{(+)}\cdot\mathbf{E}^{(-)} - \frac{1}{2}\mathbf{d}^{(-)}\cdot\mathbf{E}^{(+)} \\ &= -\frac{1}{2}\left[\alpha(\omega)\mathbf{E}^{(+)}\right]\cdot\mathbf{E}^{(-)} - \frac{1}{2}\left[\alpha(\omega)\mathbf{E}^{(-)}\right]\cdot\mathbf{E}^{(+)} \\ &= -\text{Re}[\alpha(\omega)]\left|E^{(+)}\right|^2 \\ &= -\frac{\eta_0}{2}\text{Re}[\alpha(\omega)]I(\mathbf{r}). \end{aligned} \tag{1.65}$$

and in terms of the intensity,

$$V_{\text{dipole}} = -\frac{\eta_0}{2}\text{Re}[\alpha(\omega)]I(\mathbf{r}). \tag{1.66}$$
(dipole potential)

Here, η_0 is the **vacuum wave impedance**

$$\eta_0 := \sqrt{\frac{\mu_0}{\epsilon_0}} = \mu_0 c = \frac{1}{\epsilon_0 c} \approx 377\ \Omega, \tag{1.67}$$
(wave impedance of vacuum)

and we are regarding the electric-field envelope $E^{(+)}(\mathbf{r})$ to be a slowly varying function of position. Recall that the intensity in vacuum is given in terms of the real and complex field amplitudes E_0 and $E_0^{(+)}$ by

$$I = |E_0|^2/2\eta_0 = 2|E_0^{(+)}|^2/\eta_0. \tag{1.68}$$
(intensity related to field amplitude)

Putting in the explicit form for the polarizability, we can write the dipole potential as

$$V_{\text{dipole}} = \frac{-e^2}{2m\epsilon_0 c}\,\frac{\omega_0^2 - \omega^2}{(\omega_0^2-\omega^2)^2 + \gamma^2\omega^2}\,I(\mathbf{r}). \tag{1.69}$$

Thus, the atom sees a spatial potential proportional to $I(\mathbf{r})$ and to $(n-1)$. This potential-shift effect (also known as the **ac Stark shift**) is the atomic counterpart of the phase shift (due to $n-1$) of a beam propagating through a vapor. Both effects follow from the coupling of the field to the atom.

The corresponding force is given by the potential gradient

$$\mathbf{F}_{\text{dipole}} = -\nabla V_{\text{dipole}} \propto \nabla I(\mathbf{r}). \tag{1.70}$$

Thus, the dipole force responds to intensity *gradients*. If the dipole is viewed as two slightly separated, opposite charges, there is only a net force if the two charges see a different electric field, which is only possible in the ideal dipole limit if the field has a gradient.

The sign of the dipole potential is set solely by the detuning of the field from the atomic resonance. Defining the detuning $\Delta := \omega - \omega_0$, we can write the dipole potential as

$$V_{\text{dipole}} = \frac{e^2}{2m\epsilon_0 c} \frac{(\omega_0+\omega)\Delta}{[(\omega_0+\omega)\Delta]^2 + \gamma^2\omega^2} I(\mathbf{r}). \tag{1.71}$$

Everything in this expression is positive except for the factor of Δ in the numerator. Thus, for positive Δ ($\omega > \omega_0$, or *blue detuning*), $V_{\text{dipole}} > 0$, while for negative Δ ($\omega < \omega_0$, or *red detuning*), $V_{\text{dipole}} < 0$. That is, a bright spot in space (e.g., due to a tightly focused Gaussian beam) will repel an atom for blue detunings, forming a potential barrier, while for red detunings, the spot attracts atoms and forms a potential well.

The sign dependence of V_{dipole} makes sense in terms of the phase lag (1.20). Recall that for small frequencies ($\Delta < 0$), the phase lag of the dipole behind the field is smaller than $\pi/2$, while for large frequencies ($\Delta > 0$), the phase lag is between $\pi/2$ and π. Since $V_{\text{dipole}} \propto -\mathbf{d}\cdot\mathbf{E}$, the phase lag is important because then $\mathbf{d}$ and $\mathbf{E}$ are mostly aligned or mostly opposed for $\Delta < 0$ and $\Delta > 0$, respectively. Thus, $V_{\text{dipole}} \gtrless 0$ for $\Delta \gtrless 0$.

1.4.1.1 Dipole Potential: Standard Form

By writing the dipole potential in a more standard form, we can see that it matches the result of a quantum calculation, at least in the limit of low intensity. To do this, we first need to patch the classical result of Eq. (1.69) as before by including the oscillator strength and summing over all transitions:

$$V_{\text{dipole}} = -\sum_j \frac{e^2 f_{0j}}{2m\epsilon_0 c} \frac{\omega_{j0}^2 - \omega^2}{(\omega_{j0}^2-\omega^2)^2 + \gamma_j^2\omega^2} I(\mathbf{r}). \tag{1.72}$$
(corrected dipole potential)

Now to put this in more standard form, we need to define the saturation intensity for the atom. When we encounter rate equations in the next chapter, we will see that it is sensible to define an intensity scale known as the **saturation intensity**, given by

$$I_{\text{sat}} := \frac{\hbar\omega_0\gamma}{2\sigma_0}. \tag{1.73}$$
(saturation intensity)

(The damping rate γ here will correspond to the Einstein A coefficient in the rate-equation treatment.) Briefly, the saturation intensity is relevant here in that this classical model is valid—that is, it agrees with quantum predictions—if either the driving intensity is small ($I \ll I_{\text{sat}}$) or the detuning from any resonance is large ($|\omega - \omega_{j0}| \gg \gamma_j$). For the maximum possible resonant cross section of $\sigma_0 = 3\lambda^2/2\pi$ (where there is no average over the dipole orientation), the saturation intensity is $I_{\text{sat}} = 1.10$ mW/cm^2 for ^{133}Cs on the D_2 transition (852 nm), while for ^{87}Rb on the same transition (780 nm), the saturation intensity is $I_{\text{sat}} = 1.67$ mW/cm^2. We can also write the saturation intensity in terms of the oscillator strength by using Eq. (1.37), with the result

$$I_{\text{sat},j} = \frac{\hbar\omega_{0j} m\epsilon_0 c\gamma_j^2}{2e^2 f_{0j}}. \tag{1.74}$$

Even though the above numerical values are often quoted for the saturation intensity, this is actually a context-dependent quantity. A safe but cumbersome approach is to use the quantum-mechanical formalism for angular momentum to directly calculate the cross section and thus saturation intensity.[7]

[7] Daniel A. Steck, "Cesium D Line Data," 2003. Available online at `http://steck.us/alkalidata`.

Using Eq. (1.74), we can write the dipole potential (1.72) as

$$V_{\text{dipole}} = -\sum_j \frac{\hbar\omega_{j0}\gamma_j^2}{4} \frac{\omega_{j0}^2 - \omega^2}{(\omega_{j0}^2 - \omega^2)^2 + \gamma_j^2\omega^2} \frac{I(\mathbf{r})}{I_{\text{sat},j}}.$$

(quantum dipole potential, small intensity) (1.75)

This is the general expression for any frequency, so long as the intensity is small. To simplify this, we can look at the functional form far away from all resonances ($|\omega_{j0} - \omega| \gg \gamma_j$ for all j) so that

$$\begin{aligned} V_{\text{dipole}} &= \sum_j \frac{\hbar\omega_{j0}\gamma_j^2}{4} \frac{1}{\left(\omega^2 - \omega_{j0}^2\right)} \frac{I(\mathbf{r})}{I_{\text{sat},j}} \\ &= \sum_j \frac{\hbar\gamma_j^2}{8} \left(\frac{1}{\omega - \omega_{j0}} - \frac{1}{\omega + \omega_{j0}}\right) \frac{I(\mathbf{r})}{I_{\text{sat},j}}. \end{aligned}$$

(1.76) (far off resonance)

The first term in the parentheses is the inverse of the detuning, and represents the Stark shift due to the atomic resonances. The second term can be interpreted as the weak, additional Stark shift due to resonances at the corresponding *negative frequencies*. This secondary shift is always negative (like a red detuning), and accounts for part of the **Bloch–Siegert shift** (Section 5.9), as well as other effects such as the **Lamb shift** (Section 13.12) and the **Casimir–Polder effect** (Chapters 13-14). Note that this expression also recovers the *dc* Stark shift (or equivalently, the dc polarizability up to some universal factor) when $\omega = 0$, when both terms contribute equal, negative energy shifts.

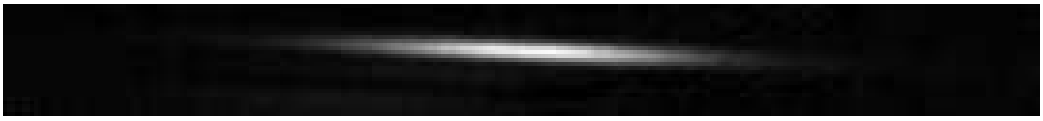
If one resonance is dominant (that is, the laser is tuned far away from resonance, but much closer to one than all the others), then we can make the rotating-wave approximation and neglect the second term in the parentheses of Eq. (1.76) to obtain

$$V_{\text{dipole}} = \frac{\hbar\gamma^2}{8\Delta} \frac{I(\mathbf{r})}{I_{\text{sat}}},$$

(1.77) (far off single dominant resonance)

where again $\Delta = \omega - \omega_0$ is the detuning from resonance. Note that for a far-detuned, linearly polarized laser creating this potential, it turns out that $\sigma_0 = \lambda_0^2/2\pi$ is the appropriate resonant cross section, so the above saturation intensity values should be multiplied by 3 before being used in this formula.

Typically, a focused, red-detuned, Gaussian laser beam is used to make a **dipole trap** or *far-off resonance trap* (FORT)[8] for atoms via the dipole force.[9] Below is an example image of about 10^5 ^{87}Rb atoms confined in a dipole trap formed by a 10 W, 1090 nm Gaussian laser beam (far below the 780 and 794 nm main resonances) focused to a 31 μm beam waist ($1/e^2$ radius), implying a Rayleigh length (depth of focus along the beam direction) of 2.8 mm.



The dipole trap clearly runs from left to right with a slight downward angle; the dimensions of the image are 270 × 29 CCD pixels (6.59 × 0.71 mm). This is an **absorption image**, where the shadow cast by the atoms in a brief, resonant laser probe is imaged and recorded on a CCD camera. (The image greyscale is inverted so the atoms appear bright rather than dark.)

But now the important question to address is under what conditions the trap is *stable*, since as the atom scatters photons, it heats up until it boils out of the trap. So we will need to take a closer look at the *radiation* of the Lorentz atom.

[8] Steven L. Rolston, Christoph Gerz, Kristian Helmerson, P. S. Jessen, Paul D. Lett, William D. Phillips, R. J. Spreeuw, and C. I. Westbrook, "Trapping atoms with optical potentials," *Proceedings of SPIE* **1726**, 205 (1992) (doi: 10.1117/12.130392).

[9] The first observation of atoms trapped in a dipole-force potential was Steven Chu, J. E. Bjorkholm, A. Ashkin, and A. Cable, "Experimental Observation of Optically Trapped Atoms," *Physical Review Letters* **57**, 314 (1986) (doi: 10.1103/PhysRevLett.57.314).

1.4.1.2 Photon Scattering Rate

Now we can compute the rate of photon scattering as a way to get to the rate of momentum transfer. We can write the total radiated power from Eq. (1.51) in terms of the polarizability as

$$P_{\text{rad}} = \frac{\omega^4 |\alpha(\omega)|^2}{6\pi\epsilon_0^2 c^4} I, \tag{1.78}$$
(total radiated power)

where we used $d^{(+)} = \alpha(\omega) E_0^{(+)} e^{-i\omega t}$.

As a brief aside, though, we can write down the scattering cross section from the total radiated power, given the defining relation $P_{\text{rad}} = \sigma I$:

$$\sigma_{\text{Rayleigh}} = \frac{\omega^4 |\alpha(\omega)|^2}{6\pi\epsilon_0^2 c^4}. \tag{1.79}$$
(Rayleigh scattering cross section)

The overall scaling is as ω^4 (neglecting the small modification due to the polarizability). This is the usual explanation for why the sky is blue and sunsets are red: blue wavelengths are preferentially scattered by the atmosphere, while red wavelengths are preferentially transmitted.

We can continue by writing out explicitly the polarizability in Eq. (1.78), using Eq. (1.32):

$$P_{\text{rad}} = \frac{e^2\omega^4}{6\pi m^2 \epsilon_0^2 c^2} \left| \sum_j \frac{f_{0j}}{\omega_{j0}^2 - \omega^2 - i\gamma_j\omega} \right|^2 I. \tag{1.80}$$

Using Eq. (1.58) to eliminate the oscillator strengths,

$$\begin{aligned} P_{\text{rad}} &= 6\pi c^2 \left| \sum_j \frac{\omega^2}{\omega_{j0}^2} \frac{\gamma_j}{\omega_{j0}^2 - \omega^2 - i\gamma_j\omega} \right|^2 I \\ &= \frac{\hbar}{2} \left| \sum_j \frac{\omega^2}{\sqrt{\omega_{j0}}} \frac{\gamma_j^{3/2}}{\omega_{j0}^2 - \omega^2 - i\gamma_j\omega} \sqrt{\frac{I}{I_{\text{sat},j}}} \right|^2, \end{aligned} \tag{1.81}$$

where we used $\sigma_{0j} = 3\lambda_0^2/2\pi$ to write the saturation intensity as

$$I_{\text{sat},j} = \frac{\hbar\omega_{j0}\gamma_j}{2\sigma_{0j}} = \frac{\hbar\omega_{j0}^3\gamma_j}{4\pi c^2}. \tag{1.82}$$

The photon scattering rate R_{sc} is the radiated power divided by the photon energy $\hbar\omega$:

$$R_{\text{sc}} = \frac{P_{\text{rad}}}{\hbar\omega} = \left| \sum_j \frac{\omega^{3/2}}{\sqrt{2\omega_{j0}}} \frac{\gamma_j^{3/2}}{\omega_{j0}^2 - \omega^2 - i\gamma_j\omega} \sqrt{\frac{I}{I_{\text{sat},j}}} \right|^2. \tag{1.83}$$
(photon scattering rate)

Again, this expression simplifies greatly for certain detunings. Near one resonance, we can ignore the contribution of the others:

$$\begin{aligned} R_{\text{sc}} &\approx \frac{\omega^3}{2\omega_0} \frac{\gamma^3}{|\omega_0^2 - \omega^2 - i\gamma\omega|^2} \frac{I}{I_{\text{sat}}} \\ &= \frac{\omega^3}{2\omega_0} \frac{\gamma^3}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2} \frac{I}{I_{\text{sat}}} \end{aligned} \tag{1.84}$$

Using Eq. (1.32) restricted to a single resonance, we find

$$R_{\text{sc}} = \frac{\eta_0}{\hbar} \frac{\omega^2}{\omega_0^2} \text{Im}[\alpha] I(\mathbf{r}), \tag{1.85}$$
(single dominant resonance)

which shows the connection of the scattering rate (and hence the radiation pressure force below) to the absorptive part of the polarizability. But as we have seen, this is only true near a *single* resonance.

1.4.1.3 Optical Theorem

Actually, what we just said isn't *quite* true: the scattering rate *can* be generally written in terms of the imaginary part of the polarizability. The problem is actually with the Lorentz model, when we tried to mock it up to model multiple transitions. To see this, we start with the susceptibility expression (1.17), $\chi(\omega) = (N/\epsilon_0)\alpha(\omega)$. Now for a rarefied vapor, we can use Eq. (1.23) for the complex refractive index to write

$$\tilde{n}(\omega) \approx 1 + \frac{\chi(\omega)}{2} = 1 + \frac{N}{2\epsilon_0}\alpha(\omega). \tag{1.86}$$

Then the absorption coefficient from (1.28) for the vapor is

$$a(\omega) = \frac{2\omega}{c}\mathrm{Im}[\tilde{n}(\omega)] = \frac{N\omega}{\epsilon_0 c}\mathrm{Im}[\alpha(\omega)]. \tag{1.87}$$

The power *absorbed* from an incident, monochromatic field of intensity $I(\omega)$ by the atomic vapor is given by $P_{\mathrm{abs}}(\omega) = \sigma(\omega)I(\omega)$ in terms of the cross-section σ, which is related to the absorption coefficient by $a(\omega) = \sigma(\omega)N$. (Note how the units work out in these two expressions.) Putting these together, the power absorbed is

$$P_{\mathrm{abs}}(\omega) = \frac{a(\omega)I(\omega)}{N} = \frac{\omega}{\epsilon_0 c}\mathrm{Im}[\alpha(\omega)]\, I(\omega). \tag{1.88}$$

For energy to be conserved, the power emitted from Eq. (1.78) must match this expression for the power absorbed. Equating these, we arrive at the **optical theorem**

$$\mathrm{Im}[\alpha(\omega)] = \frac{1}{4\pi\epsilon_0}\frac{2\omega^3}{3c^3}|\alpha(\omega)|^2. \tag{1.89}$$
(optical theorem)

Of course, this statement of energy conservation in principle only applies in steady state (since we neglect energy stored in the atom), but we have already implied we are considering steady-state behavior just by writing down the polarizability. Thus, the result above from Eq. (1.85),

$$R_{\mathrm{sc}} = \frac{\eta_0}{\hbar}\frac{\omega^2}{\omega_0^2}\mathrm{Im}[\alpha]I(\mathbf{r}), \tag{1.90}$$
(general scattering rate)

is generally true, and it emphasizes that the *imaginary* part of the polarizability represents absorption or loss, and equivalently scattering. We will return to this statement later in Section 14.1.4.1 and see that the imaginary parts of generalized response functions *always* represent loss.

However, as we have already noted, the Lorentz polarizability for multiple transitions ("electrons") does not satisfy the optical theorem.[10] We can see this by looking at Eq. (1.83), where the terms in the sum [representing the same sum in $\alpha(\omega)$] are combined and then squared, so that there are cross terms involving different resonances. But when considering the equivalent expression in terms of $\mathrm{Im}[\alpha(\omega)]$, this scattering rate involves a sum over terms of the form of Eq. (1.84), which always involve a single resonance. There are no cross terms in this case. In the first case, the cross terms reflect the fact that the fields radiated on each transition can *interfere*, which is why field amplitudes are added and then squared in Eq. (1.83). This is apparently not captured in the imaginary part of $\alpha(\omega)$, which indicates a defect in our polarizability expression. This problem is somewhat academic, however, since optical atomic resonances have separations much larger than their widths, and the difference between the two expressions is usually negligible. For practical purposes, the Lorentz polarizability (1.32) is just fine.

[10] For a discussion of quantum-mechanical polarizabilities and their compatibility with the optical theorem, see Paul R. Berman, Robert W. Boyd, and Peter W. Milonni, "Polarizability and the optical theorem for a two-level atom with radiative broadening," *Physical Review A* **74**, 053816 (2006) (doi: 10.1103/PhysRevA.74.053816).

1.4.1.4 Scaling

Far away from the dominant resonance ($|\Delta| \gg \gamma$), but still close enough for the resonance to still dominate, we find that

$$R_{\rm sc} \approx \frac{\gamma^3}{8\Delta^2}\frac{I}{I_{\rm sat}} = \frac{\gamma}{\hbar\Delta}V_{\rm dipole}, \qquad \text{(far off single dominant resonance)} \tag{1.91}$$

where in the last formula we have used Eq. (1.77) for the dipole potential. This is a fundamentally important scaling law for making productive use of dipole forces, as we will now discuss.

The result (1.91) for $R_{\rm sc}$, along with (1.77) for $V_{\rm dipole}$, are of prime importance in the design of an optical dipole trap. The photon scattering rate represents *heating* of the atoms, as we will discuss, because of the random nature of photon emission. But the scattering rate and dipole potential scale as

$$R_{\rm sc} \propto \frac{I}{\Delta^2};\;\; V_{\rm dipole} \propto \frac{I}{\Delta}. \qquad \text{(far off single dominant resonance)} \tag{1.92}$$

These scaling laws are actually interrelated by a nice qualitative argument. From the scaling of $R_{\rm sc}$, we conclude that the radiated *field* is proportional to $\sqrt{I}/\Delta$. Now recall that the dipole potential scales as the phase refractive index $n-1$, which at a microscopic level arises due to the interference of the radiated field with the *forward* field (thus changing the phase velocity of the transmitted field), whose amplitude is $\sqrt{I}$. Hence $n-1$ and thus the dipole potential scale as $(\sqrt{I}/\Delta)\sqrt{I}$, or as I/Δ.

So, for a given desired potential depth, the scattering (heating) rate can be made very small by making the detuning Δ large. The resulting decrease in trap depth is compensated by increasing the intensity. Thus, dipole traps with small heating rates and hence long lifetimes (up to minutes for dipole traps created by CO_2 laser light) can be created in this way. For example, a CO_2-laser dipole trap has been realized for Cs atoms,[11] where a beam waist of 110 μm with 25 W of power gives a trap 120 μK deep, deep enough to confine laser-cooled atoms. The intensities here are incredibly high compared to the saturation intensity (1 mW/cm^2) because the laser is so far off resonance ($\lambda = 10.6$ μm vs. a resonance wavelength $\lambda_0 = 852$ nm). In this regime the above scaling laws are invalid, and in fact the laser field can be treated as static to good approximation.

Note that for a linearly polarized, far-detuned dipole trap where these scalings are valid, the same saturation intensities are to be used to calculate the dipole force and scattering rate, as discussed above.

1.4.2 Radiation Pressure

Now we will examine the forces due to absorption and reemission of the incident light. Each photon carries a momentum of $\hbar k$. Thus, the photon scattering rate (1.84) implies a rate of momentum transfer and thus a force due to *radiation pressure* of

$$\mathbf{F}_{\rm rad} = \hbar\mathbf{k}R_{\rm sc}, \tag{1.93}$$

Note that even though we have invoked the concept of the photon here, which will be convenient when discussing the heating rate, everything here is really classical, since the classical field momentum is related to the absorbed beam power: $F = dp/dt = P_{\rm abs}/c = \sigma I/c$.

To get a sense of scale of the momenta involved, we can compute the **recoil velocity** $v_{\rm r}$, defined as the velocity corresponding to one photon recoil momentum $\hbar k$:

$$v_{\rm r} = \frac{\hbar k}{m}. \tag{1.94}$$

For ^{133}Cs at 852 nm, $v_{\rm r} = 3.5$ mm/s, and for ^{87}Rb at 780 nm, $v_{\rm r} = 5.9$ mm/s, so the recoil velocity is orders of magnitude smaller than typical room-temperature velocities.

Close to a single resonance, the scattering rate from Eq. (1.84) is

$$R_{\rm sc} = \frac{(\gamma/2)^3}{\Delta^2 + (\gamma/2)^2}\,\frac{I}{I_{\rm sat}}, \tag{1.95}$$

[11] H. Engler, T. Weber, M. Mudrich, R. Grimm, and M. Weidemüller, "Very long storage times and evaporative cooling of cesium atoms in a quasielectrostatic dipole trap," *Physical Review A* **62**, 031402 (2000) (doi: 10.1103/PhysRevA.62.031402).

so that the force due to radiation pressure becomes

$$\mathbf{F}_{\text{rad}} = \frac{\hbar \mathbf{k}_0 (\gamma/2)^3}{\Delta^2 + (\gamma/2)^2} \frac{I}{I_{\text{sat}}}. \tag{1.96}$$
(radiation pressure)

Again, depending on the polarization and exactly how close the detuning is (i.e., whether or not the hyperfine structure is resolved), the appropriate value of the saturation intensity might be very different, so some caution is necessary in applying these formulae.

1.4.3 Laser Cooling: Optical Molasses

Now let's explore how we can use the radiation-pressure force to cool atoms. The simplest setup we can consider is an atom moving with velocity $\mathbf{v}$, exposed to identical but counterpropagating laser fields along the velocity direction.



The radiation-pressure force on the atom due to the two fields from Eq. (1.96) is

$$F_{\text{rad}} = \hbar k (\gamma/2)^3 \left(\frac{1}{\Delta_1^2 + (\gamma/2)^2} - \frac{1}{\Delta_2^2 + (\gamma/2)^2} \right) \frac{I}{I_{\text{sat}}}, \tag{1.97}$$

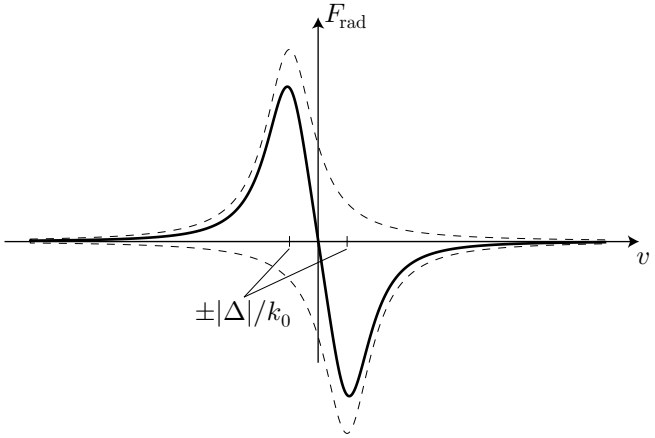
where $\Delta_{1,2}$ are the effective detunings of the two lasers. The detunings of the two lasers are the same in the laboratory frame, but the idea behind Doppler cooling is to tune the lasers *below* the atomic resonance, so that the beam that opposes the atomic velocity is Doppler-shifted into resonance, thus tending to stop the atom. With the pictured setup, the frequency of laser 1 is Doppler shifted (red shifted) by $-kv$, while the frequency of laser 2 is Doppler shifted (blue shifted) by $+kv$. Since the detunings are given by $\Delta_{1,2} = \omega_{1,2} - \omega_0$, we can write

$$\begin{aligned} \Delta_1 &= \Delta - kv \\ \Delta_2 &= \Delta + kv, \end{aligned} \tag{1.98}$$

where $\Delta = \omega - \omega_0$ is the detuning in the laboratory frame. Then the force is

$$F_{\text{rad}} = \hbar k (\gamma/2)^3 \left(\frac{1}{(\Delta - kv)^2 + (\gamma/2)^2} - \frac{1}{(\Delta + kv)^2 + (\gamma/2)^2} \right) \frac{I}{I_{\text{sat}}}. \tag{1.99}$$

Regarded as a function of velocity, this expression is the difference of two Lorentzians, each displaced by $|\Delta|/k$ from zero velocity. This force is plotted for $\Delta = -\gamma/2$, and the two offset Lorentzians are shown as dashed lines.



For small velocity [$v \ll \max(|\Delta|, \gamma)/k$], we can expand to lowest order in v to obtain the viscous damping (''friction'') force:

$$F_{\text{rad}} = \frac{\hbar k^2 \gamma^3}{2} \frac{\Delta}{[\Delta^2 + (\gamma/2)^2]^2} \frac{I}{I_{\text{sat}}} v. \qquad (1.100)$$
(optical molasses, small v)

Because this force is damping for $\Delta < 0$, typically leading to heavily overdamped motion for trapped atoms, this light configuration is called **optical molasses**.[12] The maximum damping rate occurs for $\Delta = -\gamma/2\sqrt{3}$, although it turns out that the optimal detuning is actually something else for reasons we will soon discuss.

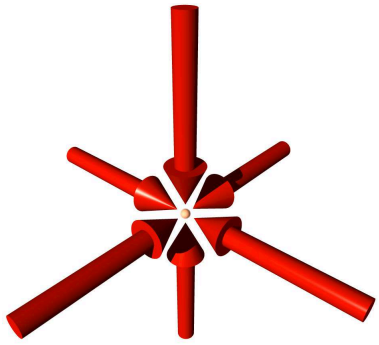
The velocity capture range is the range in velocity for which the force is appreciable. Thus, the capture range is on the order of

$$\pm \frac{|\Delta|}{k} \sim \pm \frac{\gamma}{2k} = \pm \frac{\gamma \lambda}{4\pi}, \qquad (1.101)$$
(capture velocity range)

assuming $\Delta \sim -\gamma/2$. For both ^{133}Cs ($\gamma = 32.8 \times 10^6$ s^{-1}, $\lambda_0 = 852$ nm) and ^{87}Rb ($\gamma = 38.1 \times 10^6$ s^{-1}, $\lambda_0 = 780$ nm), the capture range is about ± 2 m/s. Thus, only fairly slowly moving atoms can be cooled at all with this method. Traditionally to load atomic traps, atoms were slowed by other methods from hot atomic beams to below the capture velocity and then trapped. However, it is possible to load a trap from room-temperature vapor with this method by capturing only the small fraction of the atoms with small enough velocity.

1.4.3.1 Doppler Cooling Limit

For laser cooling in three dimensions, it is sufficient to simply combine three of the above one-dimensional setups, one along each axis.[13]



Then we can write the force vector for small velocities as,

$$\mathbf{F}_{\text{rad}} = \frac{\hbar k^2 \gamma^3}{2} \frac{\Delta}{[\Delta^2 + (\gamma/2)^2]^2} \frac{I}{I_{\text{sat}}} \mathbf{v}. \qquad (1.102)$$

where I is still the intensity of a *single* beam.

Our treatment so far makes it appear as though the atomic velocity may be damped completely away. However, we have only considered the *average* cooling force. There are also *fluctuations* of the cooling force that lead to a temperature limit. We will now derive this temperature limit, the **Doppler limit**, for the cooling mechanism presented here.

Let's look at the variance of the velocity distribution:

$$\frac{d}{dt} \langle v^2 \rangle = 2 \left\langle \mathbf{v} \cdot \frac{d\mathbf{v}}{dt} \right\rangle = \frac{2}{m_{\text{A}}} \left\langle \mathbf{v} \cdot \frac{d\mathbf{p}}{dt} \right\rangle = \frac{2}{m_{\text{A}}} \langle \mathbf{v} \cdot \mathbf{F}_{\text{rad}} \rangle . \qquad (1.103)$$

[12] Steven Chu, J. E. Bjorkholm, A. Ashkin, and A. Cable, ''Experimental Observation of Optically Trapped Atoms,'' *Physical Review Letters* **57**, 314 (1986) (doi: 10.1103/PhysRevLett.57.314).

[13] Graphics by Windell Oskay.

Here, $m_{\rm A}$ is the atomic mass, and the angle brackets denote an ensemble average. With the small-velocity expression (1.100) for the average cooling force, this equation of motion becomes

$$\frac{d}{dt}\langle v^2\rangle = \frac{\hbar k^2\gamma^3}{m_{\rm A}}\frac{\Delta}{[\Delta^2+(\gamma/2)^2]^2}\,\frac{I}{I_{\rm sat}}\,\langle v^2\rangle. \tag{1.104}$$

Again, according to this differential equation, the velocity damps to zero for $\Delta < 0$.

Now we will include the force fluctuations heuristically, since the fluctuations are quantum–mechanical in origin (although there is a more general connection between damping and fluctuations known as the **fluctuation–dissipation relation**; see Problem 5.26 and Section 14.3.8.1). In the course of scattering a photon from one of the laser beams, there is a photon absorption and a photon emission. Each absorption leads to a momentum "kick" of magnitude $\hbar k$, and the direction is random but along one of the six beams. The emission is also in a random direction (not in a dipole-radiation pattern if we assume all polarizations to be equally present), leading to a second kick of magnitude $\hbar k$ in a random direction. Thus, a scattering event is effectively equivalent to two steps in a random walk in velocity space, where the step size is $\hbar k/m_{\rm A}$. These scattering events happen at the scattering rate

$$R_{\rm sc} = \frac{(\gamma/2)^3}{\Delta^2+(\gamma/2)^2}\,\frac{6I}{I_{\rm sat}}, \tag{1.105}$$

since there are six beams present. Recall that for a random walk (see Section 17.1, p. 697), each step increases $\langle v^2\rangle$ by a fixed amount, given by the variance after one step starting from the origin. The three-dimensional probability distribution for a single scattering event is confined to a shell of radius $\hbar k/m_{\rm A}$ in velocity space for either the absorption or emission event. The probability distribution is also inversion symmetric in either case. Thus if the atom is initially at rest, then after one step in the random walk, we can write

$$\langle v^2_{\rm initial}\rangle = 0 \quad\longrightarrow\quad \langle v^2_{\rm final}\rangle = \left(\frac{\hbar k}{m_{\rm A}}\right)^2, \tag{1.106}$$

so that $\langle v^2\rangle$ increases at the rate

$$2R_{\rm sc}\left(\frac{\hbar k}{m_{\rm A}}\right)^2. \tag{1.107}$$

Including the heating rate in Eq. (1.104), we find

$$\frac{d}{dt}\langle v^2\rangle = \frac{\hbar k^2\gamma^3}{m_{\rm A}}\frac{\Delta}{[\Delta^2+(\gamma/2)^2]^2}\,\frac{I}{I_{\rm sat}}\,\langle v^2\rangle + \frac{3\gamma^3}{2}\frac{1}{\Delta^2+(\gamma/2)^2}\,\frac{I}{I_{\rm sat}}\left(\frac{\hbar k}{m_{\rm A}}\right)^2. \tag{1.108}$$

In steady state, we can set the right-hand side to zero, with the result

$$\langle v^2\rangle = \frac{3\hbar\gamma}{4m_{\rm A}}\frac{1+(2\Delta/\gamma)^2}{(-2\Delta/\gamma)}. \tag{1.109}$$

This is an expression for the equilibrium kinetic energy, which we can convert to a temperature via

$$\frac{1}{2}m_{\rm A}\langle v^2\rangle = \frac{3}{2}k_{\rm B}T, \tag{1.110}$$

where $k_{\rm B}$ is the Boltzmann constant. This gives

$$k_{\rm B}T = \frac{\hbar\gamma}{4}\frac{1+(2\Delta/\gamma)^2}{(-2\Delta/\gamma)}. \tag{1.111}$$

The temperature is minimized for the detuning $\Delta = -\gamma/2$, giving the **Doppler temperature** $T_{\rm D}$:

$$k_{\rm B}T_{\rm D} = \frac{\hbar\gamma}{2}. \tag{1.112}$$
(Doppler limit)

This temperature is the best expected for Doppler cooling. For ^{133}Cs at 852 nm, $T_{\rm D} = 125\ \mu$K, and for ^{87}Rb at 780 nm, $T_{\rm D} = 146\ \mu$K. These temperatures are extremely low. We can compare these temperatures to the **recoil temperature** $T_{\rm r}$, which is the temperature corresponding to atoms with an average momentum of one photon recoil $\hbar k$ (i.e., a one-dimensional rms momentum of one photon recoil):

$$k_{\rm B} T_{\rm r} = \frac{(\hbar k)^2}{m_{\rm A}}. \qquad (1.113)$$
(recoil temperature)

For ^{133}Cs, $T_{\rm r} = 198$ nK, and for ^{87}Rb, $T_{\rm r} = 362$ nK, so the Doppler limit is $T_{\rm D} = 631\,T_{\rm r}$ for ^{133}Cs and $T_{\rm D} = 403\,T_{\rm r}$ for ^{87}Rb. Since the (one-dimensional) rms velocity is

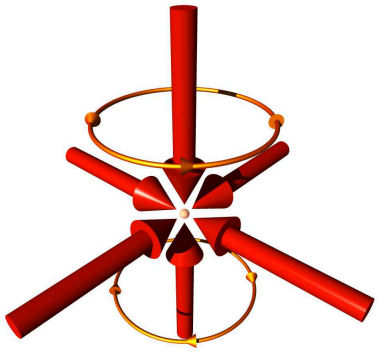
$$v_{\rm rms} = \sqrt{\frac{T_{\rm D}}{T_{\rm r}}}\left(\frac{\hbar k}{m_{\rm A}}\right), \qquad (1.114)$$

which is 8.8 cm/s for ^{133}Cs and 12 cm/s for ^{87}Rb. These velocities are about three orders of magnitude slower than room-temperature rms velocities.

It turns out that for alkali vapors, typical laser-cooled samples exhibit temperatures well below the Doppler limit. Such "sub-Doppler" cooling is due to the degenerate level structure of alkali atoms.[14] For example, ^{133}Cs can be laser cooled with the same general setup described above to about 2.5 μK.[15]

1.4.3.2 Magneto-Optical Trap

Optical molasses tends to stop atoms, making them "stuck," but it does not confine atoms to a particular place. A slight modification to the three-dimensional optical molasses is to impose the magnetic field due to two opposed current loops in the "anti-Helmholtz" configuration. This arrangement is called the **magneto-optical trap** (MOT).[16]



The magnetic field vanishes at the center point of the trap, thus defining a point for atoms to accumulate. We will not go into the operation of the trap in detail yet, but essentially the idea is very similar to laser cooling. The additional complication is that the laser beams must all be correctly (circularly) polarized to address magnetic substates in the degenerate excited level. The magnetic field gives a position-dependent "Zeeman" shift of the transition frequency, such that if the atom is away from the center of the trap, the appropriate beam comes into resonance and pushes it towards the trap center.[17]

[14] Sub-Doppler temperature were observed in some of the first laser cooling experiments. The first reported observation is P. D. Lett, W. D. Phillips, S. L. Rolston, C. E. Tanner, R. N. Watts, and C. I. Westbrook, "Optical molasses," *Journal of the Optical Society of America B* **6**, 2084 (1989). A classic treatment of sub-Doppler cooling mechanisms is J. Dalibard and C. Cohen-Tannoudji, "Laser cooling below the Doppler limit by polarization gradients: simple theoretical models," *Journal of the Optical Society of America B* **6**, 2023 (1989).

[15] C. Salomon, J. Dalibard, W. D. Phillips, A. Clairon, and S. Guellati, "Laser Cooling of Cesium Atoms below 3 μK," *Europhysics Letters* **12**, 683 (1990).

[16] Graphics by Windell Oskay.

[17] Jean Dalibard proposed the idea for the magneto-optical trap, and the MOT was first demonstrated by E. L. Raab, M. Prentiss, Alex Cable, Steven Chu, and D. E. Pritchard, "Trapping of Neutral Sodium Atoms with Radiation Pressure," *Physical Review Letters* **59**, 2631 (1987) (doi: 10.1103/PhysRevLett.59.2631).

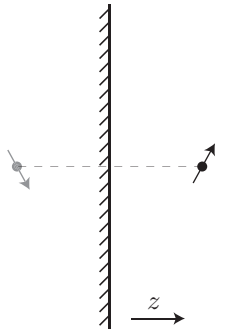
1.5 Cooperative Radiation

We will close our classical treatment with ''cooperative effects,'' where the radiation of an atom is influenced by other atoms. The two examples here serve as prototypes for the quantum-mechanical problems and to show how far one can get with classical arguments, as in the atom optics presentation above.

1.5.1 Atom–Mirror Interaction

The first effect we consider is the influence of a perfectly conducting plane on a radiating dipole.[18] This is a classical prototype for both cavity QED and the Casimir–Polder effect, which we will study much later.

The setup is a dipole located a distance z from a mirror. The field due to the boundary is equivalent to that of an image dipole at position $-z$; due to the relative locations of the constituent charges, the image dipole is reflected in the direction transverse to the plane, but the orientation is the same in the orthogonal direction.



For a dipole oscillating at frequency ω, the field amplitude is

$$\mathbf{E}^{(+)}(\mathbf{r},\omega) = \frac{1}{4\pi\epsilon_0}\left[3(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\left[\frac{1}{r^3}-i\frac{k}{r^2}\right]d^{(+)}(\omega)e^{ikr} - \frac{1}{4\pi\epsilon_0}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}\right]\frac{k^2}{r}d^{(+)}(\omega)e^{ikr}, \quad \text{(monochromatic dipole field)} \tag{1.115}$$

with $\omega = ck$. If we regard the image as the radiating dipole, we are interested in the field at the position of the atom, so that $r = 2z$ and $\hat{r} = \hat{z}$. Also, it is useful to consider separately the cases of a dipole parallel to the surface,

$$\begin{aligned}\left[3(\hat{\varepsilon}_\parallel\cdot\hat{r})\hat{r}-\hat{\varepsilon}_\parallel\right] &= -\hat{\varepsilon}_\parallel \\ \left[(\hat{\varepsilon}_\parallel\cdot\hat{r})\hat{r}-\hat{\varepsilon}_\parallel\right] &= -\hat{\varepsilon}_\parallel,\end{aligned} \tag{1.116}$$

and a dipole perpendicular to the surface,

$$\begin{aligned}\left[3(\hat{\varepsilon}_\perp\cdot\hat{r})\hat{r}-\hat{\varepsilon}_\perp\right] &= 2\hat{\varepsilon}_\perp = 2\hat{z} \\ \left[(\hat{\varepsilon}_\perp\cdot\hat{r})\hat{r}-\hat{\varepsilon}_\perp\right] &= 0,\end{aligned} \tag{1.117}$$

where recall that $\hat{\varepsilon}$, the polarization vector of the applied field, is also the unit vector representing the dipole direction. Since we are concerned with the image and not the original dipole, we make the replacements $\hat{\varepsilon}_\parallel \longrightarrow -\hat{\varepsilon}_\parallel$ and $\hat{\varepsilon}_\perp \longrightarrow \hat{\varepsilon}_\perp$, as we can see from the above diagram. Thus, the field due to the surface at the atom is

$$\begin{aligned}\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega) &= \frac{k^3}{4\pi\epsilon_0}\left[\frac{1}{(2kz)^3}-i\frac{1}{(2kz)^2}\right]\left[2\mathbf{d}^{(+)}_\perp(\omega)+\mathbf{d}^{(+)}_\parallel(\omega)\right]e^{i2kz} - \frac{k^3}{4\pi\epsilon_0}\frac{1}{2kz}\mathbf{d}^{(+)}_\parallel(\omega)e^{i2kz} \\ &= -\frac{3}{2}\gamma\frac{k^3}{k_0^2}\frac{mc}{e}\left\{\left[\frac{1}{z'^3}-i\frac{1}{z'^2}\right]\left[2\mathbf{x}^{(+)}_\perp(\omega)+\mathbf{x}^{(+)}_\parallel(\omega)\right]-\frac{1}{z'}\mathbf{x}^{(+)}_\parallel(\omega)\right\}e^{iz'},\end{aligned} \tag{1.118}$$

[18] See, e.g., H. Morawitz, ''Self-Coupling of a Two-Level System by a Mirror,'' *Physical Review* **187**, 1792 (1969) (doi: 10.1103/PhysRev.187.1792).

where recall that $\mathbf{d} = -e\mathbf{x}$,

$$z' := 2kz, \tag{1.119}$$

$k_0 = \omega_0/c$, and we have used the classical formula (1.56) for the damping rate γ.

Now we put this field back into the the Lorentz model (1.18) as the driving field:

$$\ddot{\mathbf{x}}^{(+)} + \gamma\dot{\mathbf{x}}^{(+)} + \omega_0^2\mathbf{x}^{(+)} = -\frac{e}{m}\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega)e^{-i\omega t}. \tag{1.120}$$

We consider no other driving field, since we are simply interested in how the atom damps to equilibrium. Assuming again a solution of the form

$$\mathbf{x}^{(+)}(t) = \hat{\varepsilon}x_0^{(+)}e^{-i\omega t}, \tag{1.121}$$

we have

$$\omega_0^2 - \omega^2 = i\omega\gamma - \frac{e\hat{\varepsilon}\cdot\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega_0)}{mx_0^{(+)}}. \tag{1.122}$$

Assuming a small perturbation ($|\omega - \omega_0| \ll \omega_0$), so that $\omega_0^2 - \omega^2 = (\omega_0 + \omega)(\omega_0 - \omega) \approx 2\omega_0(\omega_0 - \omega)$ (and we can write ω_0 for ω when it is isolated), we find

$$\omega = \omega_0 - i\frac{\gamma}{2} + \frac{e\hat{\varepsilon}\cdot\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega_0)}{2m\omega_0 x_0^{(+)}}. \tag{1.123}$$

Since ω is the rotation frequency of the dipole, the *real* part corresponds to the actual rotation (energy), while the *imaginary* part corresponds to damping. We can thus see that the mirror field induces shifts $\delta\omega_0$ and $\delta\gamma$ in the oscillator frequency and damping rate, respectively, according to

$$\omega = (\omega_0 + \delta\omega_0) - i\frac{(\gamma + \delta\gamma)}{2}, \tag{1.124}$$

so that

$$\begin{aligned}\delta\omega_0 &= \text{Re}\left[\frac{e\hat{\varepsilon}\cdot\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega)}{2m\omega_0 x_0^{(+)}}\right] \\ \delta\gamma &= -\text{Im}\left[\frac{e\hat{\varepsilon}\cdot\mathbf{E}^{(+)}_{\text{mirror}}(z,\omega)}{m\omega_0 x_0^{(+)}}\right].\end{aligned} \tag{1.125}$$

Evaluating these expressions for a perpendicular dipole, we find

$$\begin{aligned}\delta\omega_{0\perp} &= -\frac{3}{2}\gamma\left(\frac{\sin z'}{z'^2} + \frac{\cos z'}{z'^3}\right) \\ \delta\gamma_\perp &= -3\gamma\left(\frac{\cos z'}{z'^2} - \frac{\sin z'}{z'^3}\right).\end{aligned} \tag{1.126}$$

(dipole-mirror shifts)

For the parallel dipole,

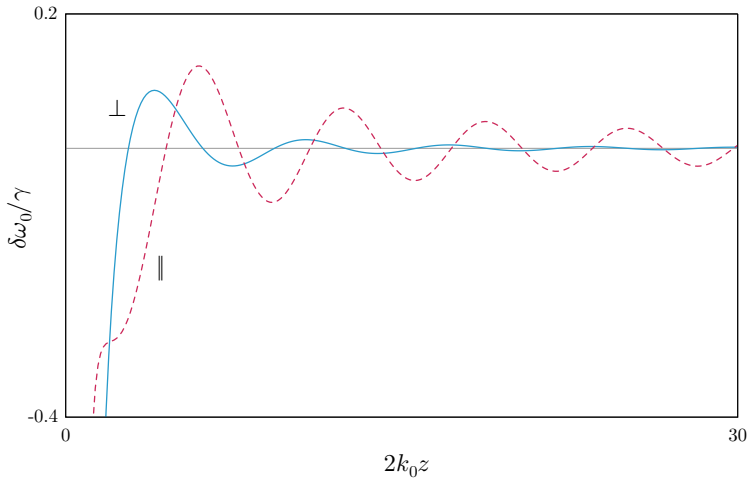
$$\begin{aligned}\delta\omega_{0\parallel} &= \frac{3}{4}\gamma\left[\left(\frac{1}{z'} - \frac{1}{z'^3}\right)\cos z' - \frac{\sin z'}{z'^2}\right] \\ \delta\gamma_\parallel &= -\frac{3}{2}\gamma\left[\left(\frac{1}{z'} - \frac{1}{z'^3}\right)\sin z' + \frac{\cos z'}{z'^2}\right].\end{aligned} \tag{1.127}$$

(dipole-mirror shifts)

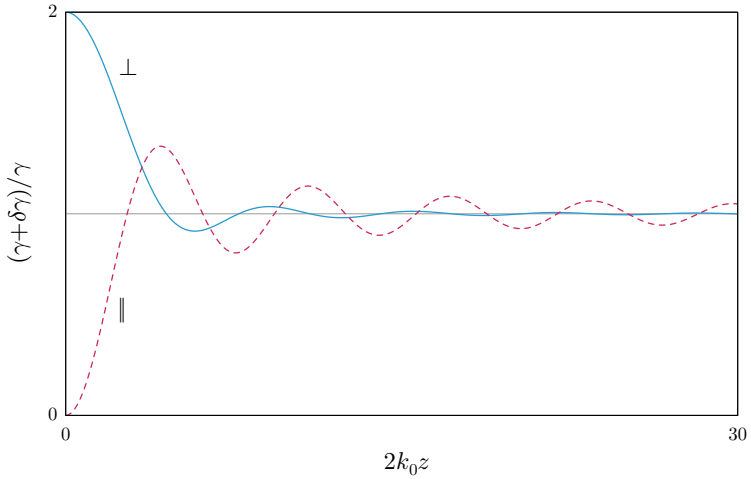
Note that here that in the near-resonant approximation,

$$z' = 2k_0 z. \tag{1.128}$$

The frequency shift (corresponding to a transition-frequency shift of a quantum-mechanical atom), is plotted here.



This effect can be interpreted as an ac Stark shift of the atom due to its own radiated field. Note that only the parallel component has a radiative $(1/z')$ component. In the near field, the frequency shift becomes very large until the dipole approximation breaks down. The decay rate has similar oscillatory behavior, but the prediction is not divergent as the atom approaches the wall.



Notice that the decay rate drops to zero as $z \longrightarrow 0$ for the parallel component; for this component, the dipole and its image are out of phase when the dipole is close to the mirror, and so there is complete destructive interference of the radiated field. For the perpendicular component, the decay rate increases to twice the free-space value as $z \longrightarrow 0$; for this component the dipole and image are in phase near the mirror surface, and there is constructive interference of the radiated and image fields, leading to a sort of **superradiance**.

It is possible to combine the expressions (1.126) and (1.127) into a more compact form that will be useful for later comparison as follows. Noting that

$$\begin{aligned}
\frac{1}{2}\delta\omega_{0\perp} - \delta\omega_{0\parallel} &= -\frac{3}{4}\gamma\frac{\cos z'}{z'} \\
\frac{1}{2}\delta\omega_{0\perp} + \delta\omega_{0\parallel} &= \frac{3}{4}\gamma\left[\left(\frac{1}{z'} - \frac{2}{z'^3}\right)\cos z' - 2\frac{\sin z'}{z'^2}\right] = -\frac{3}{4}\gamma\partial_{z'}^2\frac{\cos z'}{z'},
\end{aligned} \tag{1.129}$$

we see that we can write the total shift as

$$\delta\omega_0 = \frac{3}{4}\gamma\left[\left(\hat{\varepsilon}_\parallel^2/2 - \hat{\varepsilon}_\perp^2\right) - \left(\hat{\varepsilon}_\parallel^2/2 + \hat{\varepsilon}_\perp^2\right)\partial_{z'}^2\right]\frac{\cos z'}{z'}. \tag{1.130}$$

(transition frequency shift)

Similarly, since

$$\begin{aligned}\frac{1}{2}\delta\gamma_\perp - \delta\gamma_\parallel &= \frac{3}{2}\gamma\frac{\sin z'}{z'}\\ \frac{1}{2}\delta\gamma_\perp + \delta\gamma_\parallel &= -\frac{3}{2}\gamma\left[\left(\frac{1}{z'}-\frac{2}{z'^3}\right)\sin z' - 2\frac{\cos z'}{z'^2}\right] = \frac{3}{2}\gamma\partial_{z'}^2\frac{\sin z'}{z'},\end{aligned} \tag{1.131}$$

we have

$$\delta\gamma = -\frac{3}{2}\gamma\left[\left(\hat{\varepsilon}_\parallel^2/2-\hat{\varepsilon}_\perp^2\right) - \left(\hat{\varepsilon}_\parallel^2/2+\hat{\varepsilon}_\perp^2\right)\partial_{z'}^2\right]\frac{\sin z'}{z'}. \tag{1.132}$$

(damping-rate shift)

The polarization combinations have the following interpretation. Note that the combination $\hat{\varepsilon}_\parallel^2/2-\hat{\varepsilon}_\perp^2$ vanishes if the dipole has all three components equally excited (i.e., isotropic excitation), since it is proportional to $(\hat{x}^2+\hat{y}^2)/2-\hat{z}^2=0$. On the other hand, the combination $\hat{\varepsilon}_\parallel^2/2+\hat{\varepsilon}_\perp^2$ is proportional to $(\hat{x}^2+\hat{y}^2)/2+\hat{z}^2=2$. Thus, $\hat{\varepsilon}_\parallel^2/2-\hat{\varepsilon}_\perp^2$ gives the *anisotropic* part of the dipole excitation, while $\hat{\varepsilon}_\parallel^2/2+\hat{\varepsilon}_\perp^2$ gives the *isotropic* contribution to the dipole excitation (isotropic here is with respect to the parallel and perpendicular parts).

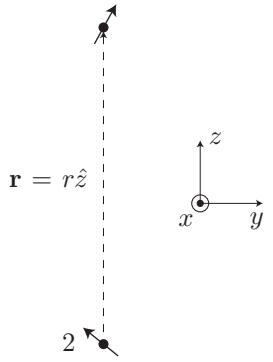
The above forms (1.130) and (1.132) are somewhat nonsensical in the classical framework, since the different shifts in the different directions lead to precession of the dipole vector; however, the interpretation is more sensible when we go over to quantum mechanics and interpret the polarization combinations as dipole matrix elements (Chapter 14). But for now, we can use these forms to arrive at the compact expressions

$$\begin{aligned}\delta\omega_{0\parallel} &= \frac{3}{8}\gamma\left(1-\partial_{z'}^2\right)\frac{\cos z'}{z'}\\ \delta\omega_{0\perp} &= -\frac{3}{4}\gamma\left(1+\partial_{z'}^2\right)\frac{\cos z'}{z'}\\ \delta\gamma_\parallel &= -\frac{3}{4}\gamma\left(1-\partial_{z'}^2\right)\frac{\sin z'}{z'}\\ \delta\gamma_\perp &= \frac{3}{2}\gamma\left(1+\partial_{z'}^2\right)\frac{\sin z'}{z'}\end{aligned} \tag{1.133}$$

for the separate shifts.

1.5.2 Two-Atom Radiation

A similar problem to the atom-wall interaction is the problem of two coupled atoms. (See Problem 6.7 for the quantum version of this problem.)



It is convenient again here to decompose the dipoles into parts that are perpendicular or parallel to the $z=0$ plane that separates to the two atoms. This is because the field due to the parallel part of one dipole couples only to the parallel part of the other, and the same is true for the perpendicular parts. This follows from the polarization properties of the radiated field. The field of dipole 1 at the location of dipole 2 has a form similar to that of (1.118), but now without the reversal of $\hat{\epsilon}_\parallel$, and with a separation of r.

$$\mathbf{E}^{(+)}_{\text{dipole 1}}(z,\omega) = -\frac{3}{2}\gamma\frac{m\omega_0}{e}\left\{\left[\frac{1}{(k_0r)^3}-i\frac{1}{(k_0r)^2}\right]\left[2\tilde{\mathbf{x}}_\perp^{(+)}(t)+\tilde{\mathbf{x}}_\parallel^{(+)}(t)\right]-\frac{1}{k_0r}\tilde{\mathbf{x}}_\parallel^{(+)}(t)\right\}e^{ik_0r}, \tag{1.134}$$

Here, we have assumed that the dipoles are oscillating at approximately their common resonance frequency ω_0, justifying the decomposition

$$\mathbf{x}_{\perp,\parallel}^{(+)}(t) = \tilde{\mathbf{x}}_{\perp,\parallel}^{(+)}(t)e^{-i\omega_0 t}, \tag{1.135}$$

where $\tilde{\mathbf{x}}_{\perp,\parallel}^{(+)}(t)$ is a slowly varying amplitude, so that

$$|\partial_t \tilde{\mathbf{x}}_{\perp,\parallel}^{(+)}(t)| \ll |\omega_0 \tilde{\mathbf{x}}_{\perp,\parallel}^{(+)}(t)|. \tag{1.136}$$

Then we can note that the approximate derivatives have the form

$$\begin{aligned} \dot{x}^{(+)} &\approx -i\omega_0 \tilde{x}^{(+)} e^{-i\omega_0 t} \\ \ddot{x}^{(+)} + \omega_0^2 x^{(+)} &\approx -i2\omega_0 \dot{\tilde{x}}^{(+)} e^{-i\omega_0 t}. \end{aligned} \tag{1.137}$$

Using the Lorentz model (1.18) again, with the above expressions for the time derivatives, we can write the equations of motion for the components of the second dipole reacting to the first dipole as

$$\begin{aligned} \dot{\tilde{z}}_2^{(+)} + \frac{\gamma}{2}\tilde{z}_2^{(+)} &= -i\frac{3}{2}\gamma\left[\frac{1}{(k_0 r)^3} - i\frac{1}{(k_0 r)^2}\right] e^{ik_0 r}\tilde{z}_1^{(+)} =: -i\Omega_z(r)\tilde{z}_1^{(+)} \\ \dot{\tilde{x}}_2^{(+)} + \frac{\gamma}{2}\tilde{x}_2^{(+)} &= -i\frac{3}{2}\gamma\left[\frac{1}{(k_0 r)^3} - i\frac{1}{(k_0 r)^2} - \frac{1}{k_0 r}\right] e^{ik_0 r}\tilde{x}_1^{(+)} =: -i\Omega_x(r)\tilde{x}_1^{(+)}. \end{aligned} \tag{1.138}$$

Here, we are writing out the explicit displacement components in coordinates, so that z is the perpendicular component, and x and y are the parallel components. Of course, $\tilde{y}_2^{(+)}$ satisfies an equation of the same form as $\tilde{x}_2^{(+)}$. Similarly, the equations of motion for the first dipole become

$$\begin{aligned} \dot{\tilde{z}}_1^{(+)} + \frac{\gamma}{2}\tilde{z}_1^{(+)} &= -i\Omega_z(r)\tilde{z}_2^{(+)} \\ \dot{\tilde{x}}_1^{(+)} + \frac{\gamma}{2}\tilde{x}_1^{(+)} &= -i\Omega_x(r)\tilde{x}_2^{(+)}, \end{aligned} \tag{1.139}$$

leading to pairs of coupled equations for each vector component of the displacement.[19]

For any component, we thus have a pair of coupled equations of the form

$$\begin{aligned} \dot{\alpha} + \frac{\gamma}{2}\alpha &= -i\Omega\beta \\ \dot{\beta} + \frac{\gamma}{2}\beta &= -i\Omega\alpha. \end{aligned} \tag{1.140}$$

We can solve these by the method of Laplace transforms. The Laplace transforms of the equations are

$$\begin{aligned} (s+\gamma/2)\mathscr{L}[\alpha] - \alpha_0 &= -i\Omega\mathscr{L}[\beta] \\ (s+\gamma/2)\mathscr{L}[\beta] - \beta_0 &= -i\Omega\mathscr{L}[\alpha], \end{aligned} \tag{1.141}$$

which we can decouple and solve. For example, the solution for α is

$$\mathscr{L}[\alpha] = \frac{(s+\gamma/2)\alpha_0}{(s+\gamma/2)^2 + \Omega^2} - i\frac{\Omega\beta_0}{(s+\gamma/2)^2 + \Omega^2}. \tag{1.142}$$

[19] Note that we are ignoring the time delay in the propagating waves between the two dipoles, which leads to "signaling" behavior between the two atoms. This amounts to a coarse-graining on time scales of order γ^{-1}, and is valid as long as $r \ll c/\gamma$. For alkali atoms, $\gamma^{-1} \sim 30$ ns, so this approximation is good as long as $r \ll 10$ m. See P. W. Milonni and P. L. Knight, "Retardation in coupled dipole–oscillator systems, *American Journal of Physics* **44**, 741 (1976) (doi: 10.1119/1.10122).

The inverse transform gives the solution in terms of the initial values α_0 and β_0:

$$\begin{aligned}\alpha(t) &= \alpha_0 e^{-\gamma t/2}\cos\Omega t - i\beta_0 e^{-\gamma t/2}\sin\Omega t \\ &= \frac{\alpha_0-\beta_0}{2}e^{-(\gamma/2-i\Omega)t} + \frac{\alpha_0+\beta_0}{2}e^{-(\gamma/2+i\Omega)t}.\end{aligned} \tag{1.143}$$

The first term here represents an *antisymmetric* or out-of-phase component in the initial excitation, while the second term represents the *symmetric* or in-phase component.

In view of the decomposition (1.135), we can see that the real and imaginary parts of the exponential frequencies are significant as shifts in the frequency and damping rate, as in the atom–mirror problem. In particular,

$$\begin{aligned}\delta\omega_{0\pm} &= \pm\mathrm{Re}[\Omega] \\ \delta\gamma_{\pm} &= \mp 2\mathrm{Im}[\Omega].\end{aligned} \tag{1.144}$$

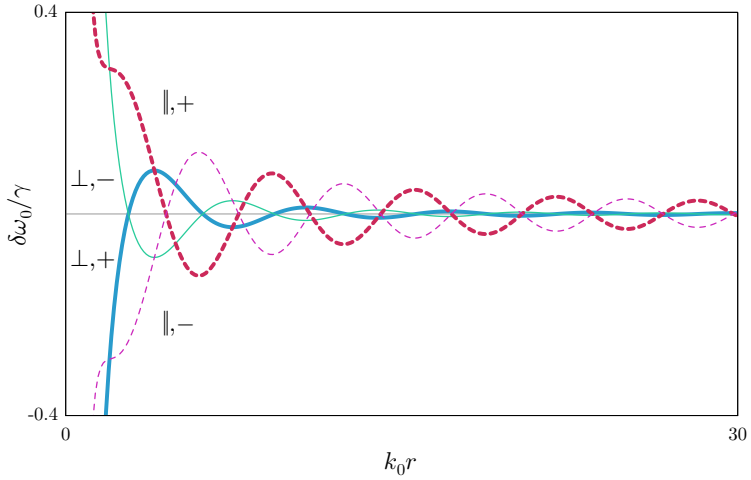
Here, the + subscript refers to the in-phase part, and the − subscript refers to the out-of-phase part. From Eq. (1.138), we can write the shifts as

$$\begin{aligned}\delta\omega_{0z\pm} &= \mp\frac{3}{2}\gamma\left[\frac{\sin k_0 r}{(k_0 r)^2}+\frac{\cos k_0 r}{(k_0 r)^3}\right] = \mp\frac{3}{4}\gamma\left(1+\partial_{r'}^2\right)\frac{\cos r'}{r'} \\ \delta\gamma_{z\pm} &= \mp 3\gamma\left[\frac{\cos k_0 r}{(k_0 r)^2}-\frac{\sin k_0 r}{(k_0 r)^3}\right] = \pm\frac{3}{2}\gamma\left(1+\partial_{r'}^2\right)\frac{\sin r'}{r'}\end{aligned} \tag{1.145}$$

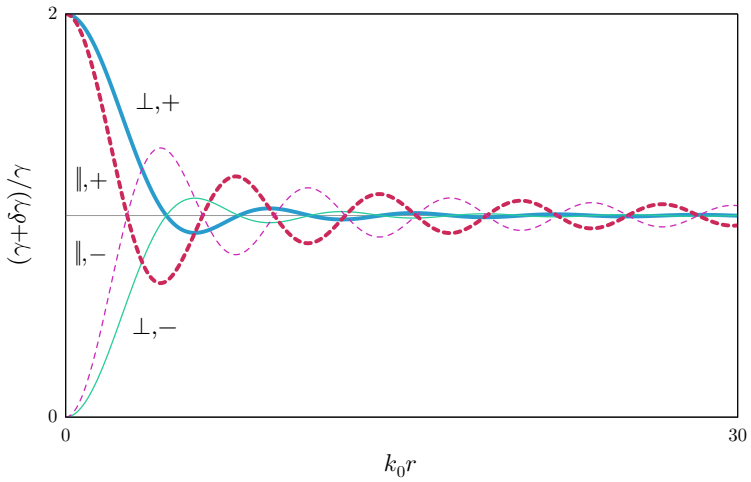
for the perpendicular components, and

$$\begin{aligned}\delta\omega_{0x\pm} &= \mp\frac{3}{4}\gamma\left[\left(\frac{1}{k_0 r}-\frac{1}{(k_0 r)^3}\right)\cos k_0 r-\frac{\sin k_0 r}{(k_0 r)^2}\right] = \mp\frac{3}{8}\gamma\left(1-\partial_{r'}^2\right)\frac{\cos r'}{r'} \\ \delta\gamma_{x\pm} &= \pm\frac{3}{2}\gamma\left[\left(\frac{1}{k_0 r}-\frac{1}{(k_0 r)^3}\right)\sin k_0 r+\frac{\cos k_0 r}{(k_0 r)^2}\right] = \pm\frac{3}{4}\gamma\left(1-\partial_{r'}^2\right)\frac{\sin r'}{r'}\end{aligned} \tag{1.146}$$

for the parallel components. To shorten the notation, we have used $r' = k_0 r$ here. Also, the y component again satisfies an equation of this same form. These relations have the same form as the atom–mirror shifts of Eqs. (1.126) and (1.127), in the sense that for the perpendicular component, the + solution matches the atom–mirror result, while for the parallel component, the − solution matches the atom–mirror case. This is what we expect from the phases of the image-dipole components relative to the source dipole. The shifts for both components and both relative phases are plotted below.



For the frequency (energy) shift, we can see that for small distances and in-phase dipoles, the perpendicular (z-z) orientation produces an attractive potential, while the parallel (x-x) orientation produces a repulsive potential, as we expect for static dipoles.



For the decay-rate shift, we see superradiant behavior for either orientation when the dipoles are in phase and subradiant behavior when the dipoles are out of phase. This is what we expect from the respective constructive or destructive interference of the emitted waves.

Note that again we can write all components together with both the symmetric and antisymmetric phases, here in coordinate-free form:

$$\delta\omega = -\frac{3}{8}\gamma\left\{[3(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2]\left(1+\partial_{r'}^2\right)-[(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2]\left(1-\partial_{r'}^2\right)\right\}\frac{\cos r'}{r'}$$
$$\delta\gamma = \frac{3}{4}\gamma\left\{[3(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2]\left(1+\partial_{r'}^2\right)-[(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2]\left(1-\partial_{r'}^2\right)\right\}\frac{\sin r'}{r'}.$$

(two-atom resonance and damping-rate shifts) (1.147)

In the far-field (the radiation zone), these simplify to

$$\begin{aligned}\delta\omega &= \frac{3}{4}\gamma\left[(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2\right]\frac{\cos r'}{r'}\\ \delta\gamma &= -\frac{3}{2}\gamma\left[(\hat{\varepsilon}_1\cdot\hat{r})(\hat{\varepsilon}_2\cdot\hat{r})-\hat{\varepsilon}_1\cdot\hat{\varepsilon}_2\right]\frac{\sin r'}{r'},\end{aligned} \tag{1.148}$$

since in this regime, only the $1/r'$ terms are important.

1.6 Exercises

Problem 1.1

Estimate the absorption coefficient of room-temperature rubidium (^{87}Rb) on the D_2 resonance at 780 nm as follows.

(a) Write down an expression for the one-dimensional atomic velocity distribution in the direction of the pumping laser beam. What is the rms velocity? Recalling that the Doppler shift is given by $\Delta\omega = k_0 v$, where v is the velocity component opposite to the pumping-beam direction, what is the corresponding full width at half maximum $\Delta\omega_D$ in frequency (the **Doppler width**)?

(b) Now write down an expression for the absorption coefficient that accounts for the Doppler broadening effect. The result involves a convolution of the frequency distribution from (a) with the natural lineshape, but you only need the result at one frequency. Noting that $1/\gamma = 26.2$ ns, you should argue that $\gamma \ll \Delta\omega_D$ to simplify your calculation.

(c) Use the above values for γ and λ_0 to compute the absorption oscillator strength f_{D_2} for this transition. Note that this is a $J = 1/2 \longrightarrow J' = 3/2$ transition, so that the degeneracy ratio $g'/g = 2$ (the prime here refers to the excited level).

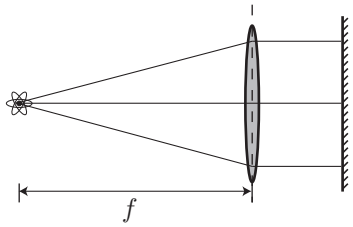
(d) Now give a numerical value for the absorption coefficient. Assume a vapor pressure for Rb of 3×10^{-7} torr at room temperature. Note that the relative abundance of ^{87}Rb is about 28% (the rest being ^{85}Rb).

(e) The answer to (d) does not predict the absorption coefficient well, because it assumes that the Doppler width is much larger than the splittings of the various hyperfine levels. In fact, this is only marginally true for the excited states, which span an effective range of 400 MHz. The ground-state doublet is in fact resolved at room temperature, since the splitting is 6.8 GHz. For what temperature range would you expect the above treatment to become valid?

We will return to the room-temperature case when we have the formalism to properly handle the hyperfine structure.

Problem 1.2

The radiation of an atom is collected and collimated by a lens of radius a and focal length f, reflected by a planar mirror, and imaged back onto the atom by the lens.



As in the direct atom–mirror interaction, the interaction of the atom with the distant mirror here causes a shift of the energy level that changes sinusoidally as the mirror is translated. This leads to potential wells for the atom due to its interaction with its own radiated field.[20]

(a) Give a *qualitative* explanation for why the sign of the potential varies with the distance to the mirror, and from your explanation predict the period of the variation.

(b) Using the classical model of the atom, give an expression for the depth of the potential wells near the focus (the maximum energy shift, in this case). Also assume for simplicity that the atom radiates

[20] This mechanical effect of a distant mirror on an atom has been observed experimentally with a trapped Ba ion. See Pavel Bushev, Alex Wilson, Jürgen Eschner, Christoph Raab, Ferdinand Schmidt-Kaler, Christoph Becher, and Rainer Blatt, "Forces between a Single Atom and Its Distant Mirror Image," *Physical Review Letters* **92** 223602 (2004).

in a spherically symmetric pattern, and that the lens is small enough that the radiation intensity is uniform in the plane of the lens. It will help to know the following result from diffraction theory: suppose a converging, spherical wave has radius of curvature f and a uniform intensity I_{aperture} over some aperture of radius a; then the maximum intensity is given by

$$I_0 = \frac{\pi^2 a^4}{\lambda^2 f^2} I_{\text{aperture}} \tag{1.149}$$

where the wave converges.[21]

(c) Use parameters appropriate to ^{87}Rb and assume the lens has a numerical aperture of 0.4 to give a numerical value to your answer in (b). Report your answer as a temperature.

Problem 1.3

Compute the frequency and decay-rate shifts for a Lorentz atom situated between two parallel mirrors separated by distance L and located a distance a from the closest mirror. You may assume that the dipole is oriented perpendicular to the mirrors. Make plots of your results for the case $a = L/2$.

Hint: how many image dipoles do you need to use? (The answer isn't two.)

Problem 1.4

Consider a vapor of atoms (with number density N), where the atoms are described by the Lorentz polarizability

$$\alpha(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{1.150}$$

We will use this model to explore the physics of **free-electron gases**, as in plasmas or metals, which corresponds to the limit $\omega_0 \longrightarrow 0$. (Note that metals have density-induced correlations that are missed by the models here.)

(a) First ignoring any damping ($\gamma \longrightarrow 0$), show that the complex refractive index for the free electron gas is

$$\tilde{n}(\omega) = \sqrt{1 - \left(\frac{\omega_{\text{p}}}{\omega}\right)^2}, \tag{1.151}$$

where $\omega_{\text{p}} := \sqrt{Ne^2/m\epsilon_0}$ is the **plasma frequency**. This is the **plasma model** for the refractive index.

(b) At an interface of the electron gas with vacuum (say, a metal–vacuum interface), the intensity (Fresnel) reflection coefficient for monochromatic light incident from the vacuum on the interface is

$$R(\omega) = \left|\frac{1 - \tilde{n}(\omega)}{1 + \tilde{n}(\omega)}\right|^2, \tag{1.152}$$

assuming normal incidence to the interface. This expression can be derived by imposing appropriate boundary conditions on the electromagnetic fields, but here you may take this expression to be given. According to the plasma model, what is $R(\omega)$ for $\omega < \omega_{\text{p}}$? What is $R(\omega)$ for $\omega \gg \omega_{\text{p}}$? What do these results imply for the reflectances of metal mirrors in the infrared and the deep UV, assuming a plasma frequency in the UV range?

(c) Now put the damping back in, to account for radiation reaction, electron collisions, and so on, and show that the complex refractive index becomes

$$\tilde{n}(\omega) = \sqrt{1 - \frac{\omega_{\text{p}}^2}{\omega^2 + i\gamma\omega}}. \tag{1.153}$$

[21] See Max Born and Emil Wolf, *Principles of Optics*, 7th (expanded) ed. (Cambridge, 1999), Eq. (22), p. 489.

This is the **Drude–Lorentz model** for the refractive index.

(d) Consider the limit of small frequency. In terms of the reflectance, how is the Drude–Lorentz model more physically sensible than the plasma model? Show as well that the Drude–Lorentz model reduces to the plasma model at high frequencies.

(e) The current density induced by the field is

$$\mathbf{j} = -Ne\dot{\mathbf{x}}. \tag{1.154}$$

Defining the **conductivity** of the free-electron gas by the relation $\mathbf{j} = \sigma\mathbf{E}$, show that the conductivity is given according to the Drude–Lorentz model by

$$\sigma(\omega) = \frac{\sigma_0}{1 - i\omega/\gamma}, \tag{1.155}$$

where σ_0 is the dc conductivity. What is σ_0?

Problem 1.5

We used the Lorentz model to describe the *linear* response of an atom to the field, but it can be modified to describe *nonlinear* optical media. Anharmonic terms added to the electron binding potential, for example, can describe basic nonlinear optics, but here we will explore nonlinear responses in free-electron gases.

(a) In the Lorentz model, we assumed a force induced by the electric field of the form $\mathbf{F} = -e\mathbf{E}$. However, the full Lorentz force, including the magnetic-field interaction, is $\mathbf{F} = -e\mathbf{E} - e\mathbf{v} \times \mathbf{B}$. The extra magnetic-field interaction is responsible for a nonlinear response at very large intensities. The classical equation of motion for a free electron (ignoring damping) is thus

$$m\dot{\mathbf{v}} = -e\mathbf{E} - e\mathbf{v} \times \mathbf{B}. \tag{1.156}$$

Assume the electron is driven by a linearly polarized, monochromatic plane wave

$$\mathbf{E}^{(+)}(\mathbf{r}, t) = \hat{\varepsilon} E_0^{(+)} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \tag{1.157}$$

with $\mathbf{B} = \hat{k} \times \mathbf{E}/c$ as required by Maxwell's equations. Now make a Fourier-series *ansatz* of the form

$$\mathbf{v}^{(+)}(t) = \sum_{j=1}^{\infty} \hat{\varepsilon}_{(j)} v_{(j)}^{(+)}(j\omega)\, e^{-ij\omega t}. \tag{1.158}$$

Put this form for $\mathbf{v}(t)$ into the electron equation of motion, ignoring the effect of the magnetic field, and obtain a zeroth-order perturbative solution for $\mathbf{v}^{(+)}(t)$.

(b) Put the zeroth-order solution back into the right-hand side of the equation of motion, and obtain the next-order solution for $\mathbf{v}^{(+)}(t)$, this time including the effect of the magnetic-field coupling. Write down the corresponding induced dipole moment at frequency 2ω. The nonlinear response thus allows **second-harmonic generation** from the driving field. In which direction is there no second-harmonic radiation? (Note that you should be able to interpret your solution as an electron tracing out a "figure eight.")

(c) Estimate the intensity at which the second-harmonic dipole moment becomes comparable to the dipole moment at the fundamental frequency.

(d) Now instead of the magnetic-field interaction, consider the relativistic interaction with the electric field, given by

$$\dot{\mathbf{p}} = \frac{d}{dt}(\gamma m v) = \frac{d}{dt}\left(\frac{mv}{\sqrt{1 - \dfrac{v^2}{c^2}}}\right) = -e\mathbf{E}, \tag{1.159}$$

where γ is the usual relativistic factor and the driving field is the same as before. Show that the second harmonic vanishes in this case.

(e) Obtain a perturbative expression for the relativistic dipole moment at frequency 3ω.

(f) At what intensity is the third-harmonic relativistic dipole comparable to the fundamental dipole?

Problem 1.6

We have already seen how the field responds to an atomic vapor, given the Lorentz model for the atoms: the real part of the polarizability results in a phase shift (dispersion), while the imaginary part leads to absorption. In this problem, you will explore a general classical model for how the electromagnetic field responds to the **polarization field** $\mathbf{P}(\mathbf{r}, t)$ of a medium. The polarization, or the dipole moment per unit volume, is the macroscopic generalization of the atomic dipole moment.

(a) Use Maxwell's equations in a source-free, dielectric medium,

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{H} &= \partial_t \mathbf{D}, \end{aligned} \tag{1.160}$$

where

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{1.161}$$

and the electric fields are related by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \tag{1.162}$$

to derive the polarization-driven wave equation

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \partial_t^2 \mathbf{E} = \frac{1}{\epsilon_0 c^2} \partial_t^2 \mathbf{P}. \tag{1.163}$$

You may assume the polarization field to be *transverse*, i.e., $\nabla \cdot \mathbf{P} = 0$. Generally speaking, the polarization is induced by the field, but here we may view it as an independent object acting to modify the field.

(b) Now assume the fields have the form

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \hat{\epsilon} \frac{E_0}{2} e^{i(kz - \omega t + \phi)} + \text{c.c.} \\ \mathbf{P}(\mathbf{r}, t) &= \hat{\epsilon} P_0^{(+)} e^{i(kz - \omega t + \phi)} + \text{c.c.}, \end{aligned} \tag{1.164}$$

where $\phi(z, t)$ is a slowly varying phase, $E_0(z, t)$ is a slowly varying (real) field amplitude, and $P_0^{(+)}(t)$ is a slowly varying (complex, varying in time only) polarization amplitude. Of course, due to a possible phase lag of the medium, the polarization phase may have a phase different from that of the field. In this case, ''slowly varying'' means

$$|\partial_t E_0| \ll \omega E_0; \qquad |\partial_z E_0| \ll k E_0, \tag{1.165}$$

with similar relations holding for ϕ and $|P_0^{(+)}|$. Put the fields into the above wave equation, making the approximation of slowly varying amplitude and phase, to derive the equations

$$\begin{aligned} \partial_z E_0 + \frac{1}{c} \partial_t E_0 &= -\frac{k}{\epsilon_0} \text{Im}\left[P_0^{(+)}\right] \\ E_0 \left(\partial_z \phi + \frac{1}{c} \partial_t \phi\right) &= \frac{k}{\epsilon_0} \text{Re}\left[P_0^{(+)}\right] \end{aligned} \tag{1.166}$$

for the effect of the polarization on the field. We again see the imaginary part causes absorption (or gain), and the real part causes a phase shift. These relations are important, for example, in laser physics in treating the effect of the gain medium on the laser field.

Note: in the slowly varying approximation, you should throw away second derivatives of E_0 and ϕ, as well as other second-derivative terms, e.g., of the form $(\partial_t E_0)(\partial_t \phi)$. This problem treats the lowest-order modification of the field due to the polarization field, so you should discard *all* derivatives of $P_0^{(+)}$. Finally, since we are after a perturbative result, you may assume that E_0 approximately satisfies the homogeneous equation

$$-k^2 E_0 + \frac{\omega^2}{c^2} E_0 \approx 0. \tag{1.167}$$

That is, ignoring the variation, E_0 represents a valid solution in the absence of the medium.

Chapter 2

Classical Coherence

Coherence theory is concerned with characterizing light, especially its fluctuation properties that influence how it will act in an interference-type experiment. This is covered to some extent by the spectral profile of the field, which is one of the themes here, but isn't the full story. Obviously things will get more interesting in the quantum case, but it is important to establish some classical results immediately so we can characterize atomic radiation.

Let's now consider an interference experiment that involves a range of frequencies. Recalling the interference of two monochromatic plane waves, we can write the superposition of the two waves as

$$E^{(+)}_{\rm sum}(\mathbf{r},t) = E^{(+)}_{10}e^{i(kz-\omega t)} + E^{(+)}_{20}e^{i(kz-\omega t)}e^{i\phi}. \tag{2.1}$$

Here, ϕ is a relative phase difference between the two waves. Recall that we are just writing the positive-frequency components, so that the physical fields also must include the negative-frequency parts.

The intensity of the superposition is

$$I_{\rm sum} = \langle|\mathbf{E}_{\rm sum}\times\mathbf{H}_{\rm sum}|\rangle = \frac{1}{\eta}\left\langle (E_{\rm sum})^2\right\rangle. \tag{2.2}$$

Writing this out explicitly in terms of the component fields,

$$I_{\rm sum} = \frac{1}{\eta}\left\langle \left(E^{(+)}_{10}e^{i(kz-\omega t)} + E^{(+)}_{20}e^{i(kz-\omega t)}e^{i\phi} + E^{(-)}_{10}e^{-i(kz-\omega t)} + E^{(-)}_{20}e^{-i(kz-\omega t)}e^{-i\phi}\right)^2\right\rangle. \tag{2.3}$$

The optical terms of the form $\exp(\pm i2\omega t)$ vanish in the time average, so we obtain

$$\begin{aligned} I_{\rm sum} &= \frac{2}{\eta}E^{(-)}_{10}E^{(+)}_{10} + \frac{2}{\eta}E^{(-)}_{20}E^{(+)}_{20} + \frac{2}{\eta}E^{(-)}_{10}E^{(+)}_{20}e^{i\phi} + \frac{2}{\eta}E^{(-)}_{20}E^{(+)}_{10}e^{-i\phi} \\ &= I_1 + I_2 + \left[\frac{2}{\eta}E^{(-)}_{10}E^{(+)}_{20}e^{i\phi} + \text{c.c.}\right]. \end{aligned} \tag{2.4}$$

Again, the interference is in the terms with the relative phase ϕ.

Suppose that the phase difference represents a difference in optical path length, in the form of a time delay τ. Then $\phi = -\omega\tau$, and so

$$I_{\rm sum} = I_1 + I_2 + \left[\frac{2}{\eta}E^{(-)}_{10}E^{(+)}_{20}e^{-i\omega\tau} + \text{c.c.}\right]. \tag{2.5}$$

Now let's handle the case of multiple frequencies. To simplify things, we'll assume that the two waves have equal amplitude and come from a common source. Then the intensity *density* at frequency ω is

$$I_{\rm sum}(\omega) = 2I(\omega) + \left[\frac{2}{\eta}|E^{(+)}_0(\omega)|^2e^{-i\omega\tau} + \text{c.c.}\right]. \tag{2.6}$$

Note that the notation here is a little funny; the frequency-dependent quantities $I(\omega)$ and $E^{(+)}(\omega)$ don't have the respective dimensions of intensity and electric field; rather, $I(\omega)\,d\omega$ and $|E^{(+)}(\omega)|^2\,d\omega$ are the intensity and (squared) electric field, respectively, in the frequency interval between ω and $\omega + d\omega$.

Now we can sum over all frequencies to find the total intensity:

$$\begin{aligned} I_{\text{total}} &= \int_0^\infty I_{\text{sum}}(\omega)\,d\omega \\ &= 2\int_0^\infty I(\omega)\,d\omega + \left[\frac{2}{\eta}\int_0^\infty |E_0^{(+)}(\omega)|^2 e^{-i\omega\tau}\,d\omega + \text{c.c.}\right]. \end{aligned} \tag{2.7}$$

Note that the frequency integral ranges only over *positive* frequencies; we've already accounted for the *negative* frequencies by including the complex conjugate terms. Thus the intensity spectrum $I_{\text{sum}}(\omega)$ is a *one-sided* spectrum, which is common when working with intensities and powers. We can now recognize the second integral in the last expression as a "one-sided" Fourier transform of $|E_0^{(+)}(\omega)|^2$, where we recall the normalization convention for ω-t Fourier transforms:

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^\infty \tilde{f}(\omega)\,e^{-i\omega t}\,d\omega, \quad \tilde{f}(\omega) = \int_{-\infty}^\infty f(t)\,e^{i\omega t}\,dt. \tag{2.8}$$
(ω-t Fourier transform)

But what is this, when we don't know the general form of the electric field?

2.1 Wiener–Khinchin Theorem

Recall the convolution theorem for functions f and g (see Section 17.1.2):

$$\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g]. \tag{2.9}$$

Writing out the convolution integral explicitly,

$$(f * g)(t) = \int_{-\infty}^\infty f(t')\,g(t-t')\,dt' = \mathscr{F}^{-1}\left[\mathscr{F}[f]\mathscr{F}[g]\right]. \tag{2.10}$$

If we make the particular choice $g(t) = f^*(-t)$, then

$$\mathscr{F}[g(t)] = \mathscr{F}[f^*(-t)] = \int_{-\infty}^\infty f^*(-t)\,e^{i\omega t}\,dt = \int_{-\infty}^\infty f^*(t)\,e^{-i\omega t}\,dt = \left(\mathscr{F}[f(t)]\right)^*. \tag{2.11}$$

Thus, Eq. (2.10) becomes

$$\int_{-\infty}^\infty f(t')\,f^*(t'-t)\,dt' = \mathscr{F}^{-1}\left[\left|\mathscr{F}[f]\right|^2\right]. \tag{2.12}$$

Inverting the transform and letting $t' \longrightarrow t' + t$, we obtain the **Wiener–Khinchin theorem**:

$$\mathscr{F}\left[\int_{-\infty}^\infty f^*(t')\,f(t+t')\,dt'\right] = \left|\mathscr{F}[f]\right|^2. \tag{2.13}$$
(Wiener–Khinchin theorem)

The function on the left-hand side,

$$\int_{-\infty}^\infty f^*(t)\,f(t+\tau)\,dt, \tag{2.14}$$
(autocorrelation function)

is the **autocorrelation function** of $f(t)$. Essentially, it compares f to itself but shifted by an amount τ by computing an overlap integral. We can understand the right-hand side by noting that $\mathscr{F}[f]$ is the *spectrum* of f, and so $|\mathscr{F}[f]|^2$ is the **energy spectral density** of f. Essentially, the energy spectrum is the square of the usual spectrum, with the phase information removed. This is consistent with the notion of energy going

as the square of a signal amplitude. Thus, the Wiener–Khinchin theorem states that the Fourier transform of the autocorrelation function gives the energy spectrum.

There is one subtle point to these definitions: for some signals, such as steady optical signals, the correlation integral diverges:

$$\int_{-\infty}^{\infty} f^*(t)\, f(t+\tau)\, dt \longrightarrow \infty. \tag{2.15}$$

In this case, we should consider a time average instead of the normal integral. For an averaging time of T,

$$\langle f^*(t)\, f(t+\tau)\rangle_T := \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)\, f(t+\tau)\, dt. \tag{2.16}$$

For bounded signals, this integral is guaranteed to converge. To be physically sensible, T should be a suitably long observation time (e.g., long enough to resolve the frequency spectrum). For such signals, we can write the Wiener–Khinchin theorem as

$$\mathscr{F}\left[\langle f^*(t)\, f(t+\tau)\rangle_T\right] = \mathscr{F}^*[f]\, \mathscr{F}_T[f]. \tag{2.17}$$

Here we have defined a finite-time Fourier transform by

$$\mathscr{F}_T[f] := \frac{1}{T}\int_{-T/2}^{T/2} f(t)\, e^{i\omega t}\, dt. \tag{2.18}$$

Defining this seems a bit funny, but it avoids problems with singular spectra. Now the Wiener–Khinchin theorem says that the Fourier transform of the (time-averaged) correlation function is the **power spectral density**, or the energy spectral density per unit time. For a *stationary* process, the correlation function is independent of t (generally for a sufficiently long averaging time T). Then we can extend the averaging time $T \longrightarrow \infty$. Denoting this long-time averaging limit as

$$\langle f^*(t)\, f(t+\tau)\rangle = \langle f^*(t)\, f(t+\tau)\rangle_{T\to\infty}\,, \tag{2.19}$$

we can thus write the Wiener–Khinchin theorem as

$$\mathscr{F}\left[\langle f^*(t)\, f(t+\tau)\rangle\right] = \lim_{T\to\infty}\frac{1}{T}\left|\int_{-T/2}^{T/2} f(t)\, e^{i\omega t}\, dt\right|^2. \tag{2.20}$$

(Wiener–Khinchin theorem, time-average form)

Again, the right-hand side is the power spectral density, and in this form it is more clear that this is the energy density per unit time.

2.2 Optical Wiener–Khinchin Theorem

In terms of stationary optical fields, the Wiener–Khinchin theorem (2.20) becomes

$$\int_0^{\infty} I(\omega)\, e^{-i\omega\tau}\, d\omega = \frac{2}{\eta}\langle E^{(-)}(t)E^{(+)}(t+\tau)\rangle. \tag{2.21}$$

(optical Wiener–Khinchin theorem)

This is because the intensity density $I(\omega) = \langle|\mathbf{S}|\rangle$ *is* the time-averaged power spectral density of the optical field. Note that from the inverse Fourier transform, there is conventionally a factor of $1/2\pi$, but it is missing here because it is already implicitly included in $I(\omega)$. We can see this from the boundary condition at $\tau = 0$, which gives the total intensity:

$$\int_0^{\infty} I(\omega)\, d\omega = \frac{2}{\eta}\langle E^{(-)}(t)E^{(+)}(t)\rangle = \frac{2}{\eta}\left\langle |E^{(+)}(t)|^2\right\rangle. \tag{2.22}$$

In optics, as in statistics and other fields, it is conventional to define a normalized correlation function

$$g^{(1)}(\tau) := \frac{\langle E^{(-)}(t)E^{(+)}(t+\tau)\rangle}{\langle E^{(-)}(t)E^{(+)}(t)\rangle},$$

(degree of first-order temporal coherence) (2.23)

so that

$$\frac{2}{\eta}\langle E^{(-)}(t)E^{(+)}(t+\tau)\rangle = \left(\int_0^\infty I(\omega)\,d\omega\right) g^{(1)}(\tau). \tag{2.24}$$

The normalization is such that $g^{(1)}(\tau = 0) = 1$. That is, a correlation value of unity indicates perfect correlation. Note that we could just as well have written the correlation function as $\langle E^{(+)}(t+\tau)E^{(-)}(t)\rangle$, but it turns out that the order of $E^{(+)}$ and $E^{(-)}$ matters in quantum mechanics, so we'll be careful to stick to the form in Eq. (2.23). Notice also that from the definition of $g^{(1)}(\tau)$, the correlation function is subject to the constraint

$$g^{(1)}(-\tau) = [g^{(1)}(\tau)]^*. \tag{2.25}$$

(time symmetry of correlation function)

In quantum optics, $g^{(1)}(\tau)$ is called the **degree of first-order temporal coherence**. The light is said to be **coherent** if $|g^{(1)}(\tau)|^2 = 1$, **incoherent** if $|g^{(1)}(\tau)|^2 = 0$, and **partially coherent** otherwise (for $\tau \neq 0$). This function can be generalized to include spatial correlation by

$$g^{(1)}(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) := \frac{\langle E^{(-)}(\mathbf{r}_1, t_1)E^{(+)}(\mathbf{r}_2, t_2)\rangle}{\sqrt{\langle |E^{(+)}(\mathbf{r}_1, t_1)|^2\rangle \langle |E^{(+)}(\mathbf{r}_2, t_2)|^2\rangle}},$$

(degree of first-order coherence) (2.26)

which is the **degree of first-order coherence**. We will focus on the case (2.23) of temporal coherence.

Returning to the interference result of Eq. (2.7), we find

$$\begin{aligned} I_{\text{total}} &= 2\int_0^\infty I_{\text{sum}}(\omega)\,d\omega + \left[\frac{2}{\eta}\int_0^\infty |E_0^{(-)}(\omega)|^2 e^{-i\omega\tau}\,d\omega + \text{c.c.}\right] \\ &= \left(\int_0^\infty I(\omega)\,d\omega\right)\left\{2 + \left[g^{(1)}(\tau) + \text{c.c.}\right]\right\}, \end{aligned} \tag{2.27}$$

and thus

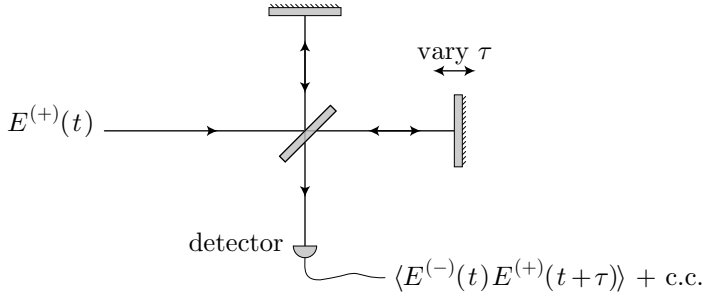
$$I_{\text{total}} = 2\left(\int_0^\infty I(\omega)\,d\omega\right)\left\{1 + \text{Re}\left[g^{(1)}(\tau)\right]\right\}.$$

(interferometer signal in terms of $g^{(1)}(\tau)$) (2.28)

It is worth keeping in mind that the form (2.13) is still relevant for pulsed fields, in which case the result here is modified so that the coefficient in front is the integrated *power* rather than intensity.

2.2.1 Michelson Interferometer

One example where the interference result (2.28) arises directly—and thus where the optical Wiener–Khinchin theorem is very useful—is in the Michelson interferometer. The interferometer splits and interferes the beam with itself, and the path-length difference τ varies with the displacement of one of the mirrors.



The photodetector at the output measures the time average of the product of the fields, up to an overall constant, and thus measures $g^{(1)}(\tau)$, just as in Eq. (2.28).

The Michelson interferometer thus gives a method for measuring the spectrum of an optical field. The idea is then to digitize the output of the photodetector as a function of the mirror displacement, effectively recording $\mathrm{Re}[g^{(1)}(\tau)]$. Computing the Fourier transform of the signal on the computer gives the spectrum $I(\omega)$. (Note that $I(\omega)$ is real, and so the imaginary parts of $g^{(1)}(\tau)$ don't contribute to the spectrum.) This is the technique behind, for example, **Fourier-transform infrared (FTIR) spectroscopy**.

2.2.2 Example: Monochromatic Light

As a simple example, let's compute the correlation function for monochromatic light and verify that the Wiener–Khinchin theorem makes sense. Let's take a monochromatic wave of the form

$$E^{(+)}(t) = E_0^{(+)} e^{-i\omega_0 t}. \tag{2.29}$$

The correlation function is

$$\left\langle E^{(-)}(t) E^{(+)}(t+\tau) \right\rangle = \left| E_0^{(+)} \right|^2 e^{-i\omega_0 \tau}. \tag{2.30}$$

In normalized form, this function becomes

$$g^{(1)}(\tau) = e^{-i\omega_0 \tau}. \tag{2.31}$$

Thus, a wave with harmonic time dependence leads to a harmonic correlation function. This is true *independent* of the phase of the input field; that is, the correlation function does not reflect any extra phase in $E_0^{(+)}$.

The power spectrum is easy to calculate via the Wiener–Khinchin theorem:

$$\mathscr{F}\left[\left\langle E^{(-)}(t) E^{(+)}(t+\tau) \right\rangle\right] = \left| E_0^{(+)} \right|^2 \mathscr{F}\left[e^{-i\omega_0 \tau}\right] = \left| E_0^{(+)} \right|^2 \delta(\omega - \omega_0). \tag{2.32}$$

Again, the harmonic time dependence produces a single frequency in the power spectrum. Of course, with our convention of positive and negative frequencies, there would be a matching $\delta(\omega + \omega_0)$ component, but this is already included in the *one*-sided spectrum $I(\omega)$.

Let's now compute the power spectrum directly, using Eq. (2.17). The normal spectrum is

$$\mathscr{F}\left[E^{(+)}(t)\right] = E_0^{(+)} \delta(\omega - \omega_0). \tag{2.33}$$

The finite-time transform is

$$\mathscr{F}_T\left[E^{(+)}(t)\right] = \frac{1}{T} \int_{-T/2}^{T/2} E_0^{(+)} e^{i(\omega - \omega_0)t}\, dt = E_0^{(+)} \mathrm{sinc}\left[(\omega - \omega_0)T/2\right]. \tag{2.34}$$

Note that the value of the sinc function is 1 at $\omega = \omega_0$. Thus, the spectrum is

$$\mathscr{F}^*\left[E^{(+)}(t)\right] \mathscr{F}_T\left[E^{(+)}(t)\right] = E_0^{(-)} \delta(\omega - \omega_0) E_0^{(+)} \mathrm{sinc}\left[(\omega - \omega_0)T/2\right] = \left| E_0^{(+)} \right|^2 \delta(\omega - \omega_0). \tag{2.35}$$

This result is consistent with Eq. (2.32). Note that without being careful with finite-time transforms, we would run into something bad involving the square of a δ-function.

2.2.3 Spectrum of Atomic Radiation

With this formalism, it is straightforward to analyze the spectrum of the scattered light from the Lorentz atom. If the atom is briefly excited, then the amplitude of the field decays exponentially with time. The

autocorrelation function does the same, and the Fourier transform of a decaying exponential is a Lorentzian lineshape (see Problem 2.1). Thus, the radiated spectrum has the form

$$s(\omega) = \frac{\gamma/2\pi}{(\omega_0 - \omega)^2 + (\gamma/2)^2}. \tag{2.36}$$

(Actually, this turns out to be the spectrum $s_+(\omega)$ that we will define below, corresponding to a coherence function $g^{(1)}(\tau) = e^{-i\omega_0\tau}e^{-\gamma\tau/2}$.) On the other hand, if the atom is driven by a monochromatic field, the dipole oscillates sinusoidally in steady state at the *driving* frequency ω_{L} (say of the driving laser), rather than the atomic resonance frequency. In this case, the radiated spectrum is monochromatic: $s(\omega) = \delta(\omega - \omega_{\text{L}})$. The above broadened spectrum is a transient effect that is swamped by this "elastic" peak (elastic since the scattered light has the same frequency as the incident light).

2.2.4 Normalized One- and Two-Sided Spectra

Now we will be a bit more precise about the nature of the spectrum, to avoid confusion with different possible conventions for the power spectrum. First, it's convenient to define a *normalized* spectral density (lineshape function)

$$s(\omega) := \frac{I(\omega)}{\displaystyle\int_0^\infty I(\omega)\, d\omega}. \tag{2.37}$$
(normalized spectral density)

Note that this spectrum extends only over *positive* frequencies, as the intensity spectrum $I(\omega)$ corresponded to *physical* frequency components. Examining the inverse Fourier transform, we combine Eqs. (2.21) and (2.24) to obtain

$$g^{(1)}(\tau) = \int_0^\infty s(\omega)\, e^{-i\omega\tau}\, d\omega.$$
(first-order coherence in terms of normalized, one-sided spectral density) (2.38)

However, the usual inverse Fourier transform has an integral extending over both positive and negative frequencies. We can obtain something of this form by considering the *real* part of the correlation function,

$$\begin{aligned}\text{Re}\left[g^{(1)}(\tau)\right] &= \frac{g^{(1)}(\tau) + \left[g^{(1)}(\tau)\right]^*}{2} = \frac{g^{(1)}(\tau) + g^{(1)}(-\tau)}{2} \\ &= \frac{1}{2}\int_0^\infty s(\omega)\, e^{-i\omega\tau}\, d\omega + \frac{1}{2}\int_0^\infty s(\omega)\, e^{i\omega\tau}\, d\omega,\end{aligned} \tag{2.39}$$

so that if we define a **two-sided spectrum** $s_\leftrightarrow(\omega)$ (for all $\omega \in \mathbb{R}$) via

$$s_\leftrightarrow(\omega) := \begin{cases} s(\omega)/2 & (\omega > 0) \\ s(\omega) & (\omega = 0) \\ s(-\omega)/2 & (\omega < 0), \end{cases}$$
(two-sided spectrum in terms of one-sided spectrum) (2.40)

which satisfies

$$s_\leftrightarrow(-\omega) = s_\leftrightarrow(\omega), \tag{2.41}$$
(symmetry of two-sided spectrum)

we obtain

$$\text{Re}\left[g^{(1)}(\tau)\right] = \int_{-\infty}^\infty s_\leftrightarrow(\omega)\, e^{-i\omega\tau}\, d\omega.$$
(first-order coherence in terms of two-sided spectrum) (2.42)

We have lost the imaginary part, but *only* the real part can contribute to a physical result: $g^{(1)}(\tau)$ must always be accompanied by its conjugate, as we saw in the interference experiment.

Inverting this last relation, we have

$$\begin{aligned} s_{\leftrightarrow}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{Re}\left[g^{(1)}(\tau)\right] e^{i\omega\tau}\, d\tau \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} g^{(1)}(\tau)\cos\omega\tau\, d\tau \\ &= \frac{1}{2\pi}\int_{0}^{\infty} g^{(1)}(\tau)\cos\omega\tau\, d\tau + \text{c.c.} \qquad (\omega\in\mathbb{R}). \end{aligned}$$

(two-sided spectral density in terms of first-order coherence) (2.43)

The one-sided spectrum can then be written in terms of the double-sided spectrum as

$$s(\omega) = \begin{cases} s_{\leftrightarrow}(\omega) + s_{\leftrightarrow}(-\omega) = 2s_{\leftrightarrow}(\omega) & (\omega > 0) \\ s_{\leftrightarrow}(\omega) & (\omega = 0) \\ 0 & (\omega < 0), \end{cases}$$

(two-sided spectrum in terms of one-sided spectrum) (2.44)

so that the one-sided spectrum simply concentrates all the power at positive and negative frequencies on the positive side. Of course, this cannot be done for *any* spectrum, but the power spectral density—the Fourier transform of the coherence function—has a special structure due to the symmetry of the correlation function. Thus, we can write the one-sided spectrum as

$$\begin{aligned} s(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{Re}\left[g^{(1)}(\tau)\right] e^{i\omega\tau}\, d\tau \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} g^{(1)}(\tau)\cos\omega\tau\, d\tau \\ &= \frac{1}{\pi}\int_{0}^{\infty} g^{(1)}(\tau)\cos\omega\tau\, d\tau + \text{c.c.} \qquad (\omega\geq 0) \end{aligned}$$

(one-sided spectral density in terms of first-order coherence) (2.45)

in terms of the coherence function. Combining Eqs. (2.42) and (2.44), we then find

$$\mathrm{Re}\left[g^{(1)}(\tau)\right] = \int_{0}^{\infty} s(\omega)\cos\omega\tau\, d\omega$$

(first-order coherence in terms of one-sided spectrum) (2.46)

for the reverse transformation.

Finally, note that it can be more convenient to associate spectra separately with $g^{(1)}(\tau)$ and its conjugate, by defining

$$s_{\pm}(\omega) := \frac{1}{4\pi}\int_{-\infty}^{\infty} g^{(1)}(\tau)\, e^{\pm i\omega\tau}\, d\tau = \frac{1}{4\pi}\int_{0}^{\infty} g^{(1)}(\tau)\, e^{\pm i\omega\tau}\, d\tau + \text{c.c.} \qquad (\omega\in\mathbb{R}),$$

(component spectral densities in terms of first-order coherence) (2.47)

so that

$$s_{+}(\omega) = s_{-}(-\omega).$$

(symmetry of component spectra) (2.48)

Then the relations

$$\begin{aligned} s_{\leftrightarrow}(\omega) &= s_{+}(\omega) + s_{-}(\omega) = s_{+}(\omega) + s_{+}(-\omega) \qquad (\omega\in\mathbb{R}) \\ s(\omega) &= 2[s_{+}(\omega) + s_{-}(\omega)] = 2[s_{+}(\omega) + s_{+}(-\omega)] \qquad (\omega\geq 0) \end{aligned}$$

(component spectral densities in terms of first-order coherence) (2.49)

recover the total spectra (2.43) and (2.45), respectively.

2.3 Visibility

The example in Eq. (2.31) of the correlation function for a monochromatic wave is special in the sense that the correlation function does not decay with τ. This is important, because the correlation function is the magnitude of the interference terms. To quantify this better, we can define the *fringe visibility* as

$$\mathscr{V} := \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{2.50}$$
(fringe visibility)

where $I_{\max}$ and $I_{\min}$ are the maximum and minimum intensities achieved for phase variations on the order of several π. For example, complete interference results in intensity variation from 0 to some maximum, and so $\mathscr{V} = 1$. For no interference, the intensity does not vary with phase and so $\mathscr{V} = 0$. Partial coherence is represented by intermediate values of the visibility.

Writing out the explicit phase of the correlation function,

$$g^{(1)}(\tau) = |g^{(1)}(\tau)|e^{i\phi(\tau)}, \tag{2.51}$$

and so Eq. (2.28) becomes

$$I_{\text{total}} = 2\left(\int_0^\infty I(\omega)\, d\omega\right)\left[1 + |g^{(1)}(\tau)|\cos\phi(\tau)\right]. \tag{2.52}$$

The cosine varies from -1 to 1, so the visibility is just the magnitude of the correlation function:

$$\mathscr{V} = |g^{(1)}(\tau)|. \tag{2.53}$$
(visibility in terms of coherence)

For monochromatic light, $\mathscr{V} = 1$. Actually, this is only true if the amplitudes of the input waves are equal. For two monochromatic waves of *unequal* amplitude, the visibility becomes

$$\mathscr{V} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}. \tag{2.54}$$
(monochromatic, unbalanced interference)

The important thing to note is that for interference of monochromatic light, the visibility is independent of τ.

2.4 Coherence Time, Coherence Length, and Uncertainty Measures

As we have just seen, a peculiarity of monochromatic light is that the coherence does not decay with τ. In the generic case, where light is composed of a range of frequencies, the visibility drops as the phase difference increases, since $g^{(1)}(\tau) \longrightarrow 0$ as $\tau \longrightarrow \infty$. Intuitively, this is because as light waves with different wavelengths propagate, the monochromatic components tend to dephase. Mathematically, we can express this as an uncertainty relationship. Recall from Section 2.2.4 that the coherence $\mathrm{Re}[g^{(1)}(\tau)]$ and the normalized (two-sided) spectral density (lineshape function) form a Fourier-transform pair. Actually, $2\pi s_{\leftrightarrow}(\omega)$ is the second half of the pair, due to the form of the transforms (2.42) and (2.43): recall that in the ω-t convention, there is normally a factor of $1/2\pi$ in front of the *inverse*-transform integral, as in Eqs. (2.8).

Thus, if we define the root-mean-square (rms) widths of these two functions, regarding $\mathrm{Re}[g^{(1)}(\tau)]^2$ and $[s_{\leftrightarrow}(\omega)]^2$ as (unnormalized) probability distributions, we can write down an **uncertainty relation**, as in quantum mechanics:

$$\delta\omega_{\text{rms}}\,\delta\tau_{\text{rms}} \geq \frac{1}{2}. \tag{2.55}$$
(rms uncertainty relation)

For many distributions, we can say that the equality is more or less satisfied to

$$\delta\omega_{\text{rms}}\,\delta\tau_{\text{rms}} \sim \frac{1}{2}, \tag{2.56}$$

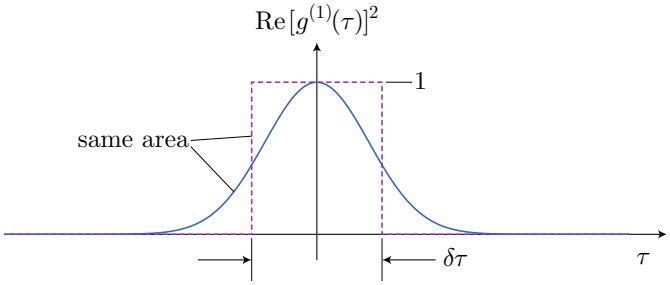
and thus see that the "widths" of $g^{(1)}(\tau)$ and $s_{\leftrightarrow}(\omega)$ are inversely related. The problem is that for some useful functions in optics, these uncertainties can diverge (e.g., for Lorentzian functions) or vary by orders of magnitude. The uncertainty inequality is always satisfied, of course, but as a practical relation for the temporal and frequency widths, this is less useful.

Thus we will adopt other uncertainty conventions in time and frequency[1] We can define the **coherence time** as the power-equivalent width of the correlation function:

$$\delta\tau := \int_{-\infty}^{\infty} \left|\mathrm{Re}\left[g^{(1)}(\tau)\right]\right|^2 d\tau. \tag{2.57}$$

(coherence time, power-equivalent width)

The idea here is to first note that $g^{(1)}(\tau = 0) = 1$. Thus the width $\delta\tau$ is the width of a box of unit height, with the same area as $|\mathrm{Re}[g^{(1)}(\tau)]|^2$. That is, $\delta\tau$ is the width of a unit-height box signal with the same *power* as $|\mathrm{Re}[g^{(1)}(\tau)]|^2$.



We'll take a slightly different convention for the width of the frequency spectrum. Let's define the *average value of the one-sided normalized spectrum* as

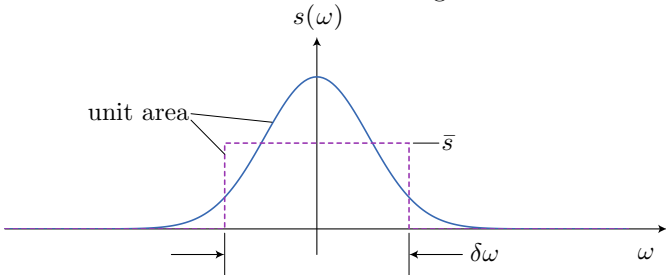
$$\bar{s} = \int_0^{\infty} s^2(\omega)\, d\omega = 2\int_{-\infty}^{\infty} s^2_{\leftrightarrow}(\omega)\, d\omega. \tag{2.58}$$

That is, we're regarding $s(\omega)$ as a probability distribution (because it has the appropriate normalization), and then we're calculating the expectation value of $s(\omega)$ with respect to itself. Then we'll define the effective frequency width by the reciprocal of the average value:

$$\delta\omega := (\bar{s})^{-1} = \left[\int_0^{\infty} s^2(\omega)\, d\omega\right]^{-1}. \tag{2.59}$$

(spectral width)

Thus, $\delta\omega$ is the width of the box function of unit area and height $\bar{s}$.



Note that we have constructed the diagram as if $s(\omega)$ is two-sided, but the argument carries through for the one-sided case as well, being careful to keep track of factors of 2. The definitions for $\delta\tau$ and $\delta\omega$ *look* like inverses, but they're really quite different because of the different normalizations of the two functions.

The big advantage of these definitions is that they are related in a simple way. We can write

$$\delta\tau\,\delta\omega = \frac{\displaystyle\int_{-\infty}^{\infty} \left|\mathrm{Re}\left[g^{(1)}(\tau)\right]\right|^2 d\tau}{\displaystyle 2\int_{-\infty}^{\infty} s^2_{\leftrightarrow}(\omega)\, d\omega}. \tag{2.60}$$

[1] as in Max Born and Emil Wolf, *Principles of Optics*, 7th (expanded) ed. (Cambridge, 1999).

We can use **Parseval's theorem** to evaluate this ratio, which states that the signal power is equivalently measured in either the time or frequency basis:

$$\int_{-\infty}^{\infty} |f(t)|^2 \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\tilde{f}(\omega)\right|^2 \, d\omega. \tag{2.61}$$
(Parseval's theorem)

Noting again that $\mathrm{Re}[g^{(1)}(\tau)]$ and $2\pi s_{\leftrightarrow}(\omega)$ form a Fourier-transform pair, and recalling the factor of two for using the one-sided spectrum, we can use this to write

$$\delta\tau \, \delta\omega = \pi. \tag{2.62}$$
(uncertainty relation)

This "uncertainty relation" is a strict equality, valid for any functions as long as the measures exist and are finite. Neat, eh?

The point of all this is, for time (optical path length) delays larger than the coherence time $\delta\tau$, the fringe visibility is mostly gone. This coherence time corresponds to a physical path length difference

$$\ell_c := c \, \delta\tau, \tag{2.63}$$
(coherence length)

which is called the **coherence length**.

Here are some examples. For a He-Ne laser, the laser line width is $\delta\nu = \delta\omega/2\pi \sim 1$ GHz. This corresponds to a coherence time of $\delta\tau \sim 1$ ns, or a coherence length $\ell_c \sim 30$ cm. On the other hand for a light bulb that spans the visible wavelength range of 400-700 nm, the line width is

$$\delta\nu \sim \frac{\nu \, \delta\lambda}{\lambda} = \frac{c \, \delta\lambda}{\lambda^2} = 300 \text{ THz}. \tag{2.64}$$

This gives a coherence time $\delta\tau \sim 3$ fs and a coherence length $\ell_c \sim 1$ μm. So in fact it is possible to see interference of white light in a Michelson, but it's very difficult because the path lengths must be matched to μm accuracy. On the other hand, it's much easier to observe interference or record a hologram with light from a He-Ne laser, because it remains coherent on the scale of about a foot.

2.5 Interference Between Two Partially Coherent Sources

In general, we can now look at the interference pattern between two partially coherent sources, represented by the two fields $E_1^{(+)}(t)$ and $E_2^{(+)}(t)$. The second field has an adjustable time delay of τ. Then the intensity of the superposition of these waves is

$$\begin{aligned} I &= \frac{2}{\eta} \left\langle \left| E_1^{(+)}(t) + E_2^{(+)}(t+\tau) \right| \right\rangle \\ &= \frac{2}{\eta} \left\langle \left| E_1^{(+)}(t) \right| \right\rangle + \frac{2}{\eta} \left\langle \left| E_2^{(+)}(t+\tau) \right| \right\rangle + \left[\frac{2}{\eta} \left\langle \left| E_1^{(-)}(t) E_2^{(+)}(t+\tau) \right| \right\rangle + \text{c.c.} \right] \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2} \, \mathrm{Re}\left[g_{12}^{(1)}(\tau) \right], \end{aligned} \tag{2.65}$$

where $g_{12}^{(1)}(\tau)$ is the **normalized cross-correlation function**:

$$g_{12}^{(1)}(\tau) := \frac{\left\langle \left| E_1^{(-)}(t) E_2^{(+)}(t+\tau) \right| \right\rangle}{\left\langle \left| E_1^{(-)}(t) E_2^{(+)}(t) \right| \right\rangle}. \tag{2.66}$$
(normalized cross-correlation)

Again, the visibility is

$$\mathscr{V} = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2} \left| g_{12}^{(1)}(\tau) \right|. \tag{2.67}$$
(visibility for two different fields)

This tells us that very different beams, resulting in little correlation, don't interfere very well.

2.6 Second-Order Coherence

The **degree of second-order coherence** is the autocorrelation function for the *intensity*, rather than the field. We can define this function as

$$g^{(2)}(\tau) := \frac{\langle E^{(-)}(t)E^{(-)}(t+\tau)E^{(+)}(t+\tau)E^{(+)}(t)\rangle}{\langle E^{(-)}(t)E^{(+)}(t)\rangle^2}.$$

(degree of second-order coherence) (2.68)

Classically, this expression becomes

$$g^{(2)}(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I\rangle^2}. \tag{2.69}$$

For the first-order coherence, we had that $0 \leq |g^{(1)}(\tau)| \leq 1$, since $E^{(+)}(t)$ is always at least as correlated with $E^{(+)}(t)$ as with $E^{(+)}(t+\tau)$, but this constraint is somewhat different for $g^{(2)}(\tau)$, because the normalization convention is somewhat different.

First let's consider the variance of $I(t)$, which is manifestly nonnegative:

$$\left\langle \Big(I(t) - \langle I(t)\rangle \Big)^2 \right\rangle \geq 0. \tag{2.70}$$

Multiplying this out, this constraint becomes

$$\langle I^2(t)\rangle \geq \langle I(t)\rangle^2, \tag{2.71}$$

which implies the boundary condition

$$g^{(2)}(0) = \frac{\langle I^2\rangle}{\langle I\rangle^2} \geq 1$$

(2.72) (classical bunching constraint)

for the second-order coherence. There is no corresponding upper bound, and this argument fails for $\tau \neq 0$. This inequality is *classical*, and we will see that it is possible for a quantum field to violate this condition (see Section 5.7.5 or Problem 8.12). This inequality effectively provides a boundary between quantum and classical statistical behavior.

We can also start with the inequality

$$\Big[I(t) - I(t+\tau) \Big]^2 \geq 0, \tag{2.73}$$

which when multiplied out yields

$$I^2(t) + I^2(t+\tau) \geq 2I(t)\, I(t+\tau). \tag{2.74}$$

Taking the time average of this relation then gives

$$\langle I^2(t)\rangle \geq \langle I(t)I(t+\tau)\rangle. \tag{2.75}$$

This implies

$$g^{(2)}(0) \geq g^{(2)}(\tau),$$

(2.76) (classical constraint)

so the second-order coherence is a nonincreasing function, at least in the vicinity of $\tau = 0$.

The second-order coherence is related to how ''concentrated'' or ''bunched'' the intensity function is. For example, a monochromatic wave has $I(t) = I$, which gives $g^{(2)} = 1$. Again, the monochromatic wave is coherent at second order. Note that we can generalize the correlation functions in an obvious way to arbitrarily high order, where it turns out that a monochromatic wave always takes on the value unity and is thus coherent to all orders. However, the first- and second-order degrees of coherence are the most important and relevant for experiments.

For the opposite extreme of a periodic train of short pulses,

$$I(t) = A\sum_n \delta(t - nT), \tag{2.77}$$

we can see that

$$\langle I\rangle = \frac{A}{T} \tag{2.78}$$

and

$$\langle I^2\rangle = \frac{A^2}{T}\sum_n \delta(\tau - nT), \tag{2.79}$$

so that

$$g^{(2)}(\tau) = T\sum_n \delta(\tau - nT). \tag{2.80}$$

Note that in deriving Eq. (2.79), we have used the relation

$$\int \delta(t)\delta(t+\tau)\,dt = \lim_{m\to\infty}\int me^{-\pi m^2 t^2}\delta(t+\tau)\,dt = \lim_{m\to\infty} me^{-\pi m^2\tau^2} = \delta(\tau), \tag{2.81}$$

where we have used the definition of the delta function as a limit of a sequence of normalized Gaussians of decreasing width. This relation also makes sense as a convolution, since the delta function is the identity kernel of the convolution operation. Thus, we see that very large values of $g^{(2)}(\tau)$ correspond to temporally concentrated intensity. This is the classical manifestation of *photon bunching*.

2.6.1 Thermal Light

To model light from a thermal source, we will assume that $E^{(+)}(t)$ fluctuates as a stationary Gaussian random process. That is, $E^{(+)}(t)$ is a complex Gaussian random variable whose statistics are time-independent, and $E^{(+)}(t)$ is correlated with the itself at other times as required by the power spectrum.

Gaussian random processes are fundamental in modeling noisy systems, as they allow for tremendous simplifications. We will call Z a **complex Gaussian random variable** if $Z = X + iY$, where X and Y are independent and identically distributed Gaussian random variables. The joint probability density of X and Y is

$$f(x,y) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}, \tag{2.82}$$

if we assume that $\langle X\rangle = \langle Y\rangle = 0$. Then we can write the probability density of Z as

$$\begin{aligned} f(z) &= \frac{1}{2\pi\sigma^2}e^{-|z|^2/2\sigma^2} \\ &= \frac{1}{\pi\sigma_z^2}e^{-|z|^2/\sigma_z^2}, \end{aligned} \tag{2.83}$$

where

$$\sigma_z^2 = \langle |Z|^2\rangle = 2\sigma^2 \tag{2.84}$$

is the variance of Z. Generalizing this to N complex Gaussian variables $Z_1,\ldots,Z_N$, the probability density becomes

$$f(z_\alpha) = \frac{1}{\pi^N \det S_{\alpha\beta}}e^{-z_\alpha^*(S^{-1})_{\alpha\beta}z_\beta}, \tag{2.85}$$

where summations are implied by repeated indices, and

$$S_{\alpha\beta} := \langle Z_\alpha Z_\beta^*\rangle \tag{2.86}$$

is the **covariance matrix**. For such complex Gaussian variables, one can show that high-order moments factor as[2]

$$\langle Z^*_{\alpha_1} Z^*_{\alpha_2} \cdots Z^*_{\alpha_N} Z_{\beta_M} \cdots Z_{\beta_1} \rangle = \sum_{\substack{\text{all } N! \\ \text{pairings}}} \langle Z^*_{\alpha_1} Z_{\beta_1} \rangle \langle Z^*_{\alpha_2} Z_{\beta_2} \rangle \cdots \langle Z^*_{\alpha_N} Z_{\beta_N} \rangle$$

(Gaussian moment factorization) (2.87)

if $M = N$, with the moment vanishing otherwise. We can always factor high-order moments because the Gaussian distribution (2.85) itself only depends on the quadratic moments. In particular, we need the factorization for the fourth-order moment

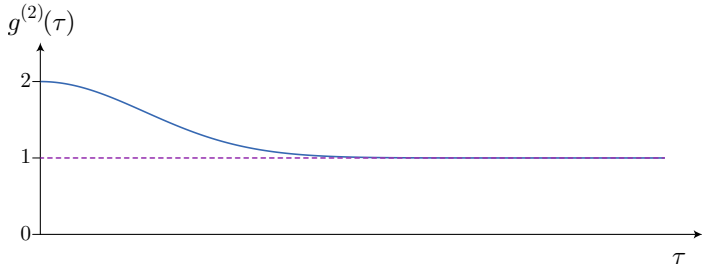
$$\begin{aligned} \langle Z^*_1 Z^*_2 Z_2 Z_1 \rangle &= \langle Z^*_1 Z_1 \rangle \langle Z^*_2 Z_2 \rangle + \langle Z^*_1 Z_2 \rangle \langle Z^*_2 Z_1 \rangle \\ &= \langle |Z_1|^2 \rangle \langle |Z_2|^2 \rangle + |\langle Z^*_1 Z_2 \rangle|^2. \end{aligned} \tag{2.88}$$

Applying this to the second-order coherence function (2.68), we find the important relation

$$g^{(2)}(\tau) = 1 + |g^{(1)}(\tau)|^2, \tag{2.89}$$

(thermal-field coherence constraint)

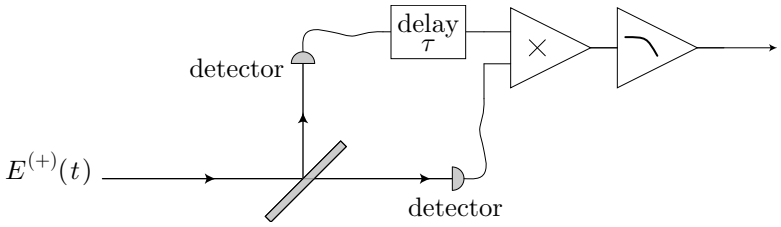
which again is valid for light with Gaussian fluctuations in the field. Recalling that $g^{(1)}(0) = 1$, we now have the boundary condition $g^{(2)}(0) = 2$ for thermal light. Additionally, $g^{(1)}(\tau) \longrightarrow 0$ as $\tau \longrightarrow \infty$, so $g^{(2)}(\tau) \longrightarrow 1$ as $\tau \longrightarrow \infty$.



The correlation function thus drops from 2 to 1 with τ, as shown here for light with a Gaussian power spectrum (in addition two the Gaussian fluctuations). Thus, thermal light exhibits some degree of bunching.

2.6.2 Experiment of Hanbury Brown and Twiss

One important feature of the second-order coherence function is that it can be measured with a reasonably simple setup, the famous **Hanbury-Brown–Twiss apparatus**.[3] A simplified schematic of this type of apparatus is shown here.



An input field is divided by a beam splitter, and the two components are monitored by two photodetectors. The two detector signals are fed into a signal multiplier (mixer), though only after a variable time delay is added to one of the signals. The mixer signal is fed through a low-pass filter, which can be thought of as a integrator with a running time average.

[2] see Leonard Mandel and Emil Wolf, *Optical Coherence and Quantum Optics* (Cambridge, 1995), p. 38.

[3] R. Hanbury Brown and R. Q. Twiss, "Correlation Between Photons in Two Coherent Beams of Light," *Nature* **177**, 27 (1956).

This setup, because it effectively correlates the two intensities, seems to give the $g^{(2)}(\tau)$ function as its output signal directly. But the detectors are generally ac-coupled, so that the detectors monitor $I(t) - \langle I \rangle$. Then the output signal of the mixer is

$$V_{\text{mixer}} \propto [I(t) - \langle I \rangle][I(t+\tau) - \langle I \rangle], \tag{2.90}$$

and the output signal of the low-pass filter is

$$V_{\text{low-pass}} \propto \langle [I(t) - \langle I \rangle][I(t+\tau) - \langle I \rangle] \rangle = \langle I(t) I(t+\tau) \rangle - \langle I \rangle^2. \tag{2.91}$$

After proper normalization, we can see that the output is just $g^{(2)}(\tau) - 1$. Thus, for thermal light this setup should give some signal for $\tau = 0$ that decays to zero for large delays.

The setup here makes it somewhat intuitive how the quantum field will violate the inequality (2.72). If the quantum field arrives as a stream of separated, individual photons, then at any given time, only one of the photodetectors can ''click.'' That is, a single photon causes a detection event on one detector or the other, but necessarily not both. This implies that $g^{(2)}(\tau = 0)$ can go to zero in such a situation. This is the phenomenon of **antibunching**.

The treatment here is simplified, since taking into account the response time of the detectors complicates things considerably. The original experiment used the same idea to measure the *spatial* correlation of intensities, which can be used to measure the size of the optical source. This has important applications, for example, in astronomy, where the signals from two separated telescopes can be mixed in the same way. In such a **stellar interferometry** arrangement, the measured correlation function gives a measure of the diameter of the distant star. For our purposes, the quantum-mechanical version of this experiment that we will return to later demonstrates some of the uniquely quantum-mechanical features of light.

2.7 Phase Noise

2.7.1 Spectra of Phase and Frequency Fluctuations

An important form of noise in oscillators and lasers is **phase noise**, for example where the time dependence of a signal has the form

$$f(t) \sim \cos[\omega_0 t + \phi(t)], \tag{2.92}$$

where $\phi(t)$ is a stochastic process. The total phase of this signal is $\phi_{\text{total}} = \omega_0 t + \phi(t)$, and the instantaneous frequency is just the time derivative of the phase:

$$\omega(t) = \frac{d\phi_{\text{total}}}{dt} = \omega_0 + \frac{d\phi(t)}{dt}. \tag{2.93}$$

Thus, the phase noise translates into frequency noise as well.

Given a signal with time dependence of the form $\exp[-i\phi(t)]$, let us define an unnormalized, one-sided spectral density of phase fluctuations via

$$S_\phi(\omega) := \int_{-\infty}^{\infty} \langle \phi(t)\, \phi(t+\tau) \rangle \cos\omega\tau \, d\tau \qquad (\omega \geq 0),$$

(spectral density of phase fluctuations (one-sided)) (2.94)

in analogy with Eq. (2.45). Correspondingly, if we are instead interested in the spectral density of *frequency* fluctuations

$$S_\omega(\omega) := \int_{-\infty}^{\infty} \left\langle \dot{\phi}(t)\, \dot{\phi}(t+\tau) \right\rangle \cos\omega\tau \, d\tau \qquad (\omega \geq 0).$$

(spectral density of frequency fluctuations (one-sided)) (2.95)

These spectra are related by

$$S_\omega(\omega) = \omega^2 S_\phi(\omega)$$

(phase to frequency spectrum conversion) (2.96)

(see Problem 2.4). Thus, for example, if $\phi(t)$ is a random-walk process, then $\dot{\phi}$ is a white-noise process, so that $S_\omega(\omega)$ is independent of frequency (i.e., $\phi(t) \propto W(t)$ is a Wiener process, whose derivative gives white noise, as in Chapter 17). Inverting the above relations, we find

$$\langle \phi(t)\, \phi(t+\tau) \rangle = \frac{1}{\pi} \int_0^\infty S_\phi(\omega) \cos \omega\tau \, d\omega,$$

(phase correlation in terms of phase spectral density) (2.97)

as we find by analogy to Eqs. (2.45) and (2.46), noting the difference in the 2π factor. Also,

$$\langle \phi(t)\, \phi(t+\tau) \rangle = \frac{1}{\pi} \int_0^\infty S_\omega(\omega) \frac{\cos \omega\tau}{\omega^2} \, d\omega,$$

(phase correlation in terms of freqency spectral density) (2.98)

as follows from Eq. (2.96)

2.7.2 Variance of Phase Fluctuations

A useful quantity to consider is the variance of the phase fluctuation in a time τ. Thus, we define

$$\Delta\phi(\tau) := \phi(t+\tau) - \phi(t), \tag{2.99}$$

which we may regard as independent of t in a statistical sense, under the assumption of a stationary process. Then

$$\begin{aligned} \left\langle [\Delta\phi(\tau)]^2 \right\rangle &= 2\langle \phi^2(t) \rangle - 2\langle \phi(t)\phi(t+\tau) \rangle \\ &= \frac{2}{\pi} \int_0^\infty S_\omega(\omega) \frac{(1-\cos\omega\tau)}{\omega^2} \, d\omega, \end{aligned} \tag{2.100}$$

where we used Eq. (2.98). Thus, we have the result

$$\left\langle [\Delta\phi(\tau)]^2 \right\rangle = \frac{4}{\pi} \int_0^\infty S_\omega(\omega) \frac{\sin^2(\omega\tau/2)}{\omega^2} \, d\omega.$$

(variance of phase fluctuations related to frequency-noise spectrum) (2.101)

Note that this integral gives a finite result even for white noise, which has an unnormalized spectrum, due to the $1/\omega^2$ factor, but it diverges for a "$1/f$" spectrum.

2.7.3 Spectrum of the Signal

Supposing the phase of a signal fluctuates, with a particular spectrum of frequency fluctuations. What is the effect on the spectrum of the signal itself?[4] Suppose that we have an optical signal of the form

$$E^{(+)}(t) \sim e^{-i\omega_0 t} e^{-i\phi(t)} \tag{2.102}$$

(technically, we could be mixing positive- and negative-frequency components, but this doesn't really matter). We then want the correlation function

$$g^{(1)}(\tau) \sim \left\langle E^{(-)}(t)\, E^{(+)}(t+\tau) \right\rangle, \tag{2.103}$$

which when normalized looks like

$$g^{(1)}(\tau) = e^{-i\omega_0\tau} \left\langle e^{-i\Delta\phi(\tau)} \right\rangle. \tag{2.104}$$

Making the assumption that $\Delta\phi(\tau)$ represents Gaussian noise (which, from the central-limit theorem, is guaranteed essentially provided that the variance of $\Delta\phi(\tau)$ is finite), we can rewrite this as

$$g^{(1)}(\tau) = e^{-i\omega_0\tau} e^{-\langle [\Delta\phi(\tau)]^2 \rangle/2}. \tag{2.105}$$

[4] D. S. Elliott, Rajarshi Roy, and S. J. Smith, "Extracavity laser band-shape and bandwidth modification," *Physical Review A* **26**, 12 (1982) (doi: 10.1103/PhysRevA.26.12).

(See Problem 2.5 for the intermediate steps here.) Now using Eqs. (2.45) for the one-sided, normalized spectrum,

$$\begin{aligned} s(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty} g^{(1)}(\tau)\cos\omega\tau\, d\tau \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} d\tau \cos\omega_0\tau\cos\omega\tau\, e^{-\langle[\Delta\phi(\tau)]^2\rangle/2} \\ &= \frac{2}{\pi}\int_{0}^{\infty} d\tau \cos\omega_0\tau\cos\omega\tau\, e^{-\langle[\Delta\phi(\tau)]^2\rangle/2}, \end{aligned} \tag{2.106}$$

and so

$$s(\omega) = \frac{2}{\pi}\int_0^\infty d\tau\cos\omega_0\tau\cos\omega\tau\,\exp\left[-\frac{2}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\,d\omega'\right] \qquad (\omega \geq 0).$$

(spectrum of the signal) (2.107)

Thus, we have the spectral density for the signal itself, in terms of the spectral density for the frequency fluctuations. This is the spectrum that is more likely to be analyzed, e.g., on a spectrum analyzer (the phase- or frequency-fluctuation spectra would be observed directly only with a phase or frequency detector, for example, using a phase-locked loop to track the signal).

2.7.3.1 Example: White Noise

Suppose we take the simple case of a flat, white-noise frequency-fluctuation spectrum:

$$S_\omega(\omega') = \gamma. \tag{2.108}$$

Then the exponent in Eq. (2.107) is

$$-\frac{2}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\,d\omega' = -\frac{2\gamma}{\pi}\int_0^\infty d\omega'\,\frac{\sin^2(\omega'\tau/2)}{\omega'^2} = -\frac{\gamma|\tau|}{2}. \tag{2.109}$$

So

$$\begin{aligned} s(\omega) &= \frac{2}{\pi}\int_0^\infty d\tau\cos\omega_0\tau\cos\omega\tau\, e^{-\gamma|\tau|/2} \\ &= \frac{1}{\pi}\int_0^\infty d\tau\left[\cos(\omega-\omega_0)\tau + \cos(\omega+\omega_0)\tau\right]e^{-\gamma\tau/2} \\ &= \left[\frac{1}{2\pi}\int_0^\infty d\tau\, e^{i(\omega-\omega_0)\tau-\gamma\tau/2} + \text{c.c.}\right] + (\omega_0 \longrightarrow -\omega_0) \\ &= \frac{1}{2\pi}\left[\frac{1}{-i(\omega-\omega_0)+\gamma/2} + \text{c.c.}\right] + (\omega_0 \longrightarrow -\omega_0), \end{aligned} \tag{2.110}$$

and thus

$$s(\omega) = \frac{\gamma/2\pi}{(\omega-\omega_0)^2+(\gamma/2)^2} + \frac{\gamma/2\pi}{(\omega+\omega_0)^2+(\gamma/2)^2},$$

(heterodyne spectrum, white frequency noise) (2.111)

which is a properly normalized, one-sided Lorentzian spectrum with a full width at half maximum of γ (the second term being the negative-frequency "mirror image" of the positive-frequency Lorentzian).

2.8 Optical Linewidth Measurements

2.8.1 Photodetection Spectrum

The spectrum analyzer measures the power spectrum of the detector photocurrent $I_{\text{det}}(t)$, which by the Wiener–Khinchin theorem is the Fourier transform of the autocorrelation function

$$G_{\text{analyzer}}(\tau) = \frac{1}{Z_{\text{in}}}\langle I_{\text{det}}(t)\, I_{\text{det}}(t+\tau)\rangle\,, \tag{2.112}$$

where Z_{in} is the input impedance of the spectrum analyzer (typically 50 Ω), which yields the appropriate units of power. The photocurrent is related to the optical intensity at the detector by

$$I_{\text{det}}(t) = \eta_I A_{\text{det}} I(t), \tag{2.113}$$

where $I(t)$ is the optical intensity, averaged over the detector surface, A_{det} is the detector area, and η_I is the detector response (with units of A/W). Recalling that intensity is related to field by $I = 2|E^{(+)}|/\eta_0$, where $\eta_0 = 1/\epsilon_0 c$ is the impedance of free space (377 Ω), we can thus write

$$G_{\text{analyzer}}(\tau) = \frac{4\eta_I^2 A_{\text{det}}^2}{\eta_0^2 Z_{\text{in}}} \left\langle E_{\text{det}}^{(-)}(t)\, E_{\text{det}}^{(-)}(t+\tau)\, E_{\text{det}}^{(+)}(t+\tau)\, E_{\text{det}}^{(+)}(t) \right\rangle, \tag{2.114}$$

where $E_{\text{det}}^{(+)}(t)$ is the optical field at the detector. Thus, we find

$$S_{\text{analyzer}}(\omega) = \frac{4\eta_I^2 A_{\text{det}}^2}{\pi\eta_0^2 Z_{\text{in}}} \int_{-\infty}^{\infty} d\tau \left\langle E_{\text{det}}^{(-)}(t)\, E_{\text{det}}^{(-)}(t+\tau)\, E_{\text{det}}^{(+)}(t+\tau)\, E_{\text{det}}^{(+)}(t) \right\rangle e^{i\omega\tau} \tag{2.115}$$

for the spectrum-analyzer signal, where we introduce the factor of π to match the normalization convention of Eq. (2.45). This ensures that the integrated spectrum gives the total measured electrical power (at least the ac part, since we dropped the dc components), which is just $G_{\text{analyzer}}(\tau = 0)$. We can then also write

$$S_{\text{analyzer}}(\omega) = \frac{8\eta_I^2 A_{\text{det}}^2}{\pi\eta_0^2 Z_{\text{in}}} \int_{0}^{\infty} d\tau \left\langle E_{\text{det}}^{(-)}(t)\, E_{\text{det}}^{(-)}(t+\tau)\, E_{\text{det}}^{(+)}(t+\tau)\, E_{\text{det}}^{(+)}(t) \right\rangle \cos\omega\tau, \quad \text{(detected signal on spectrum analyzer)} \tag{2.116}$$

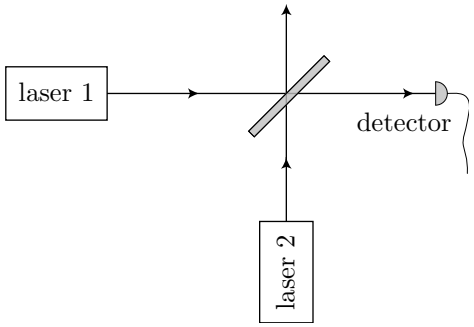
due to the even symmetry of the correlation function. Spectrum analyzers typically use normal (not angular) frequencies, so we can write

$$S_{\text{analyzer}}(\nu) = \frac{16\eta_I^2 A_{\text{det}}^2}{\eta_0^2 Z_{\text{in}}} \int_{-\infty}^{\infty} d\tau \left\langle E_{\text{det}}^{(-)}(t)\, E_{\text{det}}^{(-)}(t+\tau)\, E_{\text{det}}^{(+)}(t+\tau)\, E_{\text{det}}^{(+)}(t) \right\rangle \cos 2\pi\nu\tau, \quad \text{(detected signal on spectrum analyzer)} \tag{2.117}$$

where using the transformation $S(\omega)\, d\omega = S(\nu)\, d\nu$ (with $\omega = 2\pi\nu$) introduces an overall factor of 2π here.

2.8.2 Heterodyne Spectroscopy

In a heterodyne measurement, we beat two *independent* lasers together, but assume they are statistically identical—otherwise, it is not possible to attribute unbalanced fluctuations to the proper laser.



Then the signal field is the superposition of the two fields,

$$E_{\text{det}}^{(+)}(t) = \sqrt{\eta_1} E_1^{(+)}(t) + \sqrt{\eta_2} E_2^{(+)}(t), \tag{2.118}$$

where η_1 and η_2 are the intensity transmission coefficients through the beam splitter and any other optical components from the respective laser source to the detector. Then we may write the fields as

$$E_1^{(+)}(t) = E_{01}^{(+)} e^{-i\omega_1 t - i\phi_1(t)}, \qquad E_2^{(+)}(t) = E_{02}^{(+)} e^{-i\omega_2 t - i\phi_2(t)}, \tag{2.119}$$

where again the phase noises on the two lasers are independent but statistically identical. We also assume the two lasers have nearly the same center frequency, so we define the detuning

$$\Delta := \omega_2 - \omega_1, \qquad |\Delta| \ll \omega_1, \omega_2. \tag{2.120}$$

However, we also assume that the frequency width induced by the phase fluctuations is small compared to $|\Delta|$. Now we must substitute this expression into the correlation function (2.114), noting that of the 16 total terms, the 10 that rotate in time with frequencies that are optical *or* $|\Delta|$ will simply average to zero in the relatively slow spectrum-analyzer response time. Thus, we have

$$G_{\text{analyzer}}(\tau) \propto \eta_1^2 |E_{01}^{(+)}|^4 + \eta_2^2 |E_{02}^{(+)}|^4 + 2\eta_1\eta_2 |E_{01}^{(+)} E_{02}^{(+)}|^2 + \eta_1\eta_2 |E_{01}^{(+)} E_{02}^{(+)}|^2 \left\langle e^{-i\Delta\tau + i\Delta\phi_1(\tau) - i\Delta\phi_2(\tau)} + \text{c.c.} \right\rangle, \tag{2.121}$$

where the phase increments $\Delta\phi_{1,2}(\tau) = \phi_{1,2}(t+\tau) - \phi_{1,2}(t)$ are defined as before. We will also assume the dc components to be blocked (not to mention far enough away from the spectrum of interest that we can ignore them, since we have assumed Δ to be a sufficiently large radio frequency), so that

$$G_{\text{analyzer}}(\tau) \propto \frac{\eta_0^2 \eta_1 \eta_2 I_1 I_2}{4} \left\langle e^{-i\Delta\tau + i\Delta\phi_1(\tau) - i\Delta\phi_2(\tau)} + \text{c.c.} \right\rangle, \tag{2.122}$$

where $I_{1,2}$ are the (stationary) output intensities of the two lasers. We can again use the relation for Gaussian phase increments,

$$\left\langle e^{-i\Delta\phi(\tau)} \right\rangle = e^{-\langle [\Delta\phi(\tau)]^2 \rangle / 2}, \tag{2.123}$$

as in Eq. (2.105), so that

$$\begin{aligned} G_{\text{analyzer}}(\tau) &\propto \frac{\eta_0^2 \eta_1 \eta_2 I_1 I_2}{4} \left(e^{-i\Delta\tau} e^{-\langle [\Delta\phi_1(\tau) - \Delta\phi_2(\tau)]^2 \rangle / 2} + \text{c.c.} \right) \\ &= \frac{\eta_0^2 \eta_1 \eta_2 I_1 I_2}{4} \left(e^{-i\Delta\tau} e^{-\langle [\Delta\phi(\tau)]^2 \rangle} + \text{c.c.} \right) \\ &= \frac{\eta_0^2 \eta_1 \eta_2 I_1 I_2}{2} \cos\Delta\tau \, e^{-\langle [\Delta\phi(\tau)]^2 \rangle}, \end{aligned} \tag{2.124}$$

where we have used the independence of $\Delta\phi_1(\tau)$ and $\Delta\phi_2(\tau)$, and used the statistical identity to write $\Delta\phi_1(\tau) = \Delta\phi_2(\tau) =: \Delta\phi(\tau)$. Then the spectrum-analyzer signal is

$$S_{\text{analyzer}}(\omega) = \frac{4\eta_1\eta_2\eta_I^2 A_{\text{det}}^2 I_1 I_2}{\pi Z_{\text{in}}} \int_0^\infty d\tau \, \cos\Delta\tau \cos\omega\tau \exp\left[-\frac{4}{\pi} \int_0^\infty S_\omega(\omega') \frac{\sin^2(\omega'\tau/2)}{\omega'^2} \, d\omega' \right], \tag{2.125}$$

(heterodyne beat signal on spectrum analyzer)

where we have used Eq. (2.101) for the mean-square phase increment. Note that this spectrum has the same form as the spectrum of the optical signal in Eq. (2.107), except that ω_0 is replaced by Δ, and there is a factor of two in the exponent, meaning that the spectral part of the Fourier transform is squared. This means that the spectrum on the analyzer is the actual optical spectrum, but convolved with itself. This broadens the apparent line to a degree that depends on the precise form of the line shape.

We can rewrite the main result here as

$$S_{\text{analyzer}}(\nu) = \frac{8\eta_1\eta_2\eta_I^2 A_{\text{det}}^2 I_1 I_2}{Z_{\text{in}}} \int_0^\infty d\tau \, \cos 2\pi\delta\nu\tau \cos 2\pi\nu\tau \exp\left[-\frac{1}{\pi^3} \int_0^\infty S_\nu(\nu') \frac{\sin^2(\pi\nu'\tau)}{\nu'^2} \, d\nu' \right], \tag{2.126}$$

(heterodyne beat signal on spectrum analyzer)

in terms of the more experimentally relevant frequency ν. Here, we have rewritten the spectrum of frequency fluctuations as $S_\omega(\omega)\, d\omega = S_\nu(\nu)\, d\nu$, and the frequency difference as $\Delta = 2\pi\delta\nu$. In either case, we may also obtain the normalized forms of the (one-sided) spectra by dividing by $G_{\text{analyzer}}(\tau = 0) = 2\eta_1\eta_2\eta_I^2 A_{\text{det}}^2 I_1 I_2 / Z_{\text{in}}$ (again, after removing dc terms).

2.8.2.1 Example: White Noise in Heterodyne Spectroscopy

Returning to the white-noise example in Section 2.7.3.1,

$$S_\omega(\omega') = \gamma, \tag{2.127}$$

the exponent in Eq. (2.125) is

$$-\frac{4}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\,d\omega' = -\frac{4\gamma}{\pi}\int_0^\infty d\omega'\,\frac{\sin^2(\omega'\tau/2)}{\omega'^2} = -\gamma|\tau|. \tag{2.128}$$

This is the same result as in Section 2.7.3.1, with the replacement $\gamma \longrightarrow \gamma/2$. Then the normalized form of Eq. (2.125) is

$$s_{\text{analyzer}}(\omega) = \frac{2}{\pi}\int_0^\infty d\tau \cos\Delta\tau\cos\omega\tau\, e^{-\gamma|\tau|}, \tag{2.129}$$

carries through to the same result with the same rescaling of γ, so that

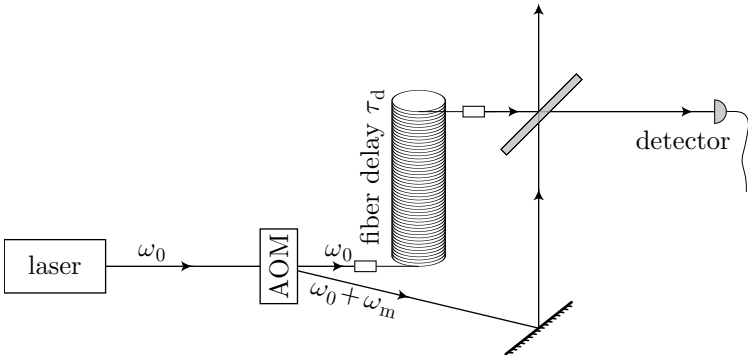
$$s_{\text{analyzer}}(\omega) = \frac{\gamma/\pi}{(\omega-\omega_0)^2+\gamma^2} + \frac{\gamma/\pi}{(\omega+\omega_0)^2+\gamma^2}, \tag{2.130}$$

(heterodyne spectrum, white frequency noise)

which is a properly normalized, one-sided Lorentzian spectrum with a full width at half maximum of 2γ. This is what we expect for a self-convolution of a Lorentzian distribution.

2.8.3 Self-Heterodyne Spectroscopy

Heterodyne spectroscopy is conceptually simple, but it may be that two identical copies of a laser are not available. An alternative approach in this case is to beat the output of a laser with a *time-delayed* and frequency-shifted version of itself.[5] If the delay is long compared to the coherence time, then this is something like a heterodyne of two independent sources. The setup is shown here; the beam is split by an acousto-optic modulator (AOM), with the zeroth order beam coupling into a fiber for a delay, and the first order being shifted by the AOM frequency ω_{m}.



The fiber is assumed to cause a well-defined time delay of the undiffracted beam, of around 5 μs for a 1 km fiber. (A fiber delay of the first-order beam just amounts to taking a negative τ_{d}.) For delays achievable in

[5]The results derived here were developed, e.g., in P. Gallion, F. J. Mendieta, and R. Leconte, ''Single-frequency laser phase-noise limitation in single-mode optical-fiber coherent-detection systems with correlated fields,'' *Journal of the Optical Society of America* **72**, 1167 (1982) (doi: 10.1364/JOSA.72.001167); Philippe B. Gallion and Guy Debarge, ''Quantum Phase Noise and Field Correlation in Single Frequency Semiconductor Laser Systems,'' *IEEE Journal of Quantum Electronics* **QE-20**, 343 (1984) (doi: 10.1109/JQE.1984.1072399); Linden B. Mercer, ''$1/f$ Frequency Noise Effects on Self-Heterodyne Linewidth Measurements,'' *Journal of Lightwave Technology* **9**, 485 (1991) (doi: 10.1109/50.76663); Hanne Ludvigsen, Mika Tossavainen, and Matti Kaivola, ''Laser linewidth measurements using self-homodyne detection with short delay,'' *Optics Communications* **155**, 180 (1998) (doi: 10.1016/S0030-4018(98)00355-1).

a fiber (∼km), though, there will in general be some residual level of correlation to deal with, which will be reflected in the spectrum. The combined field at the detector is

$$E^{(+)}_{\rm det}(t) = \sqrt{\eta_1}E^{(+)}(t) + \sqrt{\eta_2}E^{(+)}(t-\tau_{\rm d}), \tag{2.131}$$

where $\tau_{\rm d}$ is the delay due to the fiber (we will ignore the other small delays in the optical system). We will assume the undelayed field to be shifted in frequency by the acousto-optic modulator by an amount $\omega_{\rm m}$, and the delayed field to be unshifted. Thus, writing out the explicit field phases, we have

$$E^{(+)}_{\rm det}(t) = \sqrt{\eta_1}E^{(+)}_0 e^{-i(\omega_0+\omega_{\rm m})t-i\phi(t)} + \sqrt{\eta_2}E^{(+)}_0 e^{-i\omega_0(t-\tau_{\rm d})-i\phi(t-\tau_{\rm d})}. \tag{2.132}$$

Comparing to Eq. (2.119), we see that this problem is equivalent to the heterodyne case, but with the replacements $E^{(+)}_{01,02} \longrightarrow E^{(+)}_0$, $\phi_1(t) \longrightarrow \phi(t)$, $\phi_2(t) \longrightarrow \phi(t-\tau_{\rm d}) - i\omega_0\tau_{\rm d}$, $\omega_1 \longrightarrow \omega_0+\omega_{\rm m}$, and $\omega_2 \longrightarrow \omega_0$. Thus, adapting Eq. (2.122) for the correlation signal after the dc block, we have for the present case

$$G_{\rm analyzer}(\tau) \propto \frac{\eta_0^2\eta_1\eta_2 I_0^2}{4}\left\langle e^{i\omega_{\rm m}\tau+i[\phi(t+\tau)-\phi(t)]-i[\phi(t-\tau_{\rm d}+\tau)-\phi(t-\tau_{\rm d})]} + \text{c.c.}\right\rangle. \tag{2.133}$$

Again using Eq. (2.123), we have

$$G_{\rm analyzer}(\tau) \propto \frac{\eta_0^2\eta_1\eta_2 I_0^2}{4}\left(e^{i\omega_{\rm m}\tau}e^{-\langle\{[\phi(t+\tau)-\phi(t)]-[\phi(t-\tau_{\rm d}+\tau)-\phi(t-\tau_{\rm d})]\}^2\rangle/2} + \text{c.c.}\right). \tag{2.134}$$

The phase expectation value can then be transformed (Problem 2.6) so that

$$\begin{aligned} G_{\rm analyzer}(\tau) &\propto \frac{\eta_0^2\eta_1\eta_2 I_0^2}{4}\left(e^{i\omega_{\rm m}\tau}e^{-\langle[\Delta\phi(\tau)]^2\rangle-\langle[\Delta\phi(\tau_{\rm d})]^2\rangle+\langle[\Delta\phi(\tau+\tau_{\rm d})]^2\rangle/2+\langle[\Delta\phi(\tau-\tau_{\rm d})]^2\rangle/2} + \text{c.c.}\right) \\ &= \frac{\eta_0^2\eta_1\eta_2 I_0^2}{2}\cos\omega_{\rm m}\tau\, e^{-\langle[\Delta\phi(\tau)]^2\rangle-\langle[\Delta\phi(\tau_{\rm d})]^2\rangle+\langle[\Delta\phi(\tau+\tau_{\rm d})]^2\rangle/2+\langle[\Delta\phi(\tau-\tau_{\rm d})]^2\rangle/2}. \end{aligned} \tag{2.135}$$

Using Eq. (2.101), we then find (see Problem 2.7)

$$G_{\rm analyzer}(\tau) \propto \frac{\eta_0^2\eta_1\eta_2 I_0^2}{2}\cos\omega_{\rm m}\tau \exp\left[-\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_{\rm d}/2)}{\omega'^2}\,d\omega'\right]. \tag{2.136}$$

Thus, the spectrum (2.116) is

$$S_{\rm analyzer}(\omega) = \frac{4\eta_1\eta_2\eta_I^2 A_{\rm det}^2 I_0^2}{\pi Z_{\rm in}}\int_0^\infty d\tau\,\cos\omega_{\rm m}\tau\cos\omega\tau\exp\left[-\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_{\rm d}/2)}{\omega'^2}\,d\omega'\right], \tag{2.137}$$

(self-heterodyne beat signal on spectrum analyzer)

which is essentially the same as the heterodyne signal (2.125), if we identify $\omega_{\rm m} = \Delta$ and note the extra factor of $2\sin^2(\omega\tau_{\rm d}/2)$ in the integrand, due to the extra coherence in beating the signal with a time-delayed version of itself. Note that this factor reduces to unity if it is replaced by its average value, which we expect if this coherence is lost. For example, in the limit of large $\tau_{\rm d}$, this factor is a rapidly oscillating function of ω. So long as the oscillations are rapid on the scale of the structure of the spectrum and of the relevant values of τ (i.e., $\tau_{\rm d}$ is much larger than the coherence time of the signal), it is a good approximation to replace this factor by unity, so that this expression collapses to the pure heterodyne result. Finally, we can again write the main result as

$$S_{\rm analyzer}(\nu) = \frac{8\eta_1\eta_2\eta_I^2 A_{\rm det}^2 I_0^2}{Z_{\rm in}}\int_0^\infty d\tau\,\cos 2\pi\nu_{\rm m}\tau\cos 2\pi\nu\tau\exp\left[-\frac{2}{\pi^3}\int_0^\infty S_\nu(\nu')\frac{\sin^2(\pi\nu'\tau)\sin^2(\pi\nu'\tau_{\rm d})}{\nu'^2}\,d\nu'\right], \tag{2.138}$$

(self-heterodyne beat signal on spectrum analyzer)

in terms of the more experimentally relevant frequency $\nu = \omega/2\pi$.

2.8.3.1 Example: White Noise in Self-Heterodyne Spectroscopy

Returning once again to the white-noise example in Section 2.7.3.1,

$$S_\omega(\omega') = \gamma, \tag{2.139}$$

the exponent in Eq. (2.137) is

$$\begin{aligned} -\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_\mathrm{d}/2)}{\omega'^2}\,d\omega' &= -\frac{8\gamma}{\pi}\int_0^\infty d\omega'\,\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_\mathrm{d}/2)}{\omega'^2} \\ &= -\gamma\left(|\tau| + |\tau_\mathrm{d}| - \frac{|\tau-\tau_\mathrm{d}|}{2} - \frac{|\tau+\tau_\mathrm{d}|}{2}\right). \end{aligned} \tag{2.140}$$

which follows most easily by comparing Eq. (2.135) with Eq. (2.124), noting that we adapt the result simply by making several time-offset copies of the heterodyne-exponent result. Then the normalized form of Eq. (2.137) is

$$s_\mathrm{analyzer}(\omega) = \frac{2}{\pi}\int_0^\infty d\tau\cos\omega_\mathrm{m}\tau\cos\omega\tau\,\exp\left[-\gamma\left(|\tau| + |\tau_\mathrm{d}| - \frac{|\tau-\tau_\mathrm{d}|}{2} - \frac{|\tau+\tau_\mathrm{d}|}{2}\right)\right]. \tag{2.141}$$

Now since $\tau \geq 0$ and the exponent is an even function of τ_d, we can write

$$\tau + |\tau_\mathrm{d}| - \frac{|\tau-\tau_\mathrm{d}|}{2} - \frac{|\tau+\tau_\mathrm{d}|}{2} = \begin{cases} \tau + |\tau_\mathrm{d}| - (|\tau_\mathrm{d}| - \tau)/2 - (\tau + |\tau_\mathrm{d}|)/2 = \tau & (\tau < |\tau_\mathrm{d}|) \\ \tau + |\tau_\mathrm{d}| - (\tau - |\tau_\mathrm{d}|)/2 - (\tau + |\tau_\mathrm{d}|)/2 = |\tau_\mathrm{d}| & (\tau > |\tau_\mathrm{d}|). \end{cases} \tag{2.142}$$

Thus,

$$\begin{aligned} s_\mathrm{analyzer}(\omega) &= \frac{2}{\pi}\int_0^{|\tau_\mathrm{d}|} d\tau\cos\omega_\mathrm{m}\tau\cos\omega\tau\,e^{-\gamma\tau} + \frac{2}{\pi}e^{-\gamma|\tau_\mathrm{d}|}\int_{|\tau_\mathrm{d}|}^\infty d\tau\cos\omega_\mathrm{m}\tau\cos\omega\tau \\ &= \frac{2}{\pi}\int_0^{|\tau_\mathrm{d}|} d\tau\cos\omega_\mathrm{m}\tau\cos\omega\tau\left(e^{-\gamma\tau} - e^{-\gamma|\tau_\mathrm{d}|}\right) + \frac{2}{\pi}e^{-\gamma|\tau_\mathrm{d}|}\int_0^\infty d\tau\cos\omega_\mathrm{m}\tau\cos\omega\tau \\ &= \frac{1}{\pi}\int_0^{\tau_\mathrm{d}} d\tau\left[\cos(\omega-\omega_\mathrm{m})\tau + \cos(\omega+\omega_\mathrm{m})\tau\right]\left(e^{-\gamma\tau} - e^{-\gamma|\tau_\mathrm{d}|}\right) \\ &\qquad + \frac{1}{\pi}e^{-\gamma|\tau_\mathrm{d}|}\int_0^\infty d\tau\left[\cos(\omega-\omega_\mathrm{m})\tau + \cos(\omega+\omega_\mathrm{m})\tau\right] \\ &= \frac{1}{\pi}\int_0^{\tau_\mathrm{d}} d\tau\cos(\omega-\omega_\mathrm{m})\tau\left(e^{-\gamma\tau} - e^{-\gamma|\tau_\mathrm{d}|}\right) + \frac{1}{\pi}e^{-\gamma|\tau_\mathrm{d}|}\int_0^\infty d\tau\cos(\omega-\omega_\mathrm{m})\tau \\ &\qquad + (\omega_\mathrm{m} \longrightarrow -\omega_\mathrm{m}). \end{aligned} \tag{2.143}$$

Then using the integral representation of the delta function for the second integral and evaluating the first integral,

$$\begin{aligned}
s_{\text{analyzer}}(\omega) &= \left[\frac{1}{2\pi}\int_0^{|\tau_\text{d}|} d\tau \left(e^{i(\omega-\omega_\text{m})\tau-\gamma\tau} - e^{i(\omega-\omega_\text{m})\tau-\gamma|\tau_\text{d}|}\right) + \text{c.c.}\right] \\
&\qquad + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) + (\omega_\text{m} \longrightarrow -\omega_\text{m}) \\
&= \frac{1}{2\pi}\left[\frac{e^{i(\omega-\omega_\text{m})|\tau_\text{d}|-\gamma|\tau_\text{d}|}-1}{i(\omega-\omega_\text{m})-\gamma} - \frac{\left(e^{i(\omega-\omega_\text{m})|\tau_\text{d}|}-1\right)e^{-\gamma|\tau_\text{d}|}}{i(\omega-\omega_\text{m})} + \text{c.c.}\right] \\
&\qquad + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) + (\omega_\text{m} \longrightarrow -\omega_\text{m}) \\
&= \frac{\gamma/\pi}{(\omega-\omega_\text{m})^2+\gamma^2} + \frac{1}{2\pi}\left[\frac{e^{i(\omega-\omega_\text{m})|\tau_\text{d}|-\gamma|\tau_\text{d}|}}{i(\omega-\omega_\text{m})-\gamma} - \frac{e^{i(\omega-\omega_\text{m})|\tau_\text{d}|-\gamma|\tau_\text{d}|}}{i(\omega-\omega_\text{m})} + \text{c.c.}\right] \\
&\qquad + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) + (\omega_\text{m} \longrightarrow -\omega_\text{m}) \\
&= \frac{\gamma/\pi}{(\omega-\omega_\text{m})^2+\gamma^2} + \frac{e^{-\gamma|\tau_\text{d}|}}{2\pi}\left[\frac{e^{i(\omega-\omega_\text{m})|\tau_\text{d}|}}{i(\omega-\omega_\text{m})-\gamma} + \text{c.c.}\right] - \frac{e^{-\gamma|\tau_\text{d}|}\sin[(\omega-\omega_\text{m})|\tau_\text{d}|]}{\pi(\omega-\omega_\text{m})} \\
&\qquad + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) + (\omega_\text{m} \longrightarrow -\omega_\text{m}) \\
&= \frac{\gamma/\pi}{(\omega-\omega_\text{m})^2+\gamma^2}\left[1 - e^{-\gamma|\tau_\text{d}|}\left(\cos[(\omega-\omega_\text{m})|\tau_\text{d}|] - \frac{(\omega-\omega_\text{m})}{\gamma}\sin[(\omega-\omega_\text{m})|\tau_\text{d}|]\right)\right] \\
&\qquad - \frac{e^{-\gamma|\tau_\text{d}|}\sin[(\omega-\omega_\text{m})|\tau_\text{d}|]}{\pi(\omega-\omega_\text{m})} + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) \\
&\qquad + (\omega_\text{m} \longrightarrow -\omega_\text{m}),
\end{aligned} \tag{2.144}$$

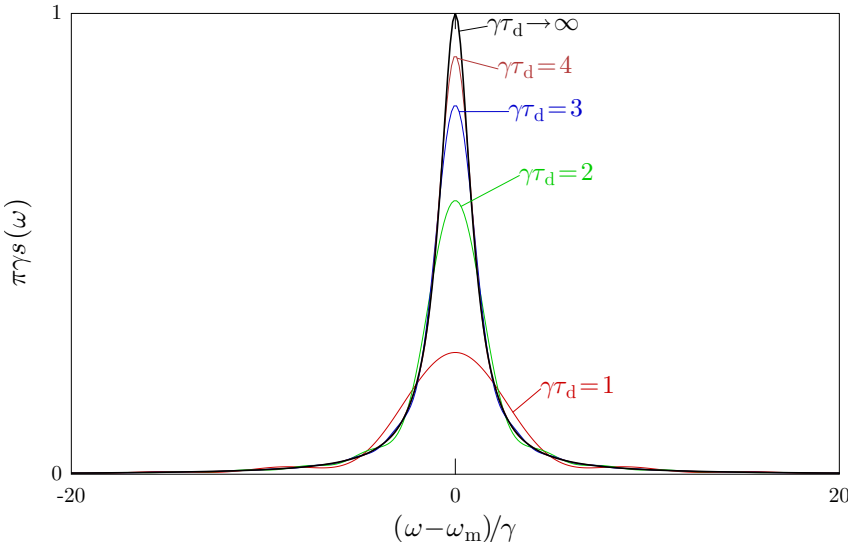
and so

$$\begin{aligned}
s_{\text{analyzer}}(\omega) &= \frac{\gamma/\pi}{(\omega-\omega_\text{m})^2+\gamma^2}\left[1 - e^{-\gamma|\tau_\text{d}|}\left(\cos[(\omega-\omega_\text{m})|\tau_\text{d}|] + \frac{\gamma}{(\omega-\omega_\text{m})}\sin[(\omega-\omega_\text{m})|\tau_\text{d}|]\right)\right] \\
&\qquad + e^{-\gamma|\tau_\text{d}|}\delta(\omega-\omega_\text{m}) \\
&\qquad + (\omega_\text{m} \longrightarrow -\omega_\text{m}).
\end{aligned} \tag{2.145}$$

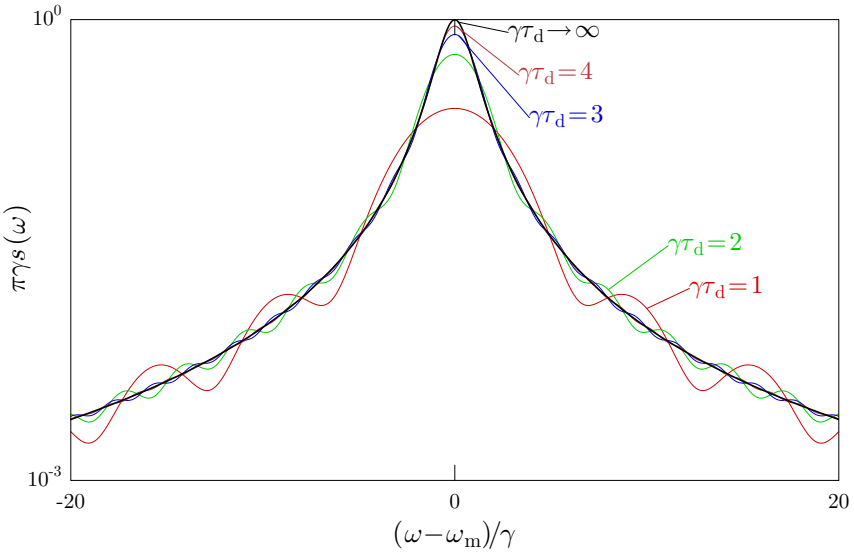
(self-heterodyne spectrum, white frequency noise)

The first term here is a modified Lorentzian spectrum, where oscillations in frequency are superimposed on the Lorentzian envelope. These oscillations become finer and smaller with increasing $|\tau_\text{d}|$, and vanish in the limit $|\tau_\text{d}| \longrightarrow \infty$, when we recover the heterodyne result (Lorentzian with a full width at half maximum of 2γ). The other terms likewise vanish in this limit. The delta-function term is the most obvious result of residual correlations between the optical signal and its time-delayed copy. Recalling that the correlation goes away exponentially with the delay, this is equivalent to an ensemble of optical signals that are exactly monochromatic and phase coherent, except for a phase jump at a time τ to a completely random phase, where τ is a random time with exponential probability distribution. Then the light tends to correlate perfectly at short time delays, and not at all for long ones; the exponential function here is just what we expect in the ensemble average, so the correlation decays away exponentially on the time scale of the coherence time.

The spectrum is plotted below for several coherence times, including the heterodyne ($\tau_\text{d} \longrightarrow \infty$) Lorentzian limit. The oscillations in the tails are only slightly visible, but what is clearly evident is a strong impact on the apparent width of the line.



On a logarithmic vertical scale, the effects near the line center are mitigated, but the oscillations in the tails are more apparent.



Both plots were made under the assumption of a narrow line, $\gamma \gg \omega_m$, so that we can ignore the contribution of the component centered at $-\omega_m$ (γ should also implicitly be small compared to the optical frequency).

2.8.3.2 Calculation of General Self-Heterodyne Spectra

In both the heterodyne and self-heterodyne setups, the Fourier-tranform integral to obtain the measured spectrum cannot be done analytically for arbitrary frequency-noise spectra, and thus they must be carried out numerically. For narrow spectral (laser) lines, the width of the line is much smaller than the optical frequency, and thus we can ignore any counterrotating terms in the spectrum (that is, any terms in the self-heterodyne spectrum centered about $-\omega_m$). For example, we can then rewrite the (normalized) self-heterodyne spectrum (2.137) as

$$\begin{aligned} s_{\text{analyzer}}(\omega) &= \frac{1}{\pi}\int_0^\infty d\tau\, \cos[(\omega-\omega_m)\tau]\exp\left[-\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_d/2)}{\omega'^2}\,d\omega'\right] \\ &= \frac{1}{2\pi}\int_{-\infty}^\infty d\tau\, e^{i\Delta\tau}\exp\left[-\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_d/2)}{\omega'^2}\,d\omega'\right], \end{aligned} \tag{2.146}$$

where $\Delta := \omega - \omega_{\rm m}$. In general, the mean-square phase integral in the exponent must be computed (analytically, in the ideal case), and then the resulting exponential function must be transformed in the remaining integral, which is just a Fourier transform, giving a centered spectrum about $\omega = \omega_{\rm m}$.

2.8.3.3 Self-Heterodyne Spectrum of $1/f$ Noise

As an example, consider a frequency-noise spectrum consisting of white-noise and $1/f$-noise components:

$$S_\omega(\omega) = \gamma + \frac{k}{\omega}.$$

(model for semiconductor-laser frequency noise) (2.147)

This spectrum can accurately model phase noise in semiconductor lasers, for example.[6] Let's start by calculating the mean-square phase increment, from Eq. (2.101):

$$\begin{aligned}\left\langle [\Delta\phi(\tau)]^2 \right\rangle &= \frac{4}{\pi}\int_0^\infty S_\omega(\omega)\frac{\sin^2(\omega\tau/2)}{\omega^2}\,d\omega \\ &= \frac{4}{\pi}\int_0^\infty \left(\gamma + \frac{k}{\omega}\right)\frac{\sin^2(\omega\tau/2)}{\omega^2}\,d\omega.\end{aligned} \tag{2.148}$$

We have already seen [Eq. (2.109)] that the white-noise part leads to $\gamma|\tau|$. Now we still need to evaluate the $1/f$ part, however:

$$\left\langle [\Delta\phi(\tau)]^2 \right\rangle = \gamma|\tau| + \frac{4k}{\pi}\int_0^\infty d\omega\,\frac{\sin^2(\omega\tau/2)}{\omega^3}. \tag{2.149}$$

Unfortunately, though, this integral has a $1/\omega$ divergence at $\omega = 0$. This essentially means we have non-Gaussian statistics, and our assumption of Gaussian noise has broken down. Nevertheless, we will proceed with the self-homodyne spectrum, which involves an extra factor of $\sin^2(\omega\tau_{\rm d}/2)$ and thus has no divergence. Recall from Eq. (2.135) that the presence of this extra factor is equivalent to considering the combination

$$\left\langle [\Delta\phi(\tau)]^2 \right\rangle + \left\langle [\Delta\phi(\tau_{\rm d})]^2 \right\rangle - \frac{\left\langle [\Delta\phi(\tau+\tau_{\rm d})]^2 \right\rangle}{2} - \frac{\left\langle [\Delta\phi(\tau-\tau_{\rm d})]^2 \right\rangle}{2}. \tag{2.150}$$

Note in particular that the delta function we saw in the Lorentzian spectrum is a general feature of self-heterodyne spectra: as $\tau \longrightarrow \infty$, this combination tends to reduce to $\left\langle [\Delta\phi(\tau_{\rm d})]^2 \right\rangle$, which is a constant offset that yields a delta function at $\omega = \omega_{\rm m}$ in the Fourier transform.

In the self-heterodyne case, the integral in the exponential can be carried out, with the result

$$\begin{aligned}&\frac{8}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)\sin^2(\omega'\tau_{\rm d}/2)}{\omega'^2}\,d\omega' \\ &\qquad = \gamma\left(|\tau| + |\tau_{\rm d}| - \frac{|\tau-\tau_{\rm d}|}{2} - \frac{|\tau+\tau_{\rm d}|}{2}\right) \\ &\qquad\quad - \frac{k}{4\pi}\left[2\tau^2\log\tau^2 + 2\tau_{\rm d}^2\log\tau_{\rm d}^2 - (\tau+\tau_{\rm d})^2\log(\tau+\tau_{\rm d})^2 - (\tau-\tau_{\rm d})^2\log(\tau-\tau_{\rm d})^2\right].\end{aligned} \tag{2.151}$$

[6] Linden B. Mercer, "$1/f$ Frequency Noise Effects on Self-Heterodyne Linewidth Measurements," *Journal of Lightwave Technology* **9**, 485 (1991) (doi: 10.1109/50.76663).

Here, we have used

$$\begin{aligned}\int_x^\infty d\omega\,\frac{\sin^2(\omega\tau/2)}{\omega^3} &= \frac{1}{2}\int_x^\infty d\omega\,\frac{1-\cos\omega\tau}{\omega^3}\\ &= \frac{1}{4x^2} - \frac{\tau^2}{2}\int_{x\tau}^\infty d\omega\,\frac{\cos\omega}{\omega^3}\\ &= \frac{1}{4x^2} - \frac{\cos x\tau}{4x^2} + \frac{\tau^2}{2}\int_{x\tau}^\infty d\omega\,\frac{\sin\omega}{\omega^2}\\ &= \frac{1}{4x^2} - \frac{\cos x\tau}{4x^2} + \frac{\tau\sin x\tau}{4x} + \frac{\tau^2}{4}\int_{x\tau}^\infty d\omega\,\frac{\cos\omega}{\omega}\\ &= \frac{1}{4x^2} - \frac{\cos x\tau}{4x^2} + \frac{\tau\sin x\tau}{4x} - \frac{\tau^2}{4}\mathrm{Ci}(x\tau)\\ &= c\tau^2 - \frac{\tau^2}{4}\log x\tau + O(x^2),\end{aligned} \tag{2.152}$$

where $\mathrm{Ci}(x)$ is the cosine integral [see Eq. (13.27)], and c is a constant whose value is unimportant. In the combination of terms we have in (2.150), both the c terms and the x-dependence of the logarithms cancel, avoiding any dc divergences. Putting the exponent (2.151) into Eq. (2.146), we have

$$s_{\mathrm{analyzer}}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} e^{-\gamma\min\{|\tau|,|\tau_\mathrm{d}|\}} |\tau|^{k\tau^2/\pi} |\tau_\mathrm{d}|^{k\tau_\mathrm{d}^2/\pi} |\tau+\tau_\mathrm{d}|^{-k(\tau+\tau_\mathrm{d})^2/2\pi} |\tau-\tau_\mathrm{d}|^{-k(\tau-\tau_\mathrm{d})^2/2\pi}.$$

(self heterodyne signal for white and $1/f$ noise) (2.153)

This is a Fourier transform, which is straightforward to calculate numerically. Note again that as $\tau \longrightarrow \infty$, the correlation function becomes the constant value $e^{-\gamma\tau_\mathrm{d}}|\tau_\mathrm{d}|^{k\tau_\mathrm{d}^2/\pi}$. It may be convenient to subtract this away before the Fourier transform to avoid dealing with the delta function in the spectrum. But note that this amplitude diverges with τ_d—the heterodyne limit is problematic, as we have already discussed. Physically, the $1/f$ spectrum must be cut off at low frequencies, as we will discuss in the next section. An alternative strategy for handling the delta function is to include an explicit cutoff of the tail of the integrand:

$$\begin{aligned}&s_{\mathrm{analyzer}}(\omega)\\ &\quad= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} e^{-\gamma\min\{|\tau|,|\tau_\mathrm{d}|\}-\delta\omega_\mathrm{RB}\tau/2} |\tau|^{k\tau^2/\pi} |\tau_\mathrm{d}|^{k\tau_\mathrm{d}^2/\pi} |\tau+\tau_\mathrm{d}|^{-k(\tau+\tau_\mathrm{d})^2/2\pi} |\tau-\tau_\mathrm{d}|^{-k(\tau-\tau_\mathrm{d})^2/2\pi}.\end{aligned}$$

(self heterodyne signal for white and $1/f$ noise, with Lorentzian resolution bandwidth) (2.154)

Here, $\delta\omega_\mathrm{RB}$ is the resolution bandwidth (full width at half maximum) of the spectrum analyzer, assuming a Lorentzian bandpass filter. For a Gaussian filter shape to model the resolution bandwidth, we instead have

$$\begin{aligned}&s_{\mathrm{analyzer}}(\omega)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} e^{-\gamma\min\{|\tau|,|\tau_\mathrm{d}|\}-(\delta\omega_\mathrm{RB}\tau)^2/16\log 2} |\tau|^{k\tau^2/\pi} |\tau_\mathrm{d}|^{k\tau_\mathrm{d}^2/\pi} |\tau+\tau_\mathrm{d}|^{-k(\tau+\tau_\mathrm{d})^2/2\pi} |\tau-\tau_\mathrm{d}|^{-k(\tau-\tau_\mathrm{d})^2/2\pi},\end{aligned}$$

(self heterodyne signal for white and $1/f$ noise, with Gaussian resolution bandwidth) (2.155)

where now $\delta\omega_\mathrm{RB}$ is the Gaussian resolution bandwidth (full width at half maximum). In the frequency domain, these cutoffs convolve the spectrum with a Lorentzian or Gaussian, respectively, of width $\delta\omega_\mathrm{RB}$, eliminating the delta function and most closely emulating the results of a physical measurement.

2.8.3.4 Observation Time and Linewidth of $1/f$ Noise

For $1/f$ noise, we have seen that the self-heterodyne spectrum can be calculated, but there is a divergence in the heterodyne limit. In fact, the width of the self-heterodyne line increases with the delay time τ_d. Effectively, this is the time over which the self-heterodyne apparatus monitors the frequency fluctuations of the laser, and this is why the $1/f$ divergence is cut off in the self-heterodyne spectrum (2.137), where there is an extra factor of $\sin^2(\omega'\tau_\mathrm{d}/2)$ in the exponent, compared with the heterodyne spectrum (2.125). The point

is that over any finite observation time $T_{\rm obs}$, the lowest (problematic) frequencies in the $1/f$ spectrum cannot contribute; roughly speaking, only frequencies *above* the cutoff frequency $\omega_{\rm c} = 2\pi/T_{\rm obs}$ can contribute to the observed line width.

To understand the problem more specifically here, consider again the case of white frequency noise,

$$S_\omega(\omega) = \gamma. \tag{2.156}$$

Recall that this noise causes the frequency of the laser to fluctuate about its center value ω_0 with a Lorentzian distribution. But recall that the spectrum of *phase* fluctuations diverges at zero frequency,

$$S_\phi(\omega) = \frac{\gamma}{\omega^2}. \tag{2.157}$$

The phase, being the integral of the frequency, diffuses in a random walk, and does not remain near a particular phase, and is thus not a **stationary noise process**. The arbitrarily large dc noise essentially means that the time-averaged phase is not a well-defined quantity. In the same way, for $1/f$ noise, the divergence indicates that the time-averaged *frequency* is not a well-defined quantity: the center frequency itself can wander, and it wanders farther the longer it is observed. Physically, this is the case for lasers whose frequencies can drift over long times due to pressure and temperature fluctuations, relaxing material stresses, and mechanical vibration and creep. Thus, it may not be surprising that the spectral linewidth of a laser depends on the time scale over which it is observed, and we are seeing that this is the case when there is a $1/f$ component to the frequency-noise spectrum.

To account for the time of observation,[7] we must revise the calculation of Section 2.7.2 of the variance of the phase fluctuations. There, we computed the mean-square phase fluctuation $\langle[\Delta\phi(\tau)]^2\rangle$, with the measurement (time average) taken over all times (this is equivalent to the variance since $\langle\Delta\phi(\tau)\rangle = 0$). Here, we will take the angle brackets to denote a time average taken over a finite time T. We will also now explicitly compute the variance, since $\langle\Delta\phi(\tau)\rangle$ is not necessarily zero when the average is taken over a finite time interval:

$$\begin{aligned}\mathrm{Var}\,[\Delta\phi(\tau)]_t &= \Big\langle [\Delta\phi(t,t+\tau)]^2 \Big\rangle_T - \langle\Delta\phi(t,t+\tau)\rangle_T^2 \\ &= \frac{1}{T}\int_{-T/2}^{T/2} dt\ [\Delta\phi(t,t+\tau)]^2 - \left(\frac{1}{T}\int_{-T/2}^{T/2} dt\,\Delta\phi(t,t+\tau)\right)^2.\end{aligned} \tag{2.158}$$

Here, we are using the more general notation $\Delta\phi(t,t+\tau) := \phi(t+\tau) - \phi(t)$ for the phase increment, and the subscript t indicates that the time-averaging interval (measurement interval) is centered around time t. The subscript T on the angle brackets denote the finite-time average, which we will write out explicitly below (the absence of this subscript still denotes the limit $T \longrightarrow \infty$). However, this finite-time average should be averaged over all time (or equivalently, averaged over the ensemble of all possible noise realizations), to obtain the finite-time statistical variance:

$$\begin{aligned}\mathrm{Var}\,[\Delta\phi(\tau)] &= \frac{1}{T}\left\langle \int_{t-T/2}^{t+T/2} dt'\ [\Delta\phi(t',t'+\tau)]^2 \right\rangle - \frac{1}{T^2}\left\langle \left(\int_{t-T/2}^{t+T/2} dt'\,\Delta\phi(t',t'+\tau)\right)^2 \right\rangle \\ &= \frac{1}{T}\int_{t-T/2}^{t+T/2} dt' \left\langle [\Delta\phi(t,t+\tau)]^2 \right\rangle - \frac{1}{T^2}\left\langle \int_{t-T/2}^{t+T/2} dt' \int_{t-T/2}^{t+T/2} dt''\,\Delta\phi(t',t'+\tau)\Delta\phi(t'',t''+\tau) \right\rangle \\ &= \Big\langle [\Delta\phi(\tau)]^2 \Big\rangle - \frac{1}{T^2}\left\langle \int_{-\infty}^{\infty} dt' \int_{-\infty}^{\infty} dt''\, f_T(t'-t)\Delta\phi(t',t'+\tau)\Delta\phi(t'',t''+\tau) f_T(t''-t) \right\rangle.\end{aligned} \tag{2.159}$$

Here, we have defined $f_T(t)$ to be the unit-pulse function of duration T (i.e., $f_T(t) = 1$ for t between $-T/2$ and $T/2$, and $f_T(t) = 0$ otherwise). Thus, $f_T(t)/T$ is a unit-area pulse function. Computing the Fourier

[7] L. S. Cutler and C. L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," *Proceedings of the IEEE* **54**, 136 (1966) (doi: 10.1109/PROC.1966.4627).

transform of this pulse function, we find

$$\begin{aligned}\int_{-\infty}^{\infty} dt'\,\frac{f_T(t'-t)}{T}e^{i\omega t'} &= \frac{1}{T}\int_{t-T/2}^{t+T/2} dt'\,e^{i\omega t'} \\ &= \frac{1}{i\omega T}\left(e^{i\omega(t+T/2)} - e^{i\omega(t-T/2)}\right) \\ &= \frac{e^{i\omega t}\sin(\omega T/2)}{\omega T/2} = e^{i\omega t}\,\mathrm{sinc}(\omega T/2),\end{aligned} \tag{2.160}$$

where $\mathrm{sinc}\,x := (\sin x)/x$. Then inverting the Fourier transform, we find

$$\frac{f_T(t'-t)}{T} = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega(t'-t)}\,\mathrm{sinc}(\omega T/2), \tag{2.161}$$

Then we can use this result twice in Eq. (2.159):

$$\begin{aligned}\mathrm{Var}\,[\Delta\phi(\tau)] &= \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{(2\pi)^2}\left\langle \int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}dt''\,\Delta\phi(t',t'+\tau)\Delta\phi(t'',t''+\tau)\right. \\ &\qquad\left.\times\, e^{-i\omega(t'-t)}\mathrm{sinc}(\omega T/2)\,e^{-i\omega'(t''-t)}\mathrm{sinc}(\omega' T/2)\right\rangle \\ &= \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{(2\pi)^2}\lim_{T'\to\infty}\frac{1}{T'}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}dt''\int_{-T'/2}^{T'/2}dt\,\Delta\phi(t',t'+\tau) \\ &\qquad\times\,\Delta\phi(t'',t''+\tau)\,e^{-i\omega(t'-t)}\mathrm{sinc}(\omega T/2)\,e^{-i\omega'(t''-t)}\mathrm{sinc}(\omega' T/2) \\ &= \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{(2\pi)^2}\lim_{T'\to\infty}\frac{1}{T'}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\int_{-T'/2}^{T'/2}dt'\int_{-\infty}^{\infty}dt''\int_{-\infty}^{\infty}dt\,\Delta\phi(t',t'+\tau) \\ &\qquad\times\,\Delta\phi(t'',t''+\tau)\,e^{i(\omega+\omega')t}\,e^{-i\omega t'}\,e^{-i\omega' t''}\,\mathrm{sinc}(\omega T/2)\,\mathrm{sinc}(\omega' T/2).\end{aligned} \tag{2.162}$$

We now get a factor of $2\pi\delta(\omega+\omega')$ from the t integration, which takes care of the ω' integral:

$$\begin{aligned}\mathrm{Var}\,[\Delta\phi(\tau)] &= \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{2\pi}\lim_{T'\to\infty}\frac{1}{T'}\int_{-\infty}^{\infty}d\omega\int_{-T'/2}^{T'/2}dt'\int_{-\infty}^{\infty}dt''\,\Delta\phi(t',t'+\tau)\Delta\phi(t'',t''+\tau) \\ &\qquad\times\, e^{-i\omega(t'-t'')}\,\mathrm{sinc}^2(\omega T/2) \\ &= \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\lim_{T'\to\infty}\left|\frac{1}{T'}\int_{-T'/2}^{T'/2}dt'\,\Delta\phi(t',t'+\tau)\,e^{-i\omega t'}\right|^2\mathrm{sinc}^2(\omega T/2).\end{aligned} \tag{2.163}$$

Now we use the Wiener–Khinchin theorem in the form of Eq. (2.20),

$$\mathrm{Var}\,[\Delta\phi(\tau)] = \left\langle [\Delta\phi(\tau)]^2\right\rangle - \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\left[\int_{-\infty}^{\infty}dt'\langle\Delta\phi(t,t+\tau)\Delta\phi(t+t',t+t'+\tau)\rangle\,e^{i\omega t'}\right]\mathrm{sinc}^2(\omega T/2), \tag{2.164}$$

where $S_{\Delta\phi}(\omega)$ is the (one-sided) power spectral density corresponding to the signal $\Delta\phi(\tau)$. We can work out the phase expectation value here as

$$\begin{aligned}\langle\Delta\phi(t,t+\tau)\Delta\phi(t+t',t+t'+\tau)\rangle &= 2\langle\phi(t)\phi(t+t')\rangle - \langle\phi(t)\phi(t+t'+\tau)\rangle - \langle\phi(t)\phi(t-t'+\tau)\rangle \\ &= \frac{1}{\pi}\int_0^{\infty}S_\omega(\omega)\frac{2\cos\omega t' - \cos[\omega(\tau+t')] - \cos[\omega(\tau-t')]]}{\omega^2}\,d\omega \\ &= \frac{2}{\pi}\int_0^{\infty}S_\omega(\omega)\frac{\cos\omega t' - \cos\omega t'\cos\omega\tau}{\omega^2}\,d\omega \\ &= \frac{4}{\pi}\int_0^{\infty}S_\omega(\omega)\frac{\cos\omega t'\sin^2(\omega\tau/2)}{\omega^2}\,d\omega \\ &= \frac{2}{\pi}\int_{-\infty}^{\infty}S_\omega(|\omega|)\frac{\sin^2(\omega\tau/2)}{\omega^2}e^{-i\omega t'}\,d\omega,\end{aligned} \tag{2.165}$$

where we have used Eq. (2.98) to evaluate the correlation functions. Then the variance becomes

$$\begin{aligned}\mathrm{Var}\left[\Delta\phi(\tau)\right] &= \left\langle\left[\Delta\phi(\tau)\right]^2\right\rangle - \frac{1}{\pi^2}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}dt'\int_{-\infty}^{\infty}d\omega' S_\omega(|\omega'|)\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\,e^{i(\omega-\omega')t'}\,\mathrm{sinc}^2(\omega T/2)\\ &= \left\langle\left[\Delta\phi(\tau)\right]^2\right\rangle - \frac{2}{\pi}\int_{-\infty}^{\infty}d\omega\, S_\omega(|\omega|)\frac{\sin^2(\omega\tau/2)}{\omega^2}\mathrm{sinc}^2(\omega T/2)\\ &= \left\langle\left[\Delta\phi(\tau)\right]^2\right\rangle - \frac{4}{\pi}\int_{0}^{\infty}d\omega\, S_\omega(\omega)\frac{\sin^2(\omega\tau/2)}{\omega^2}\mathrm{sinc}^2(\omega T/2).\end{aligned} \tag{2.166}$$

Finally, using Eq. (2.101) for the infinite-time variance $\langle[\Delta\phi(\tau)]^2\rangle$, we obtain

$$\mathrm{Var}\left[\Delta\phi(\tau)\right]_T = \frac{4}{\pi}\int_0^\infty d\omega\, S_\omega(\omega)\frac{\sin^2(\omega\tau/2)}{\omega^2}\left[1-\mathrm{sinc}^2\left(\frac{\omega T}{2}\right)\right] \tag{2.167}$$

as its generalization for finite observation times.

Now we must be a bit more careful in interpreting the observation time. In an observation time T_{obs}, the field is measured over this time interval and then used to construct the correlation function (and thus the spectrum). To construct the correlation function at delay τ, only a time of $T_{\mathrm{obs}}-|\tau|$ is actually useable in the correlation-function time average, and delays $|\tau|>T_{\mathrm{obs}}$ are nonsensical. Thus, we should take $T=T_{\mathrm{obs}}-\tau$ in Eq. (2.167), with the rsult

$$\mathrm{Var}\left[\Delta\phi(\tau)\right]_{T_{\mathrm{obs}}} = \frac{4}{\pi}\int_0^\infty d\omega\, S_\omega(\omega)\frac{\sin^2(\omega\tau/2)}{\omega^2}\left[1-\mathrm{sinc}^2\left(\frac{\omega(T_{\mathrm{obs}}-|\tau|)}{2}\right)\right].$$

(variance of phase fluctuations related to frequency-noise spectrum) (2.168)

Note that this is the same as Eq. (2.101), except for the replacement

$$S_\omega(\omega)\longrightarrow S_\omega(\omega)\left[1-\mathrm{sinc}^2\left(\frac{\omega(T_{\mathrm{obs}}-|\tau|)}{2}\right)\right],$$

(replacement to account for observation time) (2.169)

where the sinc^2 part came from the square of the finite-time-mean fluctuation. This correction factor scales as ω^2 as $\omega\longrightarrow 0$, and so also serves to cut off the dc divergence due to the $1/f$ noise spectrum.

The normalized, one-sided spectrum of the laser, including the observation time T_{obs}, is thus given by Eq. (2.107) as

$$s(\omega)=\frac{1}{2\pi}\int_{-T_{\mathrm{obs}}}^{T_{\mathrm{obs}}}d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\mathrm{obs}}}\right)\exp\left\{-\frac{2}{\pi}\int_0^\infty S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\left[1-\mathrm{sinc}^2\left(\frac{\omega'(T_{\mathrm{obs}}-|\tau|)}{2}\right)\right]d\omega'\right\},$$

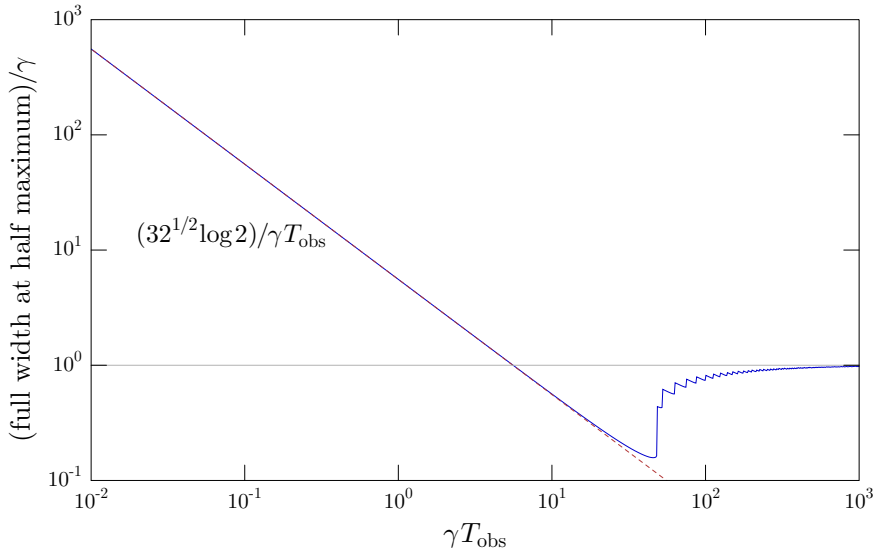
(spectrum of the signal, including observation time) (2.170)

where $\Delta=\omega-\omega_0$, and we are assuming a narrow spectral line compared to the central frequency ω_0, and we are only interested in small detunings $\Delta\ll\omega_0$, in analogy with Eq. (2.146). We have also introduced a factor of $1-|\tau|/T_{\mathrm{obs}}$ in the integral here, which bears some explanation. When we construct the temporal correlation function (2.114) for the intensity power spectrum, this is a finite-time average over the observation time T_{obs}, as in Eq. (2.16). But the intensity signal $I(t)$ itself is windowed within the observation time, so the overlap $I(t)I(t+\tau)$ is only nonzero over the time interval $T_{\mathrm{obs}}-|\tau|$. Thus, the correlation function itself should be modified for the finite observation time by multiplying by $(T_{\mathrm{obs}}-|\tau|)/T_{\mathrm{obs}}$, with the $1/T_{\mathrm{obs}}$ coming from the time average, and the correlation function is zero if $|\tau|>T_{\mathrm{obs}}$. In terms of the spectrum, the observation time imposes a rectangular temporal window, which is equivalent to convolution with $\mathrm{sinc}\,(\omega T_{\mathrm{obs}}/2)$ in the spectrum. However, we have just argued that in the correlation function, the windowing function is the triangular pulse $(T_{\mathrm{obs}}-|\tau|)/T_{\mathrm{obs}}$, which is essentially the self-convolution of the rectangular pulse (since the pulse is symmetric under time reversal). The effect of this in the power spectrum is a convolution with the Fourier transform of the triangle pulse, which is $\mathrm{sinc}^2(\omega T_{\mathrm{obs}}/2)$. For very short observation times, the correlation function is constant over the observation time, and thus the spectrum is just $\mathrm{sinc}\,(\omega T_{\mathrm{obs}}/2)$. Asymptotically, then for small T_{obs}, we expect the spectrum to have an angular FWHM of $4\alpha/T_{\mathrm{obs}}$, where $\alpha\approx 1.39156$ is the positive solution of $\mathrm{sinc}\,x=1/2$.

As a simple example, consider once again the case of white noise, $S_\omega(\omega) = \gamma$, with a finite observation time:

$$\begin{aligned} s(\omega) &= \frac{1}{2\pi}\int_{-T_{\rm obs}}^{T_{\rm obs}} d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\rm obs}}\right)\exp\left\{-\frac{2\gamma}{\pi}\int_0^\infty d\omega'\,\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\left[1-\mathrm{sinc}^2\left(\frac{\omega'(T_{\rm obs}-|\tau|)}{2}\right)\right]\right\} \\ &= \frac{1}{2\pi}\int_{-T_{\rm obs}}^{T_{\rm obs}} d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\rm obs}}\right) \\ &\qquad \times \exp\left[-\frac{\gamma}{12\,(T_{\rm obs}-|\tau|)^2}\left[(2|\tau|-T_{\rm obs})^3 - |2|\tau|-T_{\rm obs}|^3 + 2(T_{\rm obs}-|\tau|)^3\right]\right]. \end{aligned} \tag{2.171}$$

Notice that the exponent reduces to $\exp(-\gamma\tau/2)$ as $T_{\rm obs} \longrightarrow \infty$, as it should. The width of this spectrum is plotted here.

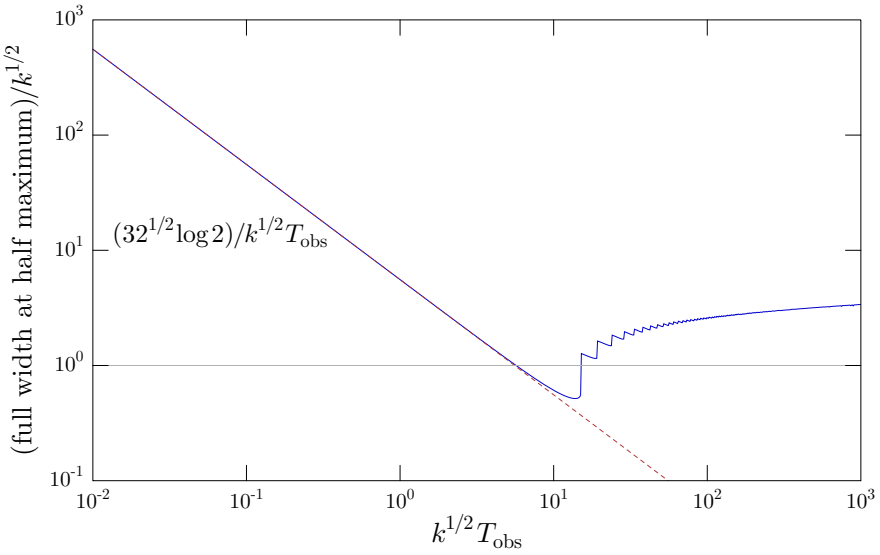


Note that for $\gamma T_{\rm obs} < 1$, the Fourier-broadened asymptotic result for small observation times is a good approximation. For large observation times, the width converges to γ as appropriate for the long-time Lorentzian shape of the line. For intermediate observation times, there is a more complicated step-like dependence on the width, due to the fringes of varying width in the spectrum from the convolution with $\mathrm{sinc}^2(\omega T_{\rm obs}/2)$. Note that the fringes cause some apparent *narrowing* of the spectrum for intermediate observation times, compared to the Lorentzian result.

As a second example, consider $1/f$ noise, $S_\omega(\omega) = k/\omega$, with a finite observation time:

$$\begin{aligned} s(\omega) &= \frac{1}{2\pi}\int_{-T_{\rm obs}}^{T_{\rm obs}} d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\rm obs}}\right)\exp\left\{-\frac{2k}{\pi}\int_0^\infty d\omega'\,\frac{\sin^2(\omega'\tau/2)}{\omega'^3}\left[1-\mathrm{sinc}^2\left(\frac{\omega'(T_{\rm obs}-|\tau|)}{2}\right)\right]\right\} \\ &= \frac{1}{2\pi}\int_{-T_{\rm obs}}^{T_{\rm obs}} d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\rm obs}}\right)\exp\left\{-\frac{2k}{\pi}\left[-\frac{7\tau^2}{48}-\frac{\tau^2}{24}\left(\frac{\tau^2}{(T_{\rm obs}-|\tau|)^2}+6\right)\log|\tau| \right.\right. \\ &\qquad + \frac{(2\tau-T_{\rm obs})^4}{48(T_{\rm obs}-|\tau|)^2}\log\left|2|\tau|-T_{\rm obs}\right| \\ &\qquad \left.\left. - \frac{(T_{\rm obs}-|\tau|)^2}{24}\log(T_{\rm obs}-|\tau|) + \frac{T_{\rm obs}^4}{48(T_{\rm obs}-|\tau|)^2}\log T_{\rm obs}\right]\right\} \\ &= \frac{1}{2\pi}\int_{-T_{\rm obs}}^{T_{\rm obs}} d\tau\, e^{i\Delta\tau}\left(1-\frac{|\tau|}{T_{\rm obs}}\right) e^{7k\tau^2/24\pi}\,|\tau|^{k\tau^2(6+\tau^2/(T_{\rm obs}-|\tau|)^2)/12\pi} \\ &\qquad \times \left|2|\tau|-T_{\rm obs}\right|^{-k(2|\tau|-T_{\rm obs})^4/24\pi(T_{\rm obs}-|\tau|)^2}(T_{\rm obs}-|\tau|)^{k(T_{\rm obs}-|\tau|)^2/12\pi}\,T_{\rm obs}^{-kT_{\rm obs}^4/24\pi(T_{\rm obs}-|\tau|)^2}. \end{aligned} \tag{2.172}$$

The width of this spectrum is plotted below as $T_{\rm obs}$ varies:



The behavior here is similar to the white-noise case, but as $T_{\rm obs}$ increases, the spectral width continues to increase. Again, for small $T_{\rm obs}$, the width matches the Fourier-broadened asymptotic result. For large $T_{\rm obs}$, the heuristic arguments (for a different model of the cutoff) lead to the asymptotic scaling[8] of $(\log \beta k^{1/2} T_{\rm obs})^{1/2}$ (for some constant β), which is very slow divergence in the line width.

We have essentially used the convolution theorem in deriving the finite-observation-time spectrum, where the windowing function $f_T(t)/T$ appeared as its Fourier transform as a high-pass filter in the frequency-noise spectrum, as well as in the form of its self-convolution as a windowing function for the correlation function. Of course, any other windowing function $f(t)$ may be used here, so long as it represents a unit-area pulse, and then, for example, the square of its Fourier transform will appear in place of sinc^2. Since $f(t)$ is normalized, the correction factor will always vanish at $\omega = 0$, taking care of the divergence due to the $1/f$ noise. A non-square windowing function could better model the observation time inherent in the scanning of a spectrum analyzer, for example, where $f(t)$ would be a scaled version of the response function of the final low-pass filter. For example, for a Gaussian window $\propto \exp[-(4\log 2)t^2/T_{\rm obs}^2]$ with a full width at half maximum of $T_{\rm obs}$, Eq. (2.170) is modified to read

$$s(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} \exp\left(-\frac{(2\log 2)t^2}{T_{\rm obs}^2}\right) \times \exp\left\{-\frac{2}{\pi}\int_0^{\infty} S_\omega(\omega')\frac{\sin^2(\omega'\tau/2)}{\omega'^2}\left[1-\exp\left(-\frac{T_{\rm obs}^2\omega'^2}{16\log 2}\right)\right]d\omega'\right\}.$$

(spectrum of the signal, including Gaussian-window observation time) (2.173)

The windowing function $\exp[-(4\log 2)t^2/2T_{\rm obs}^2]$ that appears here has $\exp[-T_{\rm obs}^2\omega^2/8\log 2]$ as its Fourier transform. If we compare this to the spectral resolution function $\exp[-(4\log 2)\,\omega^2/\delta\omega^2]$ of the spectrum analyzer, where $\delta\omega$ is the (full width at half maximum) resolution bandwidth, then the resolution bandwidth is given in terms of the observation time by $\delta\omega = \sqrt{2}(4\log 2)/T_{\rm obs} \approx 4/T_{\rm obs}$. For white noise, we can write this as a Fourier transform,

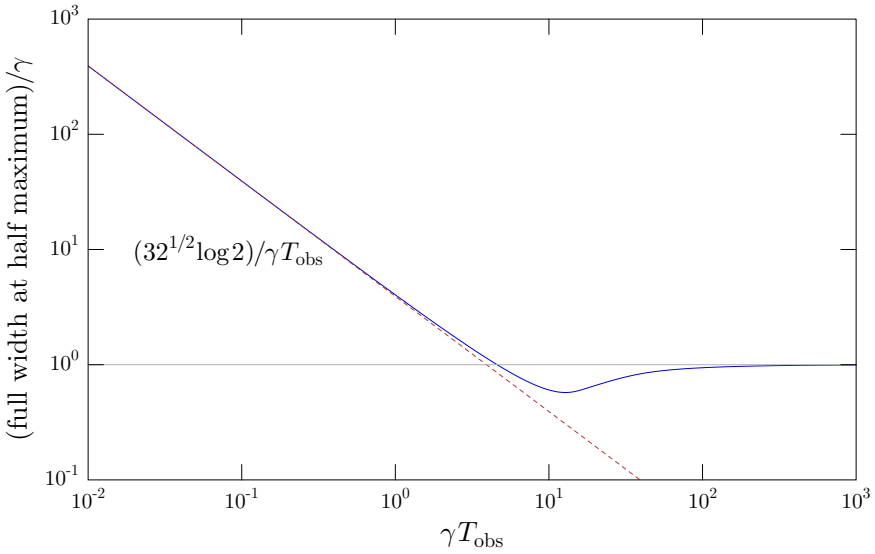
$$s(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} \exp\left(-\frac{(2\log 2)t^2}{T_{\rm obs}^2}\right) \times \exp\left\{-\frac{\gamma}{2}\left[|\tau|\,\mathrm{erfc}\left(\frac{\sqrt{4\log 2}\,|\tau|}{T_{\rm obs}}\right) - \frac{T_{\rm obs}}{\sqrt{4\pi\log 2}}\left(2^{-4\tau^2/T_{\rm obs}^2}-1\right)\right]\right\},$$

(spectrum of the signal, Gaussian-window observation time, white noise) (2.174)

where γ is as usual the long-time Lorenztian width. The behavior of this width is similar to the rectangular-window observation case, except that the function is smoother (there are no fringes associated with the

[8] Gianni Di Domenico, Stéphane Schilt, and Pierre Thomann, "Simple approach to the relation between laser frequency noise and laser line shape," *Applied Optics* **49**, 4801 (2010) (doi: 10.1364/AO.49.004801).

Gaussian convolution) and has a less pronounced dip. The short-time asymptotic form also has a slightly different coefficient: $4\sqrt{2}\log 2/T_{\text{obs}}$.

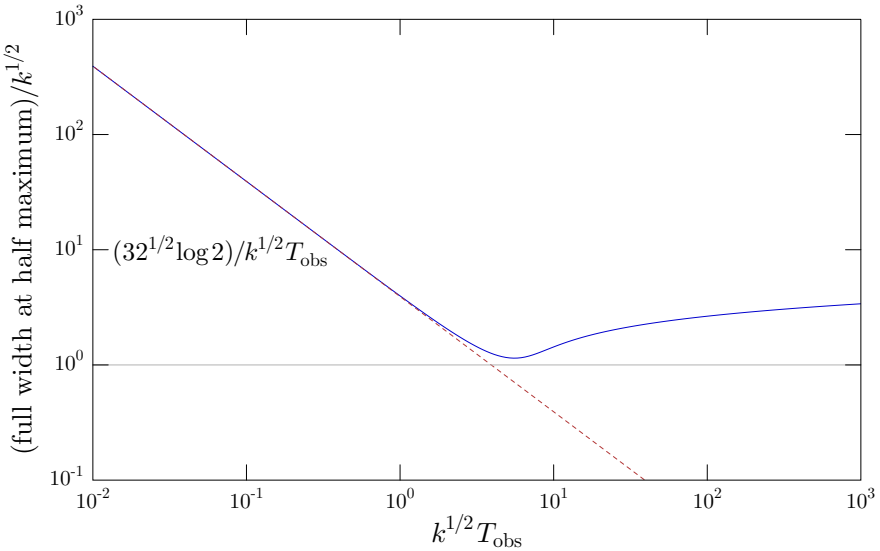


For $1/f$ noise, the expression is

$$s(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\Delta\tau} \exp\left(-\frac{(2\log 2)\tau^2}{T_{\text{obs}}^2}\right) \times \exp\left\{-\frac{k\tau^2}{3\pi T_{\text{obs}}^2}\left[(\log 2)\tau^2\, {}_2F_2\left(1,1;\frac{5}{2},3;-\frac{(4\log 2)\tau^2}{T_{\text{obs}}^2}\right) - \frac{3T_{\text{obs}}^2}{4}\left(\gamma - 3 + \log\frac{(16\log 2)\tau^2}{T_{\text{obs}}^2}\right)\right]\right\},$$

(spectrum of the signal, Gaussian-window observation time, $1/f$ noise) (2.175)

where here γ is Euler's constant, and ${}_2F_2(a_1, a_2; b_1, b_2; z)$ is a generalized hypergeometric function. The behavior is similar to the rectangular-window observation case, but again smoother and with a less pronounced minimum.



The asymptotic scaling for small T_{obs} is the same as for the white-noise case, and the scaling for large T_{obs} appears to match the rectangular-window observation case.

2.9 Exercises

Problem 2.1

In the classical model of spontaneous emission, an atom impulsively excited at $t = 0$ gives rise to radiation in the far-field of

$$E^{(+)}(\mathbf{r}, t) = \frac{E_0^{(+)}}{r}[(\hat{\varepsilon} \cdot \hat{r})\hat{r} - \hat{\varepsilon}]e^{-(\gamma/2)t - i\omega_0 t_r}\Theta(t_r), \tag{2.176}$$

where $t_r = t - r/c$ is the retarded time, and $\Theta(t)$ is the Heaviside step function. Compute the first- and second-order degrees of temporal coherence for this field, and then show that the radiated spectrum $s(\omega)$ is a Lorentzian lineshape with a full width at half maximum of γ.

Problem 2.2

In molecular spectroscopy, *Fourier-transform infrared* (FTIR) spectroscopy is an important technique. The basic idea is to use a Michelson interferometer to measure the correlation function $g^{(1)}(\tau)$ of some input (infrared) field on a detector, which is then digitized. The experimenter takes the Fourier transform on the computer to give the spectrum. This in principle gives the same spectral information as a grating spectrometer, which uses a diffraction grating and an aperture to limit the light so that only a certain range of frequencies hits the detector at any time; scanning the grating position gives the direct spectrum.

(a) Give a (qualitative) argument to justify the following statement: in the infrared, thermal detector noise is significant, so for a given measurement time the FTIR method gives an improved signal/noise ratio compared to the grating method. Assume the same detector is used in both setups.

(b) Give a different (qualitative) argument to justify the following statement: for small detector noise, a grating spectrometer system is superior to an FTIR-type system if it is important to have a large dynamic range in the measured spectrum.

Problem 2.3

Consider the Young double-slit experiment, where two slits are illuminated with classical, coherent light. The setup produces interference fringes on a distant screen due to the variation in path-length difference to the two slits. The fringe visibility $\mathscr{V}$ for a *single* detector is the one we defined in class, and $\mathscr{V} = 1$ for coherent light. We can define a **two-detector fringe visibility** for simultaneous detection by

$$\mathscr{V}^{(2)} := \frac{G^{(2)}_{\max} - G^{(2)}_{\min}}{G^{(2)}_{\max} + G^{(2)}_{\min}}, \tag{2.177}$$

where

$$G^{(2)}(x_1, x_2, \tau = 0) := \langle I(x_1, t) I(x_2, t) \rangle \tag{2.178}$$

is the unnormalized intensity correlation function for simultaneous detection at two points x_1 and x_2 on the screen (x is the direction across the fringes).

(a) What is $\mathscr{V}^{(2)}$ for this double-slit experiment?

(b) Suppose a ''phase scrambler'' is placed in front of one slit to randomize the phase of its transmitted wave. How are $\mathscr{V}$ and $\mathscr{V}^{(2)}$ changed?

Problem 2.4

Given a signal with time dependence of the form $\exp[-i\phi(t)]$, a (one-sided) phase-fluctuation spectrum defined by

$$S_\phi(\omega) := \int_{-\infty}^{\infty} \langle \phi(t)\, \phi(t + \tau) \rangle \cos \omega\tau \, d\tau, \tag{2.179}$$

and a (one-sided) frequency-fluctuation spectrum defined by

$$S_\omega(\omega) := \int_{-\infty}^{\infty} \left\langle \dot{\phi}(t)\, \dot{\phi}(t+\tau) \right\rangle \cos\omega\tau \, d\tau, \tag{2.180}$$

show that the spectra are related by

$$S_\omega(\omega) = \omega^2 S_\phi(\omega). \tag{2.181}$$

Qualitatively, what conditions must be satisfied for this relation to hold?

Problem 2.5

Given a Gaussian random variable X with zero mean and variance σ^2, show that

$$\left\langle e^{-iX} \right\rangle = e^{-\sigma^2/2}. \tag{2.182}$$

Problem 2.6

Suppose we define the exponent A of Eq. (2.134) by

$$A := \left\langle [\phi(t) - \phi(t+\tau) - \phi(t-\tau_\mathrm{d}) + \phi(t+\tau-\tau_\mathrm{d})]^2 \right\rangle. \tag{2.183}$$

Show that

$$A = 2\left\langle [\Delta\phi(\tau)]^2 \right\rangle + 2\left\langle [\Delta\phi(\tau_\mathrm{d})]^2 \right\rangle - \left\langle [\Delta\phi(\tau+\tau_\mathrm{d})]^2 \right\rangle - \left\langle [\Delta\phi(\tau-\tau_\mathrm{d})]^2 \right\rangle. \tag{2.184}$$

Do *not* make any assumptions about whether $\Delta\phi(\tau)$ is correlated at different times t, or about the probability density of $\Delta\phi(\tau)$, other than the fact that the variances above are well-defined and that $\langle \Delta\phi(\tau) \rangle = 0$.

Problem 2.7

Fill in the steps between Eq. (2.135),

$$G_\mathrm{analyzer}(\tau) \propto \frac{\eta_0^2 \eta_1 \eta_2 I_0^2}{2} \cos\omega_\mathrm{m}\tau \, e^{-\left\langle [\Delta\phi(\tau)]^2 \right\rangle - \left\langle [\Delta\phi(\tau_\mathrm{d})]^2 \right\rangle + \left\langle [\Delta\phi(\tau+\tau_\mathrm{d})]^2 \right\rangle/2 + \left\langle [\Delta\phi(\tau-\tau_\mathrm{d})]^2 \right\rangle/2}, \tag{2.185}$$

and Eq. (2.136),

$$G_\mathrm{analyzer}(\tau) \propto \frac{\eta_0^2 \eta_1 \eta_2 I_0^2}{2} \cos\omega_\mathrm{m}\tau \exp\left[-\frac{8}{\pi} \int_0^\infty S_\omega(\omega') \frac{\sin^2(\omega'\tau/2) \sin^2(\omega'\tau_\mathrm{d}/2)}{\omega'^2} \, d\omega' \right]. \tag{2.186}$$

Chapter 3

Rate-Equation Model

Before using a proper quantum model of the atom, we will use a simple model of the atom that includes discrete energy levels. However, we will not include any coherence effects, so the resulting rate equations constitute a sort of ''semi-quantum'' model of the atom. At the same time, we must treat the field with a discrete model, invoking the idea of photons, so that energy exchange between an atom and the field occurs only in multiples of $\hbar\omega$. Even with the language of photons, we will stick to a strictly semiclassical treatment, not really treating the atoms *or* the field quantum mechanically. With this rudimentary model and some simple arguments, we can derive a number of important results without the full apparatus of quantum mechanics.

3.1 Quantization

To elaborate, we will start with the observation that the energy in an electromagnetic field is *quantized*. This means that a monochromatic field of frequency ω (typically restricted to some ''quantization volume'' such as an optical cavity) has possible energies given by

$$E = \left(n + \frac{1}{2}\right)\hbar\omega, \tag{3.1}$$

where n is a nonnegative integer, representing the number of **photons** in the field. This may be familiar as the energy-level structure of the quantum harmonic oscillator. The photon number is always defined with respect to a particular mode (fixing the direction, polarization, and frequency characteristics).

The energies for atoms and molecules are also quantized, although the exact energy-level structure depends on the specific atom or molecule. If we denote the quantized energies by E_n, then the *differences* in energy levels correspond to frequencies via

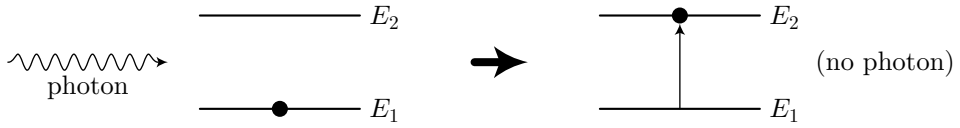
$$\Delta E_{mn} := E_m - E_n = \hbar\omega_{mn}. \tag{3.2}$$

The idea is that atoms with an energy difference ΔE_{mn} prefer to interact with *resonant* fields of frequency ω_{mn}. In this case, the energy of a single photon matches the atomic energy difference, and energy is conserved. There are different types of transitions, generally corresponding to different types of radiation. Electronic transitions in atoms are the most energetic of the type we will consider, and they correspond to visible optical frequencies. Vibrational transitions in a molecule correspond to different amplitudes and types of motion *internal* to the molecule, and generally correspond to radiation in the infrared. Rotational transitions in molecules have yet lower energy, and they correspond to microwave radiation (which enables the **maser**, the microwave predecessor to the laser).

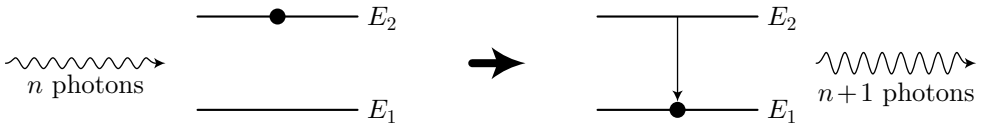
3.2 Fundamental Light–Atom Interactions

There are three fundamental interactions between light and atoms. In all cases we will consider only a two-level atom with ground-state energy E_1 and excited-state energy E_2. We will also assume resonant light, $\omega = (E_2 - E_1)/\hbar$.

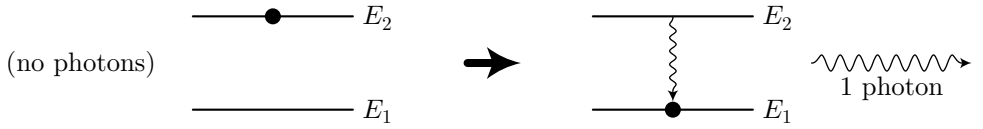
1. **Absorption (stimulated)**. In the absorption process, a photon is destroyed and the atom is promoted to the excited state. More generally, if there are n photons to start with in some resonant mode, then there are $n - 1$ photons after the absorption process.



2. **Stimulated Emission**. This process involves the atom initially being in the excited state, in the presence of n photons in some resonant mode. After the stimulated-emission event, the atom is demoted to the ground state and the field is left with $n + 1$ photons. In some sense, this process is the opposite of stimulated absorption, although absorption ending with 0 photons is possible while stimulated emission beginning with 0 photons is not.



3. **Spontaneous Emission**. This process is much like stimulated emission, but when the atom is demoted, a photon is created in some mode that is initially unpopulated. Thus, a photon can go into a wide range of possible modes by spontaneous emission. It is possible to view spontaneous emission as stimulated emission due to quantum vacuum fluctuations in addition to classical radiation reaction.



Generally, we can associate stimulated absorption and emission with a single mode that is already populated, or singled out by some other means, such as an optical cavity. We can associate spontaneous emission additionally with all other modes.

3.3 Einstein Rate Equations

Now let's consider an *ensemble* of two-level atoms interacting with light. Let $N_{1,2}$ denote the number density of atoms with energy $E_{1,2}$. Then the Einstein rate equation for the excited state is[1]

$$\frac{dN_2}{dt} = -A_{21}N_2 - B_{21}\rho(\omega)N_2 + B_{12}\rho(\omega)N_1. \tag{3.3}$$
(Einstein rate equation)

Here, $\rho(\omega)$ is the energy density of the electromagnetic field (the energy density in the frequency interval ω to $\omega + d\omega$). The first term corresponds to spontaneous emission, and we can see that it reduces the excited-state population, even in the absence of any field. The second and third terms are proportional to

[1] A. Einstein, "Zur Quantentheorie der Strahlung," *Physikalische Zeitschrift* **18**, 121 (1917), translation "On the Quantum Theory of Radiation" by Alfred Engel appears in *The Collected Papers of Albert Einstein, Volume 6, The Berlin Years: Writings, 1914-1917* (Princeton University Press, 1997) p. 200 (ISBN: 0691017344).

$\rho(\omega)$, and correspond to stimulated emission and absorption, respectively, as we can see from their overall signs. By convention, the constant A_{21} is called the **Einstein A coefficient**, while B_{21} and B_{12} are called the **Einstein B coefficients**. At this point we are simply postulating that the three processes contribute to the atomic evolution in this form, with the rate coefficients yet to be determined.

The Einstein A coefficient here represents the rate at which energy is lost from the atom. We can thus identify $A_{21} = \gamma$, where γ is the damping rate from the Lorentz atom. The connection of the Lorentz atom with the B coefficients is less clear, in part because the classical model gets this wrong (hence the necessity of patching the classical solution with the oscillator strength). We will defer this comparison until we derive the cross section for the two-level atom.

To be consistent, $N_1 + N_2$ must add up to some constant, assuming that we really have two-level atoms and that the atoms stay in place (something that even works fairly well for gas lasers as long as we modify A_{21} appropriately). Thus, $dN_2/dt = -dN_1/dt$, and so it is easy to write down the rate equation

$$\frac{dN_1}{dt} = A_{21}N_2 + B_{21}\rho(\omega)N_2 - B_{12}\rho(\omega)N_1 \tag{3.4}$$
(steady-state solution)

for the ground-state population N_1.

We can gain some valuable insight by looking at the equilibrium behavior of the rate equations. Steady state occurs when $dN_2/dt = 0$, whence it follows from Eq. (3.3) that

$$\frac{N_2}{N_1} = \frac{B_{12}\rho(\omega)}{A_{21} + B_{21}\rho(\omega)}. \tag{3.5}$$

If the energy levels are not degenerate, it turns out that $B_{12} = B_{21}$, as we will see shortly. That is, stimulated emission and absorption are exactly symmetric from the rate-equation point of view. Then we can rewrite the steady-state solution as

$$\frac{N_2}{N_1} = \frac{1}{\dfrac{A_{21}}{B_{21}\rho(\omega)} + 1}. \tag{3.6}$$

We can see from this that $N_2 < N_1$ in steady state. This result has an important result for using atoms as a gain medium for a laser: there is no steady-state population inversion in a two-level system, and hence there is no net gain of light transmitted through a medium composed of two-level atoms. This is because on average, absorption (attenuation) occurs more often than stimulated emission (amplification).

In the limit of large intensity, $\rho(\omega) \longrightarrow \infty$, the populations *equalize*. This points to an important effect that is missed by the Lorentz model: atomic saturation. For small excitation, N_2/N_1 is proportional to $\rho(\omega)$, but as the excitation increases, the slope of N_2/N_1 decreases, dropping to zero for large intensities. We will treat this point more carefully after establishing some more results regarding the rate coefficients.

3.4 Relations Between the Einstein Coefficients

Now we briefly outline Einstein's derivation of the relation between the A and B coefficients. If the energy levels are degenerate, we can define the degeneracy factors $g_{1,2}$ as the number of ways of having energy $E_{1,2}$. For example $g_{1,2} = 2J_{1,2} + 1$ for atomic angular-momentum states. Then the steady-state population ratio from Eq. (3.4) can be written also via Boltzmann statistics as

$$\frac{N_2}{N_1} = \frac{g_2}{g_1}e^{-\hbar\omega/k_\mathrm{B}T} = \frac{B_{12}\rho(\omega)}{A_{21} + B_{21}\rho(\omega)}. \tag{3.7}$$

Solving for $\rho(\omega)$,

$$\rho(\omega) = \frac{A_{21}}{B_{21}} \frac{1}{\left(\dfrac{B_{12}g_1}{B_{21}g_2}e^{\hbar\omega/k_\mathrm{B}T} - 1\right)}. \tag{3.8}$$

This is equivalent to the Planck blackbody distribution[2]

$$\rho(\omega) = \frac{8\pi h}{\lambda^3} \frac{1}{e^{\hbar\omega/k_{\mathrm{B}}T} - 1} \tag{3.9}$$

if we make the identifications

$$g_2 B_{21} = g_1 B_{12} \tag{3.10}$$
(relation between B coefficients)

and

$$\frac{A_{21}}{B_{21}} = \frac{8\pi h}{\lambda^3}. \tag{3.11}$$
(relation between A and B coefficients)

Recall that λ here is the wavelength *within* the atomic medium.

Remarkably, this simple thermodynamic argument reproduces the full quantum result that we will derive later. Essentially, this is because the Planck distribution is valid for a particular set of quantum (thermal) states (at some level, a proper summation over field modes is buried in the Planck distribution, as evidenced by the correct frequency dependence of ω^3). This is sufficient to establish the relationship between the coefficients, since they are independent of the quantum state.

3.5 Line Shape and Spectral Distributions

So far, we've considered only monochromatic light and two-level atoms with sharply defined energy levels. Now it's time to improve our model of the two-level atom and its interaction with light.

We will first introduce a **line-shape function** $s(\omega)$ to model the fact that the energy levels have some width. The line shape is defined such that $s(\omega)\,d\omega$ is the probability that a spontaneously emitted photon will have frequency between ω and $\omega + d\omega$. We can also interpret this as the *relative* probability of stimulated emission or absorption of a photon with frequency between ω and $\omega + d\omega$. Since $s(\omega)$ represents a probability density, it is appropriately normalized:

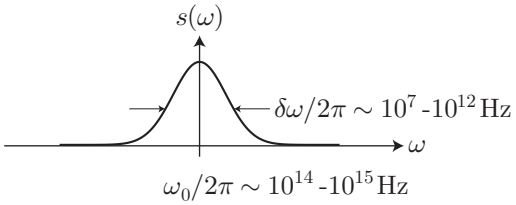
$$\int_0^\infty s(\omega)\,d\omega = 1. \tag{3.12}$$

Note that as in our discussion of coherence in Chapter 2, we are using a ''one-sided spectrum'' that ranges only over positive frequencies. In terms of the ''two-sided spectrum'' $s_\pm(\omega)$ with both positive and negative frequencies, the one-sided spectrum satisfies $s(\omega) := s_\pm(\omega) + s_\pm(-\omega)$ for $\omega \geq 0$ and $s(\omega) = 0$ for $\omega < 0$.

When we apply the line shape and sum over all frequencies, the rate equation becomes

$$\frac{dN_2}{dt} = -A_{21}N_2 - B_{21}N_2 \int_0^\infty \rho(\omega)s(\omega)\,d\omega + B_{12}N_1 \int_0^\infty \rho(\omega)s(\omega)\,d\omega. \tag{3.13}$$

Qualitatively, we can picture the line shape function as a relatively sharply peaked distribution centered around the resonant optical frequency ω_0.



Often, $s(\omega)$ turns out to be a Lorentzian, a Gaussian, or a convolution of the two (a *Voigt* profile). The line-shape function models transition width due to spontaneous emission, collisions, Doppler shifts in gas lasers, and local crystal structure effects on the dopant atoms. Note that in the absence of radiation,

[2] P. W. Milonni and M.-L. Shih, ''Zero-point energy in early quantum theory,'' *American Journal of Physics* **59**, 684 (1991) (doi: 10.1119/1.16772).

the rate equation is $dN_2/dt = -A_{21}N_2$, which has an exponentially damping solution. As we discussed before, the Fourier transform of an exponential is a Lorentzian, so the line shape for spontaneous emission (the "natural line shape") is Lorentzian, with a half-width at half maximum of A_{21}. Collisions are often modeled by a spontaneous-emission-like term, and thus also lead to Lorentzian line shapes. Doppler shifts lead to Gaussian line shapes because the Maxwell–Boltzmann velocity distribution is Gaussian. If multiple, independent broadening effects contribute, their combined effect can be modeled by the convolution of the individual line shapes.

Now we will consider two limiting cases for the light spectrum. Both are important in understanding laser operation, but the second is the more useful case for comparing to coherent quantum light–atom interactions.

3.5.1 Broadband Light

Light is broadband (relative to the transition) if $\rho(\omega)$ is much broader than $s(\omega)$. Then we can evaluate the integral in Eq. (3.13) by noting that $\rho(\omega)$ varies slowly over the width of $s(\omega)$, so that we can pull it out of the integral:

$$\int_0^\infty \rho(\omega)s(\omega)\,d\omega \approx \rho(\omega_0)\int_0^\infty s(\omega)\,d\omega = \rho(\omega_0). \tag{3.14}$$

Thus, we recover the previous rate equations, corresponding to Eq. (3.3), with sharp energy levels.

3.5.2 Nearly Monochromatic Light

For nearly monochromatic light, the field spectrum is narrow, so $s(\omega)$ is much broader than $\rho(\omega)$. Thus, we can evaluate the integral with the same slowly varying approximation as for the broadband case:

$$\int_0^\infty \rho(\omega)s(\omega)\,d\omega \approx s(\omega_{\text{field}})\int_0^\infty \rho(\omega)\,d\omega. \tag{3.15}$$

The integral on the right-hand side is the total field energy density, summed over all frequencies. Let's denote this simply by ρ. Then the rate equation becomes

$$\frac{dN_2}{dt} = -A_{21}N_2 - B_{21}N_2 s(\omega)\rho + B_{12}N_1 s(\omega)\rho, \tag{3.16}$$

where we have written $s(\omega)$ in place of $s(\omega_{\text{field}})$. The total energy density is related to the total intensity I by $\rho = I/c$, so

$$\frac{dN_2}{dt} = -A_{21}N_2 - \frac{\sigma(\omega)I}{\hbar\omega}\left[N_2 - \frac{g_2}{g_1}N_1\right]. \tag{3.17}$$

(rate equation, monochromatic light)

Here, we have defined the **absorption cross-section**

$$\sigma(\omega) = A_{21}\frac{\lambda^2}{4}s(\omega). \tag{3.18}$$

(cross section)

The cross section has the dimensions of area, and is defined such that $\sigma(\omega)I$ is the power absorbed by a single atom when irradiated by intensity I (in the weak-excitation limit). Note that for a Lorentzian line shape $s(\omega)$,

$$s(\omega) = \frac{\Delta\omega}{2\pi\left[(\omega_0-\omega)^2 + (\Delta\omega/2)^2\right]}, \tag{3.19}$$

the *resonant* cross section $\sigma(\omega_0)$ is given by

$$\sigma(\omega_0) = \frac{A_{21}}{\Delta\omega}\,\frac{\lambda^2}{2\pi}. \tag{3.20}$$

For homogeneous broadening, $\Delta\omega$ is the **natural line width** given by $\Delta\omega = A_{21}$, so that the natural cross section is

$$\sigma(\omega_0) = \frac{\lambda_0^2}{2\pi}. \tag{3.21}$$
(natural, on-resonance cross section)

This answer is consistent with a fully quantum-mechanical calculation, and this is the same cross-section that we used before [Eq. (1.35)] to derive the form of the oscillator strength. This relation also establishes the relation of the Einstein B coefficient to the classical model of the atom. Note that this answer assumes an average over all possible atomic orientations, since the blackbody distribution of Eq. (3.9) assumes isotropic radiation. For atomic dipole moments aligned with the field polarization, the resonant cross section is

$$\sigma(\omega_0) = \frac{3\lambda_0^2}{2\pi}, \tag{3.22}$$

since the coupling that would normally be ''distributed'' among three orthogonal directions is concentrated into one.

3.6 Absorption Coefficient and Saturation

Let us assume that nearly monochromatic light of frequency ω passes through a vapor of two-level atoms. Evidently, from the rate equation (3.17), the rate per unit volume at which atoms are being promoted to the excited state by the pumping field is

$$-\frac{\sigma(\omega)I}{\hbar\omega}\left[N_2 - \frac{g_2}{g_1}N_1\right]. \tag{3.23}$$

We define the absorption coefficient by

$$\frac{dI}{dz} = -a(\omega)I(z). \tag{3.24}$$
(absorption coefficient definition)

Then we multiply the expression (3.23) by the photon energy $\hbar\omega$ to obtain the rate of energy absorption per unit volume, or equivalently the rate of intensity absorption per unit length, which matches the right-hand side of Eq. (3.24). Thus, we find

$$a(\omega) = -\sigma(\omega)\left[N_2 - \frac{g_2}{g_1}N_1\right]. \tag{3.25}$$

We can get the population difference here from the steady state of the rate equation (3.17). This gives

$$\frac{N_2}{N_1} = \frac{g_2}{g_1}\left(\frac{\dfrac{\sigma I}{\hbar\omega A_{21}}}{1 + \dfrac{\sigma I}{\hbar\omega A_{21}}}\right). \tag{3.26}$$

Noting that $N_1 + N_2 = N$, which implies

$$\frac{N_2 - \dfrac{g_2}{g_1}N_1}{N} = \frac{\left(\dfrac{g_1}{g_2}\dfrac{N_2}{N_1} - 1\right)}{\left(\dfrac{g_1}{g_2}\dfrac{N_2}{N_1} + \dfrac{g_1}{g_2}\right)} = -\frac{g_2/g_1}{1 + \left(1 + \dfrac{g_2}{g_1}\right)\dfrac{\sigma I}{\hbar\omega A_{21}}}. \tag{3.27}$$

Putting this result into Eq. (3.25), we find for the absorption coefficient

$$a(\omega) = \left(\frac{g_2}{g_1}\right)\frac{\sigma(\omega)N}{1 + \left(1 + \dfrac{g_2}{g_1}\right)\dfrac{\sigma(\omega)I}{\hbar\omega A_{21}}}. \tag{3.28}$$

Note that in the case of equal degeneracies, $g_1 = g_2$, we have

$$a(\omega) = \frac{\sigma(\omega)N}{1 + 2\dfrac{\sigma(\omega)I}{\hbar\omega A_{21}}}. \tag{3.29}$$
(absorption coefficient, $g_1 = g_2$)

For small intensities, this expression is equivalent to the classical expression (1.33), which had the constant value $\sigma(\omega)N$. For large intensities, the absorption coefficient falls to zero. This is the effect of **saturation** or **optical bleaching** of the medium. On resonance, the absorption coefficient becomes

$$a(\omega_0) = \frac{\sigma_0 N}{1 + 2\dfrac{\sigma_0 I}{\hbar\omega_0 A_{21}}}, \tag{3.30}$$

where $\sigma_0 = \sigma(\omega_0)$ is the resonant cross-section. It is convenient to define the **saturation intensity** I_{sat} by

$$I_{\text{sat}} := \frac{\hbar\omega_0 A_{21}}{2\sigma_0}, \tag{3.31}$$
(saturation intensity)

so that we can write the resonant absorption coefficient as

$$a(\omega_0) = \frac{\sigma_0 N}{1 + \dfrac{I}{I_{\text{sat}}}}. \tag{3.32}$$

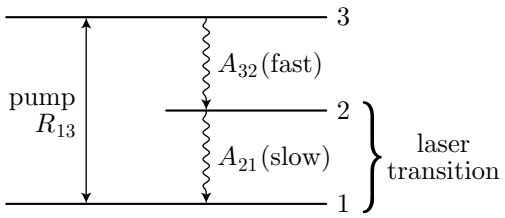
The saturation intensity gives the intensity scale over which saturation sets in. Specifically, we see that the absorption coefficient drops to half the small-signal value when $I = I_{\text{sat}}$. Again, this is one important feature of the light–matter interaction that the classical model misses: the harmonic oscillator can be excited to arbitrarily high amplitudes, but in a quantum-mechanical atom, the best excitation is when the maximum number of atoms are pumped into the excited state.

3.7 Exercises

Problem 3.1

We said that a medium of two-level atoms is no good as a laser gain medium, since the ground state ends up with more population than the excited state, and so absorption wins out over stimulated emission (i.e., loss wins out over amplification).

The simplest change that we can make to achieve a *population inversion*, where the excited state is more populated than the ground state, is to add a third level. The level scheme is shown here.



The new level (with highest energy) decays quickly, while the laser ($2 \longrightarrow 1$) transition decays slowly. That is, we will assume $A_{21} \ll R_{13}$, A_{32}. Also, for a monochromatic pump (e.g., the pump is another laser),

$$R_{13} = \frac{\sigma(\omega) I}{\hbar\omega}, \tag{3.33}$$

where $\sigma(\omega)$ is the absorption cross section for the $1 \longrightarrow 3$ transition, and ω is the frequency of the pumping field. Then we can write the rate equations for the three-level atom as

$$\begin{aligned} \frac{dN_3}{dt} &= -R_{13}(N_3 - N_1) - A_{32}N_3 \\ \frac{dN_2}{dt} &= A_{32}N_3 - A_{21}N_2 \\ \frac{dN_1}{dt} &= A_{21}N_2 + R_{13}(N_3 - N_1), \end{aligned} \tag{3.34}$$

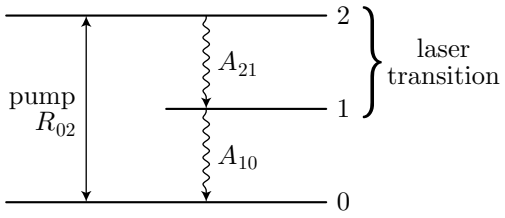
where of course one of the equations is redundant since $N_1 + N_2 + N_3$ must be a constant of the motion.

The key to why this scheme gives a population inversion on the laser transition ($1 \longrightarrow 2$) is that atoms will be promoted to level 3 after absorbing a pump photon, and they will quickly decay to the metastable level 2 before stimulated emission by the pump field returns them to level 1. In this way the pump depletes level 1 and populates level 2 without trying to return the atoms back to level 1.

(a) Got all that? Outstanding. Now find the steady-state solution in the limit where A_{32} is by far the fastest time scale in the problem. Under what conditions does a population inversion occur on the laser transition? What should you do to get the *best* possible inversion?

You can do this directly, but here is the fancy-schmancy way. Note that level 3 decays quickly, and the decay term is like a damping term for N_3. Thus, we can assume N_3 is always in quasiequilibrium with respect to N_1 and N_2, which evolve comparatively slowly. So we can take $dN_3/dt \approx 0$ to obtain an approximate ("adiabatic") expression for N_3. Now use this to *adiabatically eliminate* the N_3's in the other two rate equations. By now you should have a set of effective rate equations for a two-level atom. Finding the steady state of these new equations is a piece of cake.

(b) Now that you're comfortable with the three-level atom, let's turn it upside down and consider the *inverted* three-level laser scheme shown here.



Argue *qualitatively* that in the best possible cases (optimal pumping and decay rates), for the same pumping rate, laser cross section, and atomic number density, the small-signal gain coefficient for the usual scheme is *twice* the small-signal gain coefficient for the inverted scheme shown here. Recall that the gain corresponds to negative absorption, so we can write the gain coefficient as $\gamma(\omega) = -a(\omega) = \sigma(\omega)[N_e - N_g]$, where "e" and "g" refer to the excited and ground levels of the laser transition, respectively.

Note that you *could* go solving a new set of rate equations in steady state, but if you do some thinking, you will realize that you don't really *need* to. Just use the analysis of the two-level atom to reason out what the steady-state populations would be in the optimal cases.

(c) Give a *qualitative* argument for why the saturation intensity for the inverted three-level scheme will be *twice* that of the saturation intensity for the usual three-level scheme. Assume the laser cross sections are the same in both cases.

Chapter 4

The Quantum State

4.1 Density Operator

Traditionally, the **state vector** $|\psi\rangle$ represents the state of a quantum system. However, we will need a more general object to represent the quantum state for the purposes of studying light-matter interactions. The density operator represents the state of a quantum system in a more general way than the state vector, and equivalently represents an observer's *state of knowledge* of a system. It is particularly important to use the density operator in the quantum theory of *open* systems, where a quantum system interacts with an external system whose evolution is unknown, and in the quantum theory of measurement and information.

When a quantum state can be represented by a state vector $|\psi\rangle$, the **density operator** is defined as the product

$$\rho := |\psi\rangle\langle\psi|. \tag{4.1}$$
(density operator, pure state)

In this case, it is obvious that the information content of the density operator is equivalent to that of the state vector (except for the overall phase, which is not of physical significance).

The state vector can represent states of *coherent* superposition. The power of the density operator lies in the fact that it can represent *incoherent* superpositions as well. For example, let $|\psi_\alpha\rangle$ be a set of states (without any particular restrictions). Then the density operator

$$\rho = \sum_\alpha P_\alpha |\psi_\alpha\rangle\langle\psi_\alpha| \tag{4.2}$$
(density operator, general)

models the fact that we don't know *which* of the states $|\psi_\alpha\rangle$ the system is in, but we assign a probability or weight P_α to the quantum state $|\psi_\alpha\rangle$ in the mixture defined by ρ. Note that the weights obey

$$\sum_\alpha P_\alpha = 1 \tag{4.3}$$

for proper normalization of the density operator. Another way to say it is this: the state vector $|\psi\rangle$ represents a certain *intrinsic uncertainty* with respect to quantum observables; the density operator can represent uncertainty *beyond* the minimum required by quantum mechanics. Equivalently, the density operator can represent an *ensemble* of identical systems in possibly different states.

A state of the form (4.1) is said to be a **pure state**. One that cannot be written in this form is said to be **mixed**.

4.1.1 Example

As a simple example, consider a **qubit**, a two-level system with states $|0\rangle$ and $|1\rangle$. The density operators corresponding to the eigenstates are $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$; clearly these are pure states. Another pure state is

the superposition $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, which has the corresponding density operator

$$\rho = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1| + |0\rangle\langle 1| + |1\rangle\langle 0|\Big). \tag{4.4}$$

The density operator is the sum of the density operators for the eigenstates, plus two extra terms that indicated the *purity* of the state or the *coherence* of the superposition. An example of a mixture of the two eigenstates comes from simply removing these last two terms:

$$\rho = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1|\Big). \tag{4.5}$$

We can clearly regard this as an mixture of the form (4.2), where the probabilities are $P_{0,1} = 1/2$ for the eigenstates $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = |1\rangle$. However, we can equally well regard the same mixed state as a different mixture. That is, defining the mixed state

$$\rho' = \frac{1}{2}\Big(|+\rangle\langle +| + |-\rangle\langle -|\Big), \tag{4.6}$$

where

$$|\pm\rangle := \frac{1}{\sqrt{2}}\Big(|0\rangle \pm |1\rangle\Big). \tag{4.7}$$

it is not hard to see that $\rho = \rho'$. Thus we see that we have to be a bit careful with our above statement, where we said that a mixed state can be regarded as an association of classical probabilities with being in different pure quantum states. Just given a particular density operator, it is not not in general possible to uniquely define a pure-state decomposition of the form (4.2). Thus stating that the state is *really* in a pure state, but we don't quite know *which* one it's in, implies some *extra* information that is not contained in the density operator.

4.1.2 Evolution

Differentiating the density operator and employing the Schrödinger equation $i\hbar\partial_t|\psi\rangle = H|\psi\rangle$, we can write down the equation of motion for the density operator:

$$\begin{aligned}\partial_t\rho &= (\partial_t|\psi\rangle)\langle\psi| + |\psi\rangle\partial_t\langle\psi| \\ &= -\frac{i}{\hbar}H\rho + \frac{i}{\hbar}\rho H \\ &= -\frac{i}{\hbar}[H,\rho].\end{aligned} \tag{4.8}$$

(Schrödinger–von Neumann equation)

This is referred to as the **Schrödinger–von Neumann equation**. The derivation here assumed a pure state but carries through in the obvious way for arbitrary density operators. Of course, the point is that using the density operator allows us to write down more general evolution equations than those implied by state-vector dynamics. The more general forms are referred to as **Liouville–von Neumann equations** or **master equations**, which we can write in the form

$$\partial_t\rho = \mathcal{L}\rho. \tag{4.9}$$

(master equation, generic form)

Here, $\mathcal{L}$ is the **Liouvillian superoperator**. We use the term ''superoperator'' because the Liouvillian represents a higher-dimensional object, since it must represent the commutator above (i.e., it ''operates from both sides''). Thinking of the density operator as a two-dimensional matrix as we discuss below, the Liouvillian is effectively a rank-4 tensor.

4.1.3 Expectation Values

We can compute expectation values with respect to the density operator via the trace operation. The trace of an operator A is simply the sum over the diagonal matrix elements with respect to any complete, orthonormal set of states $|\beta\rangle$:

$$\mathrm{Tr}[A] := \sum_{\beta} \langle \beta | A | \beta \rangle \tag{4.10}$$

An important property of the trace is that the trace of a product is invariant under cyclic permutations of the product. For example, for three operators,

$$\mathrm{Tr}[ABC] = \mathrm{Tr}[BCA] = \mathrm{Tr}[CAB]. \tag{4.11}$$
(cyclic permutation invariance)

This amounts to simply an interchange in the order of summations. For example, for two operators, working in the position representation,

$$\begin{aligned} \mathrm{Tr}[AB] &= \int dx \, \langle x | AB | x \rangle \\ &= \int dx \int dx' \, \langle x | A | x' \rangle \langle x' | B | x \rangle \\ &= \int dx' \int dx \, \langle x' | B | x \rangle \langle x | A | x' \rangle \\ &= \int dx' \, \langle x' | BA | x' \rangle \\ &= \mathrm{Tr}[BA]. \end{aligned} \tag{4.12}$$

Note that this argument assumes sufficiently ''nice'' operators (it fails, for example, for $\mathrm{Tr}[xp]$). More general permutations [e.g., of the form (4.11)] are obtained by replacements of the form $B \longrightarrow BC$. Using this property, we can obviously write the expectation value with respect to a pure state as

$$\langle A \rangle = \langle \psi | A | \psi \rangle = \mathrm{Tr}[A\rho]. \tag{4.13}$$
(expectation value, pure state)

This obviously extends to the more general form (4.2) of the density operator. Taking an additional average over the ensemble of pure states,

$$\langle\langle A \rangle\rangle = \sum_{\alpha} P_\alpha \langle \psi_\alpha | A | \psi_\alpha \rangle = \mathrm{Tr}[A\rho], \tag{4.14}$$
(expectation value, ensemble)

where the double angle brackets $\langle\langle\rangle\rangle$ denote the ensemble average over expectation values. For simplicity we will drop the extra brackets and simply use single brackets for expectation values with respect to either a pure state or an ensemble ($\langle\langle\rangle\rangle \longrightarrow \langle\rangle$).

4.1.4 The Density Matrix

The physical content of the density operator is more apparent when we compute the elements $\rho_{\alpha\alpha'}$ of the *density matrix* with respect to a complete, orthonormal basis. The density matrix elements are given by

$$\rho_{\alpha\alpha'} := \langle \alpha | \rho | \alpha' \rangle. \tag{4.15}$$
(density matrix)

To analyze these matrix elements, we will assume the simple form $\rho = |\psi\rangle\langle\psi|$ of the density operator, though the arguments generalize easily to arbitrary density operators.

The diagonal elements $\rho_{\alpha\alpha}$ are referred to as *populations*, and give the measurement probability of the system in the state $|\alpha\rangle$:

$$\rho_{\alpha\alpha} = \langle \alpha | \rho | \alpha \rangle = |\langle \alpha | \psi \rangle|^2 . \tag{4.16}$$

The off-diagonal elements $\rho_{\alpha\alpha'}$ (with $\alpha \neq \alpha'$) are referred to as *coherences*, since they give information about the relative phase of different components of the superposition. For example, if we write the state vector as a superposition with explicit phases,

$$|\psi\rangle = \sum_\alpha |c_\alpha| \, e^{i\phi_\alpha} |\alpha\rangle, \tag{4.17}$$

then the coherences are

$$\rho_{\alpha\alpha'} = |c_\alpha c_{\alpha'}| \, e^{i(\phi_\alpha - \phi_{\alpha'})}. \tag{4.18}$$

Notice that for a density operator not corresponding to a pure state, the coherences in general will be the sum of complex numbers corresponding to different states in the incoherent sum. The phases will not in general line up, so that while $|\rho_{\alpha\alpha'}|^2 = \rho_{\alpha\alpha}\rho_{\alpha'\alpha'}$ for a pure state, we expect $|\rho_{\alpha\alpha'}|^2 < \rho_{\alpha\alpha}\rho_{\alpha'\alpha'}$ ($\alpha \neq \alpha'$) for a generic mixed state.

4.1.5 Purity

How can we tell a pure state from a mixed one in general? Notice that the diagonal elements of the density matrix form a probability distribution. Proper normalization thus requires

$$\mathrm{Tr}[\rho] = \sum_\alpha \rho_{\alpha\alpha} = 1. \tag{4.19}$$
(normalization)

We can do the same computation for ρ^2, and we will define the **purity** to be $\mathrm{Tr}[\rho^2]$. For a pure state, the purity is simple to calculate, since $\rho^2 = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = \rho$:

$$\mathrm{Tr}[\rho^2] = \mathrm{Tr}[\rho] = 1. \tag{4.20}$$
(purity for pure state)

(In fact $\rho^n = \rho$ in a pure state for any nonnegative n.) But for mixed states, $\mathrm{Tr}[\rho^2] < 1$. For example, for the density operator in (4.2),

$$\mathrm{Tr}[\rho^2] = \sum_\alpha P_\alpha^2, \tag{4.21}$$

if we assume the states $|\psi_\alpha\rangle$ to be orthonormal. For equal probability of being in N such states, $\mathrm{Tr}[\rho^2] = 1/N$. Intuitively, then we can see that $\mathrm{Tr}[\rho^2]$ drops to zero as the state becomes more mixed—that is, as it becomes an incoherent superposition of more and more orthogonal states.

To prove that $\mathrm{Tr}[\rho^2] < 1$ for mixed states, first note that ρ is a Hermitian operator ($\rho = \rho^\dagger$). Thus, ρ may be diagonalized by a unitary transformation, so we may write

$$\rho' = S\rho S^\dagger, \tag{4.22}$$

where ρ' is diagonal and $S^{-1} = S^\dagger$. It is easy to verify that the trace is invariant under unitary transformations, so

$$\mathrm{Tr}[\rho^2] = \mathrm{Tr}[\rho'^2] = \sum_\alpha (\rho'_{\alpha\alpha})^2 \leq \sum_\alpha \rho'_{\alpha\alpha} = 1, \tag{4.23}$$

where the inequality comes from noting that $0 \leq \rho'_{\alpha\alpha} \leq 1$, so that ${\rho'}^2_{\alpha\alpha} \leq \rho'_{\alpha\alpha}$. A diagonal pure state has only a single nonzero diagonal element, while a diagonal mixed state necessarily has more than one nonzero diagonal element. Hence, for a mixed state, $\mathrm{Tr}[\rho^2] < 1$. This follows since the diagonal matrix elements are positive,

$$\rho_{\alpha\alpha} = \langle\alpha|\rho|\alpha\rangle = \sum_k P_k |\langle\alpha|\psi\rangle_k|^2 \geq 0, \tag{4.24}$$

and so the equality occurs only for a single term in the sum.

4.2 Pictures

4.2.1 Unitary Time-Evolution Operator

The Schrödinger equation generates the time evolution of the state vector $|\psi\rangle$. It is convenient to represent time evolution in the form of an operator:

$$|\psi(t_0)\rangle \longrightarrow |\psi(t)\rangle = U(t,t_0)\,|\psi(t_0)\rangle. \tag{4.25}$$
(time-evolution operator, definition)

Here, $U(t,t_0)$ is the **unitary time-evolution operator**[1] that evolves the state from time t_0 to t. Note that the operator must be unitary to preserve the norm of the state vector. Since

$$\langle\psi(t)|\psi(t)\rangle = \langle\psi(t_0)|U^\dagger(t,t_0)U(t,t_0)|\psi(t_0)\rangle, \tag{4.26}$$

if we require this to be equal to $\langle\psi(t_0)|\psi(t_0)\rangle$ for any initial state $|\psi(t_0)\rangle$, then it follows that

$$U^\dagger(t,t_0)\,U(t,t_0) = 1, \tag{4.27}$$
(unitary condition)

and thus $U(t,t_0)$ is unitary. In other words, unitarity of the evolution is required to conserve probability.

The time-evolution operator also must have the composition property

$$U(t_2,t_0) = U(t_2,t_1)\,U(t_1,t_0), \tag{4.28}$$
(composition property)

which is sensible for the representation of time evolution. In this relation, note that with the time ordering $t_0 < t_1 < t_2$, the earliest time appears to the right, since it is that one that operates "first" on the state vector. Finally, we must have the inversion property

$$U(t,t') = U^{-1}(t',t) = U^\dagger(t',t), \tag{4.29}$$
(inversion property)

so that the inverse of an evolution operator corresponds to backwards-time evolution.

4.2.1.1 Infinitesimal Form

Again, the Schrödinger equation

$$\partial_t|\psi\rangle = -\frac{i}{\hbar}H|\psi\rangle \tag{4.30}$$

can be rewritten in differential form as

$$|\psi(t+dt)\rangle - |\psi(t)\rangle = -\frac{i}{\hbar}H|\psi(t)\rangle\,dt, \tag{4.31}$$

and thus generates the evolution over an interval dt according to

$$|\psi(t)\rangle \longrightarrow |\psi(t+dt)\rangle = \left(1 - \frac{i}{\hbar}H\,dt\right)|\psi(t)\rangle. \tag{4.32}$$

Thus, the infinitesimal time-evolution operator is given by

$$U(t+dt,t) = 1 - \frac{i}{\hbar}H\,dt. \tag{4.33}$$
(infinitesimal time-evolution operator)

We can verify the above properties for this form of the evolution operator. For example,

$$U^\dagger(t+dt,t)U(t+dt,t) = \left(1 + \frac{i}{\hbar}H\,dt\right)\left(1 - \frac{i}{\hbar}H\,dt\right) = 1 + O(dt^2) = 1. \tag{4.34}$$

A similar argument works for the composition property, which gives the form of $U(t+2\,dt,t)$.

[1] For further reading, see J. J. Sakurai, *Modern Quantum Mechanics* 2nd ed. (Addison Wesley, 1993), chapter 2, p. 68.

4.2.1.2 Differential Equation for the Evolution Operator

Now using the composition property,

$$U(t+dt,t_0) = U(t+dt,t)\,U(t,t_0) = \left(1 - \frac{i}{\hbar}H\,dt\right)U(t,t_0). \tag{4.35}$$

Thus, $U(t,t_0)$, regarded as a function of t, undergoes time translation in the same way as the state vector $|\psi(t)\rangle$. In particular, then, we can write

$$\partial_t U(t,t_0) = -\frac{i}{\hbar}HU(t,t_0),$$

(Schrödinger equation for evolution operator) (4.36)

and thus we see that the evolution operator satisfies the Schrödinger equation.

4.2.1.3 General Form

Noting again that $dt^2 = 0$ for infinitesimal time increments, we can write

$$U(t+dt,t) = 1 - \frac{i}{\hbar}H\,dt = e^{-iH\,dt/\hbar}. \tag{4.37}$$

Then the composition property extends to give the general form of the evolution operator over finite time intervals,

$$U(t,t_0) = \prod_{t_0}^{t} e^{-iH(t_\alpha)\,dt_\alpha/\hbar}, \tag{4.38}$$

where the product is over all infinitesimal time intervals dt_α between t_0 and t. The product is ordered such that earlier times are to the right of later times. In the case that $H(t)$ commutes with $H(t')$ for $t \neq t'$, then we can combine the elements of the product into a single exponential, using the relation

$$e^A e^B = e^{A+B}, \tag{4.39}$$

which holds when $[A,B] = 0$. Then

$$\begin{aligned} U(t,t_0) &= \exp\left[-\frac{i}{\hbar}\sum_{t_0}^{t} H(t_\alpha)\,dt_\alpha\right] \\ &= \exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} H(t')\,dt'\right]. \end{aligned}$$

(general form, $[H(t),H(t')] = 0$) (4.40)

If the Hamiltonian does not commute with itself at different times, then we can't use this form, and we must use the form (4.38). We can use the shorthand for this notation[2]

$$U(t,t_0) = \mathscr{T}\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} H(t')\,dt'\right],$$

(general form) (4.41)

where $\mathscr{T}$ is the **chronological operator**, which indicates that the exponential is really a time-ordered product of infinitesimal time-evolution operators. On the other hand, if the Hamiltonian is time-independent, so that $H(t) = H$,

$$U(t,t_0) = \exp\left[-\frac{i}{\hbar}H(t-t_0)\right],$$

(general form, time-independent H) (4.42)

and we see that the time-evolution operator simplifies considerably, being essentially just the exponentiated Hamiltonian operator.

[2] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory*, (Pergamon Press, 1971).

4.2.2 Schrödinger vs. Heisenberg Picture

The **Schrödinger picture** is the usual scheme that you probably learned when you first studied quantum mechanics. In the Schrödinger picture, the state vector $|\psi(t)\rangle$ evolves according to the Schrödinger equation. The operators, on the other hand, are time-independent, so that time-dependent expectation values are computed by

$$\langle A(t)\rangle = \langle\psi(t)|A|\psi(t)\rangle. \tag{4.43}$$
(Schrödinger picture)

An alternate scheme, the **Heisenberg picture**, is formulated differently. In the Heisenberg picture, the time dependence is carried by the *operators*, not the state vector. The state vectors are time-*independent* here. Thus, the expectation value in the Heisenberg picture is given as

$$\langle A(t)\rangle = \langle\psi|A(t)|\psi\rangle. \tag{4.44}$$
(Heisenberg picture)

How do we transform between the two pictures? We will use the subscripts ''S'' and ''H'' for the Schrödinger and Heisenberg pictures, respectively, so that A_{S} is the Schrödinger operator, and $A_{\text{H}}(t)$ is the Heisenberg operator. Then we can use the time-evolution operator to write

$$\begin{aligned}\langle\psi(t)|A_{\text{S}}|\psi(t)\rangle &= \langle\psi(0)|U^\dagger(t,0)A_{\text{S}}U(t,0)|\psi(0)\rangle \\ &= \langle\psi(0)|A_{\text{H}}(t)|\psi(0)\rangle,\end{aligned} \tag{4.45}$$

where we have identified the transformation between pictures as

$$A_{\text{H}}(t) = U^\dagger(t,0)A_{\text{S}}U(t,0). \tag{4.46}$$
(operator transformation)

We also identify the Heisenberg-picture state vector as the initial state vector:

$$|\psi\rangle_{\text{H}} = |\psi(0)\rangle_{\text{S}} = U^\dagger(t,0)\,|\psi(t)\rangle_{\text{S}}. \tag{4.47}$$
(state transformation)

Note that the density operator ρ is a *Schrödinger-picture* operator, since it is equivalent to the state vector. Thus, the density operator transforms as

$$\rho_{\text{H}} = U^\dagger(t,0)\,\rho_{\text{S}}(t)\,U(t,0). \tag{4.48}$$
(operator transformation)

That is, in the Schrödinger picture, the density operator is time-dependent, while it is time-independent in the Heisenberg picture, which is opposite to the behavior of operators for observables.

4.2.2.1 Heisenberg Equation of Motion

In the Heisenberg picture, the operators evolve in time, so we must derive an equation of motion for Heisenberg operators. Differentiating a Heisenberg operator A and using U as shorthand for $U(t,0)$,

$$\begin{aligned}\partial_t A_{\text{H}} &= \partial_t\left[U^\dagger A_{\text{S}}U\right] \\ &= \left[\partial_t U^\dagger\right]A_{\text{S}}U + U^\dagger A_{\text{S}}\partial_t U \\ &= \frac{i}{\hbar}U^\dagger H_{\text{S}}(t)A_{\text{S}}U - \frac{i}{\hbar}U^\dagger A_{\text{S}}H_{\text{S}}(t)U \\ &= \frac{i}{\hbar}U^\dagger H_{\text{S}}(t)UU^\dagger A_{\text{S}}U - \frac{i}{\hbar}U^\dagger A_{\text{S}}UU^\dagger H_{\text{S}}(t)U,\end{aligned} \tag{4.49}$$

and finally, we can write

$$\partial_t A_{\text{H}} = -\frac{i}{\hbar}\left[A_{\text{H}}, U^\dagger H_{\text{S}}(t) U\right]. \tag{4.50}$$
(Heisenberg-picture evolution)

Note that we assumed A_{S} to be time-independent—we are still not treating any operators with *explicit* time dependence. Recall that the Hamiltonian *generates* time evolution, and so the time dependence is externally imposed. We will thus not speak of the Heisenberg-picture Hamiltonian, although we use the "S" subscript to denote that we *introduced* it in the Schrödinger picture. Note that for a time-independent Hamiltonian such that $U(t,0) = \exp(-iHt/\hbar)$, then $U(t,0)$ commutes with the Hamiltonian, and thus

$$\partial_t A_{\text{H}} = -\frac{i}{\hbar}\left[A_{\text{H}}, H\right].$$
(Heisenberg evolution, time-independent H) (4.51)

This evolution equation has a similar form to the Schrödinger–von Neumann equation (4.8) for the density operator (differing only by a minus sign).

4.2.3 Interaction Picture

The **interaction picture** is a hybrid of the Schrödinger and Heisenberg pictures. Suppose that the Hamiltonian can be decomposed as

$$H = H_0 + V(t), \tag{4.52}$$

where V is the "interaction Hamiltonian." Then the interaction picture is essentially the Schrödinger picture with respect to V, but the Heisenberg picture with respect to H_0. That is, the state vector carries the time dependence due to V, while the operators carry the time dependence due to H_0.

For concreteness and simplicity, let's assume that H_0 is time-independent. Thus, the transformation of the state vector to the interaction picture is

$$|\psi\rangle_{\text{I}} = e^{iH_0 t/\hbar}|\psi\rangle_{\text{S}}, \tag{4.53}$$
(interaction-picture state)

and the transformation for the density operator follows similarly:

$$\rho_{\text{I}} = e^{iH_0 t/\hbar}\rho_{\text{S}}(t)\, e^{-iH_0 t/\hbar}. \tag{4.54}$$
(interaction-picture state)

The operator transforms according to

$$A_{\text{I}}(t) = e^{iH_0 t/\hbar} A_{\text{S}} e^{-iH_0 t/\hbar}. \tag{4.55}$$
(interaction-picture operator)

Then the background Hamiltonian causes the operator to evolve,

$$\partial_t A_{\text{I}} = -\frac{i}{\hbar}\left[A_{\text{I}}, H_0\right], \tag{4.56}$$
(interaction-picture evolution)

while the state evolves according to the interaction Hamiltonian

$$\begin{aligned} \partial_t |\psi\rangle_{\text{I}} &= -\frac{i}{\hbar} V_{\text{I}}(t)|\psi\rangle_{\text{I}} \\ \partial_t \rho_{\text{I}} &= -\frac{i}{\hbar}[V_{\text{I}}(t), \rho_{\text{I}}], \end{aligned} \tag{4.57}$$
(interaction-picture evolution)

where $V_{\text{I}}(t)$ is the interaction-picture form of the Schrödinger-picture operator $V(t)$ in the sense of Eq. (4.55). The interaction picture is useful in perturbation theory, where the evolution due to H_0 is already known. It is thus convenient to bury this evolution in the operators, so that it is possible to focus on the perturbation Hamiltonian V.

4.3 Wigner Distribution

One important tool for this discussion is the Wigner function (or distribution), which facilitates the description of quantum dynamics in phase space. This is particularly important in comparing quantum and classical mechanics, where the analogous classical object is the phase-space (Liouville) distribution for an ensemble in phase space. The Wigner representation allows us to work with phase-space *functions*, rather than state vectors and operators. This allows for tremendous insight and simplification for certain aspects of quantum dynamics, and quite a bit of obfuscation for some others.

What do we mean by phase space? For a single-particle, classical Hamiltonian of the form

$$H(x,p) = \frac{p^2}{2m} + V(x), \tag{4.58}$$

the canonical coordinate pair (x,p) is sufficient to completely determine the state of the particle. Thus, for this system, we can define the (x,p) plane to be the **phase space**. In general, the classical phase space is the space of all generalized coordinates needed to completely specify the state of the system. For a classical Hamiltonian system of N degrees of freedom, we can generally take the phase space to be the $2N$-tuple $(x_1,\ldots,x_N,p_1,\ldots,p_N)$ of canonical coordinates.

The Wigner function is defined in terms of the density operator as[3]

$$W(x,p) := \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx'\, e^{-ipx'/\hbar}\langle x+x'/2|\rho|x-x'/2\rangle. \tag{4.59}$$
(Wigner distribution)

We can see that in terms of a *rotated* density matrix $\rho_{\mathrm{r}}(x,x') := \langle x+x'/2|\rho|x-x'/2\rangle$, the Wigner function has the form of a Fourier transform over the second variable. Since $\rho_{\mathrm{r}}(x,x') = \rho_{\mathrm{r}}^*(x,-x')$, it is clear that the Wigner function is real-valued. Note that for a pure state, the above formula reduces to

$$W(x,p) := \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx'\, e^{-ipx'/\hbar}\psi(x+x'/2)\,\psi^*(x-x'/2). \tag{4.60}$$
(Wigner distribution)

Obviously, the information content of the Wigner function is equivalent to that of the density operator, since the Wigner transform (4.59) can be inverted: the inverse Fourier transform of $W(x,p)$ over p gives the rotated density operator, which is related by a simple coordinate rotation to the usual density operator. Note that for simplicity here we are sticking to systems of one degree of freedom; the generalizations to more dimensions is reasonably straightforward.

4.3.1 Marginal Distributions

The Wigner function is not the only possible quantum phase-space distribution, but it has several features that make it preferable to other distributions. One of its most appealing properties is that each marginal distribution of the Wigner function, where one of the variables is integrated out, results in the probability distribution corresponding to the other variable. The Wigner function itself, however, is not a joint probability distribution, since it can take on negative values, which as we will see below represent the interferences or coherences of the quantum state.

To formalize this notion, we can show that when integrated over the variable p, the Wigner function yields the correct spatial probability density:

$$\int_{-\infty}^{\infty} dp\, W(x,p) = \langle x|\rho|x\rangle\,. \tag{4.61}$$
(x marginal distribution)

[3] E. Wigner, "On the Quantum Correction For Thermodynamic Equilibrium," *Physical Review* **40**, 749 (1932) (doi: 10.1103/PhysRev.40.749). For good reviews, see also Wolfgang P. Schleich, *Quantum Optics in Phase Space* (Wiley, 2001); M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, "Distribution Functions in Physics: Fundamentals," *Physics Reports* **106**, 121 (1984) (doi: 10.1016/0370-1573(84)90160-1); and V. I. Tatarskii, "The Wigner representation of quantum mechanics," *Soviet Physics Uspekhi* **26**, 311 (1983) (doi: 10.1070/PU1983v026n04ABEH004345).

To see this, just use the definition (4.59) and the integral representation of the delta function:

$$\begin{aligned}\int_{-\infty}^{\infty} dp\, W(x,p) &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx'\, \langle x+x'/2|\rho|x-x'/2\rangle \int_{-\infty}^{\infty} dp\, e^{-ipx'/\hbar} \\ &= \int_{-\infty}^{\infty} dx'\, \langle x+x'/2|\rho|x-x'/2\rangle\, \delta(x') \\ &= \langle x|\rho|x\rangle\,.\end{aligned} \tag{4.62}$$

Similarly, we can show that integration over x gives the momentum probability density:

$$\int_{-\infty}^{\infty} dx\, W(x,p) = \langle p|\rho|p\rangle\,. \tag{4.63}$$
(p marginal distribution)

To see this, we perform the following steps: insert the definition (4.59), let $x \longrightarrow x + x'/2$, let $x' \longrightarrow x' - x$, use the expressions

$$\langle p|x'\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{-ipx'/\hbar}, \qquad \langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{ipx/\hbar}, \tag{4.64}$$

and finally use the completeness relation:

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, W(x,p) &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, e^{-ipx'/\hbar}\langle x+x'/2|\rho|x-x'/2\rangle \\ &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, e^{-ipx'/\hbar}\langle x+x'|\rho|x\rangle \\ &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, e^{-ip(x'-x)/\hbar}\langle x'|\rho|x\rangle \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, \langle p|x'\rangle\langle x'|\rho|x\rangle\langle x|p\rangle \\ &= \langle p|\rho|p\rangle\,.\end{aligned} \tag{4.65}$$

From either of these arguments it is obvious that the Wigner distribution is normalized such that

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W(x,p) = 1. \tag{4.66}$$
(normalization)

The marginal property works along other axes in phase space. This follows from the property that we will show below, that the Wigner-distribution evolution in a harmonic potential is simply a rotation, just as it is for a classical phase-space distribution. Thus, these same arguments can be applied after any phase-space rotation. In fact, this acts as one method for performing a tomographic reconstruction of the Wigner function, since the marginal distributions can be experimentally measured along many axes, then converted to the Wigner function via the (inverse) Radon transform.[4]

4.3.2 Overlap

Another intuitively appealing feature of the Wigner function is that the overlap integral of two Wigner functions yields the overlap of the two corresponding density operators. We can see this for the overlap

[4] M. G. Raymer, M. Beck, and D. F. McAlister, "Complex Wave-Field Reconstruction Using Phase-Space Tomography," *Physical Review Letters* **72**, 1137 (1994) (doi: 10.1103/PhysRevLett.72.1137).

of two states represented by $W_1(x,p)$ and $W_2(x,p)$, by again using the integral representation of the delta function:

$$\begin{aligned}
&\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\,W_1(x,p)W_2(x,p)\\
&\quad=\frac{1}{(2\pi\hbar)^2}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dx''\,e^{-ip(x'+x'')/\hbar}\langle x+x'/2|\rho_1|x-x'/2\rangle\langle x+x''/2|\rho_2|x-x''/2\rangle\\
&\quad=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dx''\,\delta(x'+x'')\langle x+x'/2|\rho_1|x-x'/2\rangle\langle x+x''/2|\rho_2|x-x''/2\rangle\\
&\quad=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\langle x+x'/2|\rho_1|x-x'/2\rangle\langle x-x'/2|\rho_2|x+x'/2\rangle\,.
\end{aligned}\tag{4.67}$$

Again letting $x\longrightarrow x+x'/2$ and then $x'\longrightarrow x'-x$,

$$\begin{aligned}
\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\,W_1(x,p)W_2(x,p)&=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\langle x+x'/2|\rho_1|x-x'/2\rangle\langle x-x'/2|\rho_2|x+x'/2\rangle\\
&=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\langle x'|\rho_1|x\rangle\langle x|\rho_2|x'\rangle\\
&=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx'\,\langle x'|\rho_1\rho_2|x'\rangle\\
&=\frac{1}{2\pi\hbar}\mathrm{Tr}[\rho_1\rho_2].
\end{aligned}\tag{4.68}$$

Recall that for pure states,

$$\mathrm{Tr}[\rho_1\rho_2]=|\langle\psi_1|\psi_2\rangle|^2,\tag{4.69}$$

where the latter expression is known as the **fidelity** of two pure states. Thus, $\mathrm{Tr}[\rho_1\rho_2]$ is a generalized overlap for mixed states, and

$$2\pi\hbar\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\;W_1(x,p)\,W_2(x,p)=\mathrm{Tr}[\rho_1\rho_2]\tag{4.70}$$
(overlap integral)

represents the same overlap in terms of the Wigner functions. Note that $\mathrm{Tr}[\rho_1\rho_2]$ also represents the more general fidelity if either of ρ_1 or ρ_2 are pure, but if both states are mixed, the fidelity is defined by $F(\rho_1,\rho_2):=\max|\langle 1|2\rangle|^2$, where the maximum is taken over all "purifications" (Section 4.4.5) $|1\rangle$ of ρ_1 and $|2\rangle$ of ρ_2.[5] This definition turns out to be equivalent to the expression[6] $F(\rho_1,\rho_2)=\{\mathrm{Tr}[\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_1}}]\}^2$.

4.3.3 Area

Recall that for the density operator, we showed that $\mathrm{Tr}[\rho^2]\le 1$, with the equality holding for pure states. Using Eq. (4.70),

$$\mathrm{Tr}[\rho^2]=2\pi\hbar\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\;W^2(x,p)\le 1.\tag{4.71}$$

Inverting this relation, we find a sort of generalized uncertainty relation,

$$\left[\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\;W^2(x,p)\right]^{-1}\ge 2\pi\hbar,\tag{4.72}$$
(area theorem)

[5]The important concept of fidelity for mixed states was introduced by Richard Jozsa, "Fidelity for Mixed Quantum States," *Journal of Modern Optics* **41**, 2315 (1994) (doi: 10.1080/09500349414552171); see also Benjamin Schumacher, "Sending entanglement through noisy quantum channels," *Physical Review A* **54**, 2614 (1996) (doi: 10.1103/PhysRevA.54.2614).

[6]Richard Jozsa, *op. cit.*

again with the equality holding for pure states. Recalling that the Wigner function is a normalized distribution, along with the definition for the width $\delta\omega$ for the normalized spectral function $s(\omega)$ [Eq. (2.59)], we see that the quantity on the left-hand side is the two-dimensional generalization of the frequency width. Thus, the quantity on the left-hand side represents the *area* of the Wigner function, and is sometimes referred to as the **Süßmann measure**[7] for the Wigner function. Again, this measure of area is more ''robust'' for certain distributions than corresponding measures based on rms widths. Eq. (4.72) therefore has the nice interpretation that the Wigner function for a pure state occupies an area of h in phase space, with mixed states occupying more area.

4.3.4 Sample Wigner Distributions

To gain more intuition, since our discussion has so far been rather abstract, we will consider a few specific examples of Wigner functions.

4.3.4.1 Gaussian State

For a Gaussian state, the definition (4.59) amounts to a Fourier transform of a Gaussian function. In this case, the Wigner function is clearly Gaussian. The most general Gaussian distribution we can write down for two dimensions (centered about the origin) is[8]

$$W(x,p) = \frac{1}{2\pi\sqrt{\det(S_{\alpha\beta})}} \exp\left[-\frac{1}{2} z_\alpha \left(S^{-1}\right)_{\alpha\beta} z_\beta\right], \tag{4.73}$$
(Gaussian state)

where $(z_\alpha) = (x,p)$ and

$$(S_{\alpha\beta}) = \frac{1}{2}\left([z_\alpha, z_\beta]_+\right) = \begin{bmatrix} V_x & C_{xp} \\ C_{xp} & V_p \end{bmatrix} \tag{4.74}$$
(covariance matrix)

is the covariance matrix, whose inverse is

$$\left(S^{-1}\right)_{\alpha\beta} = \frac{1}{V_x V_p - C_{xp}^2} \begin{bmatrix} V_p & -C_{xp} \\ -C_{xp} & V_x \end{bmatrix}. \tag{4.75}$$

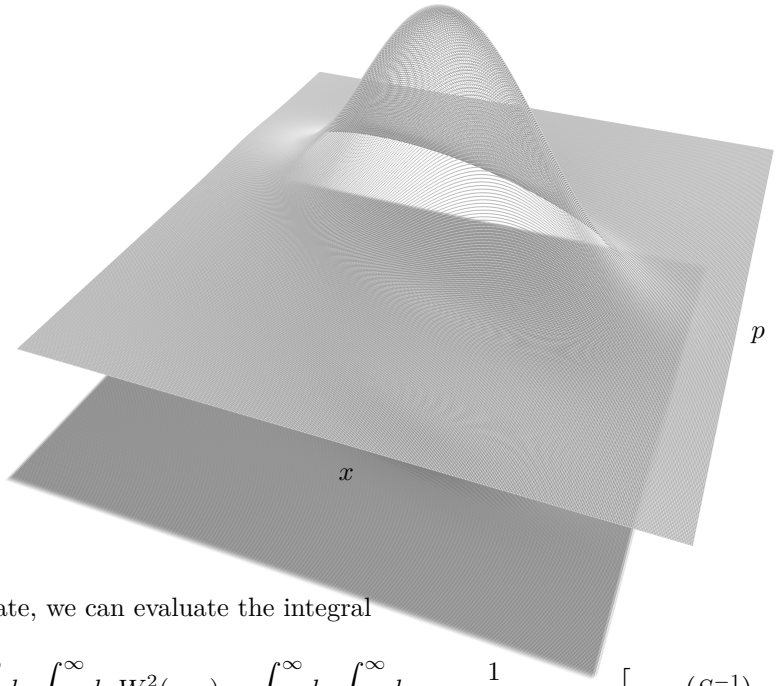
Here, $[a,b]_+ = ab + ba$ is the anticommutator bracket, $V_x = \langle x^2 \rangle$ is the position variance, $V_p = \langle p^2 \rangle$ is the momentum variance, and $C_{xp} = \langle xp + px \rangle /2$ is the covariance. This form can be derived as follows. For two *independent* Gaussian random variables, the joint distribution is just the product of two individual Gaussian distributions. Written in matrix form, this gives the above form for $W(x,p)$ where the covariance matrix is diagonal (recall that we just want the distribution with the correct marginals, so in this case classical arguments suffice.) Applying a rotation (or more generally, any linear, symplectic transformation) then gives the form where the covariance matrix is not diagonal. It is then easy to verify that this distribution is normalized and has the correct variances:

$$\begin{aligned} &\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W(x,p)\, x^2 = V_x = \langle x^2 \rangle \\ &\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W(x,p)\, p^2 = V_p = \langle p^2 \rangle \\ &\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W(x,p)\, xp = C_{xp} = \frac{1}{2}\langle xp + px \rangle . \end{aligned} \tag{4.76}$$
(Gaussian moments)

Note that in the last case, the covariance is associated with the symmetrically ordered expectation value, a point we will return to below.

[7] Wolfgang P. Schleich, *Quantum Optics in Phase Space* (Wiley, 2001).

[8] Note the slightly different form from Eq. (2.85), which is for *complex* variables.



For the Gaussian state, we can evaluate the integral

$$\begin{aligned}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W^2(x,p) &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, \frac{1}{4\pi^2 \det(S_{\alpha\beta})} \exp\left[-z_\alpha \left(S^{-1}\right)_{\alpha\beta} z_\beta\right] \\ &= \frac{1}{4\pi\sqrt{\det(S_{\alpha\beta})}} = \frac{1}{4\pi\sqrt{V_x V_p - C_{xp}^2}},\end{aligned} \tag{4.77}$$

which follows from the fact that the form of $W(x,p)$ is normalized. From Eq. (4.72), we thus find the inequality

$$V_x V_p - C_{xp}^2 \geq \frac{\hbar^2}{4}, \tag{4.78}$$
(generalized uncertainty relation)

which acts as a generalized uncertainty relation for Gaussian states, and is stronger than the usual uncertainty relation $V_x V_p \geq \hbar^2/4$, since it maintains the equality for pure Gaussian states even if they are rotated in phase space. Again, the equality holds for pure states. From Eq. (4.71), we can then see that

$$\mathrm{Tr}\left[\rho^2\right] = \frac{\hbar/2}{\sqrt{V_x V_p - C_{xp}^2}}, \tag{4.79}$$
(Gaussian-state purity)

and thus that the size of the Gaussian state is simply related to its purity.

The Gaussian state is also important in another sense. **Hudson's theorem**[9] states that the only pure states that do not take on any negative values are Gaussian, at least for systems of one degree of freedom. In this sense the Gaussian pure states are the "most classical," since they can be given a sensible classical interpretation, at least in terms of a classical probability distribution in phase space.

4.3.4.2 Coherent Superpositions

Consider the coherent superposition of two positions $\pm x_0$. To keep things physical, we will consider a superposition of two Gaussian states, each of rms width σ:

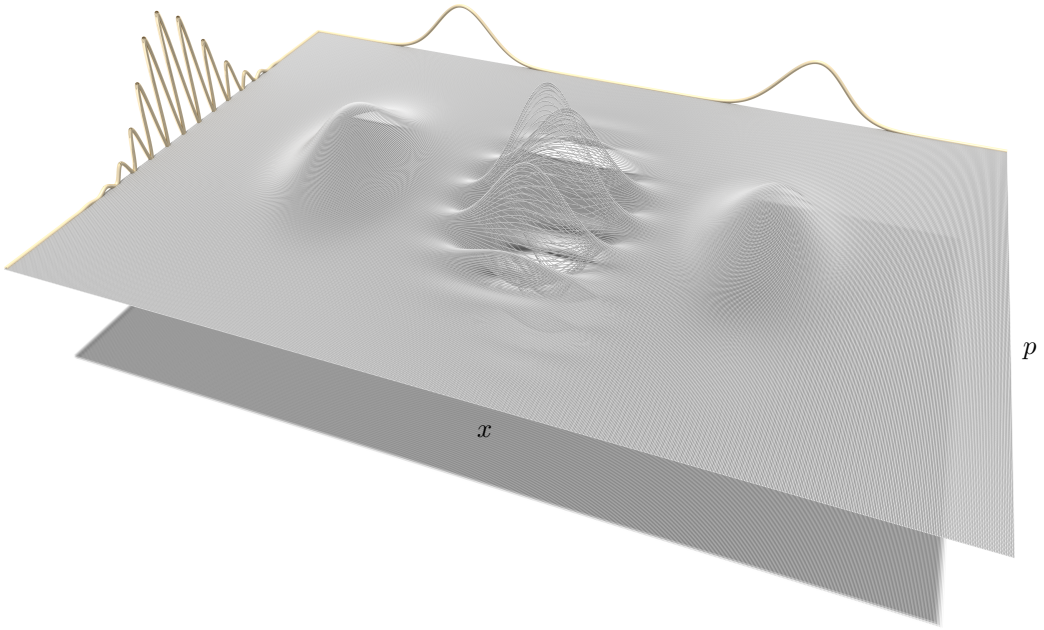
$$\psi(x) = \frac{1}{\sqrt{2}} \frac{1}{(2\pi\sigma^2)^{1/4}} \left[e^{-(x-x_0)^2/4\sigma^2} + e^{-(x+x_0)^2/4\sigma^2}\right]. \tag{4.80}$$

[9] R. L. Hudson, "When is the Wigner Quasi-Probability Density Non-Negative?" *Reports on Mathematical Physics* **6**, 249 (1974).

Putting this into the definition (4.60), we find

$$W(x,p) = \frac{1}{4\pi\hbar} e^{-2\sigma^2 p^2/\hbar^2} \left[e^{-(x-x_0)^2/2\sigma^2} + e^{-(x+x_0)^2/2\sigma^2} + 2e^{-x^2/2\sigma^2} \cos(2px_0/\hbar) \right]. \tag{4.81}$$

The first two terms in brackets are Gaussian distributions centered at $\pm x_0$, corresponding to each of the first distributions separately. The final term looks like some sort of interference term: it is a Gaussian distribution centered at the origin, but with a sinusoidal modulation in the p-direction with period $\pi\hbar/x_0$. The interpretation is that the *coherence* of the superposition is encoded in this oscillatory structure that lies between the two structures that represent the population. Indeed, had we chosen a different relative phase for the two Gaussian states, we would have simply found a different phase for the sinusoidal modulation. It is also easy to see that for an *incoherent* superposition of the two Gaussian states (i.e., the density operator being the sum of two Gaussian density operators), the Wigner function is linear in the density operator and thus the oscillatory structure would be missing.



In the wave function, coherences between components of a superposition are again ''encoded'' in the phases of the complex numbers. The Wigner function is real-valued, and thus complex phases cannot contain the same information. Instead, coherence between two phase-space regions is encoded as oscillations in the region directly between them. In this sense, negative values of the Wigner distribution are indicators of coherence.

We can also see that the oscillatory structure is necessary to recover the proper marginal distributions. The position distribution is given by integrating over p, in which case the oscillatory part vanishes, leaving just the two Gaussian states:

$$\begin{aligned} \langle x|\rho|x\rangle &= \int_{-\infty}^{\infty} dp\, W(x,p) \\ &= \frac{1}{2\sqrt{2\pi\sigma^2}} \left[e^{-(x-x_0)^2/4\sigma^2} + e^{-(x+x_0)^2/4\sigma^2} \right]^2 \\ &\approx \frac{1}{2\sqrt{2\pi\sigma^2}} \left[e^{-(x-x_0)^2/2\sigma^2} + e^{-(x+x_0)^2/2\sigma^2} \right], \end{aligned} \tag{4.82}$$

where the last expression follows through when the Gaussians are well resolved, $|x_0| \gg \sigma$, so that the overlap terms are negligible. On the other hand, the coherent superposition leads to interference fringes in

the momentum distribution (as you would expect, for example, for the far-field diffraction pattern of two Gaussian apertures):

$$\langle p|\rho|p\rangle = \int_{-\infty}^{\infty} dx\, W(x,p) = \frac{2\sigma}{\sqrt{2\pi}\hbar} e^{-2\sigma^2 p^2/\hbar^2} \cos^2(2px_0/\hbar). \tag{4.83}$$

Thus, the *orientation* of the modulation is critical: a coherent superposition of two phase-space regions implies an array of ''stripes'' pointing between the two regions, with the stripes or ridges becoming denser and more numerous as the separation increases.

4.3.4.3 Harmonic Oscillator States

For the harmonic oscillator with Hamiltonian

$$H(x,p) = \frac{p^2}{2m} + \frac{1}{2} m\omega_0^2 x^2, \tag{4.84}$$

we will see below that the quantum and classical equations of motion are identical. The classical motion of a single point particle corresponds to a closed, elliptical trajectory in phase space. This is clear from the form of the Hamiltonian, since classical trajectories correspond to surfaces of constant energy. In particular, in rescaled coordinates such that the Hamiltonian has the form

$$H'(x',p') = \frac{p'^2}{2} + \frac{x'^2}{2}, \tag{4.85}$$

the classical trajectories are circles in phase space, all rotating at the same frequency ω_0. Thus, time evolution in the rescaled harmonic-oscillator phase space is equivalent to rotation at frequency ω_0. Similarly, then, in the proper coordinates, the Wigner function simply rotates in time at frequency ω_0.

We can thus infer that in the proper coordinates, the Wigner functions for the harmonic-oscillator eigenstates must be rotationally invariant. Indeed, it can be shown that the nth eigenstate has the form[10]

$$W_n(x,p) = \frac{(-1)^n}{\pi\hbar} e^{-r^2(x,p)/\hbar} L_n\left[2r^2(x,p)/\hbar\right], \tag{4.86}$$

(harmonic-oscillator eigenstate)

where

$$r^2(x,p) = m\omega_0 x^2 + \frac{p^2}{m\omega_0}, \tag{4.87}$$

and the $L_n(x)$ are the Laguerre polynomials, given explicitly by

$$L_n(x) = \sum_{j=0}^{n} \binom{n}{j} \frac{(-x)^j}{j!}. \tag{4.88}$$

The functions $W_n(x,p)$ are thus called **Laguerre–Gaussian functions**. Naturally, the marginals of these distributions must reproduce the position and momentum distributions for the harmonic oscillator, given by

$$\begin{aligned} |\psi_n(x)|^2 &= \frac{1}{2^n n!}\sqrt{\frac{m\omega_0}{\pi\hbar}} \exp\left(-\frac{m\omega_0 x^2}{\hbar}\right) H_n^2\left(\sqrt{\frac{m\omega_0}{\hbar}}\, x\right) \\ |\phi_n(p)|^2 &= \frac{1}{2^n n!}\sqrt{\frac{1}{\pi m\omega_0\hbar}} \exp\left(-\frac{p^2}{m\omega_0\hbar}\right) H_n^2\left(\sqrt{\frac{1}{m\omega_0\hbar}}\, p\right). \end{aligned} \tag{4.89}$$

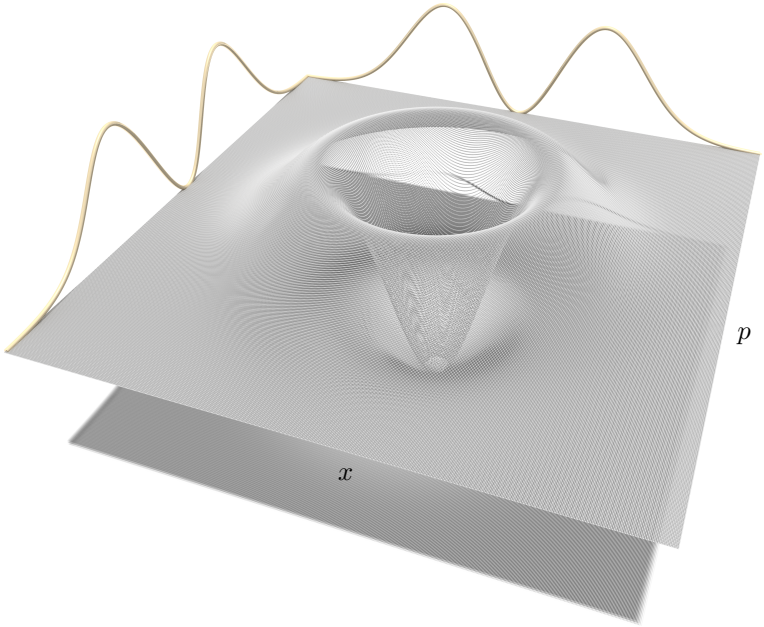
Here, the $H_n(x)$ are the Hermite polynomials, given explicitly by

$$H_n(x) = (-1)^n e^{x^2/2} \partial_x^n e^{-x^2/2}, \tag{4.90}$$

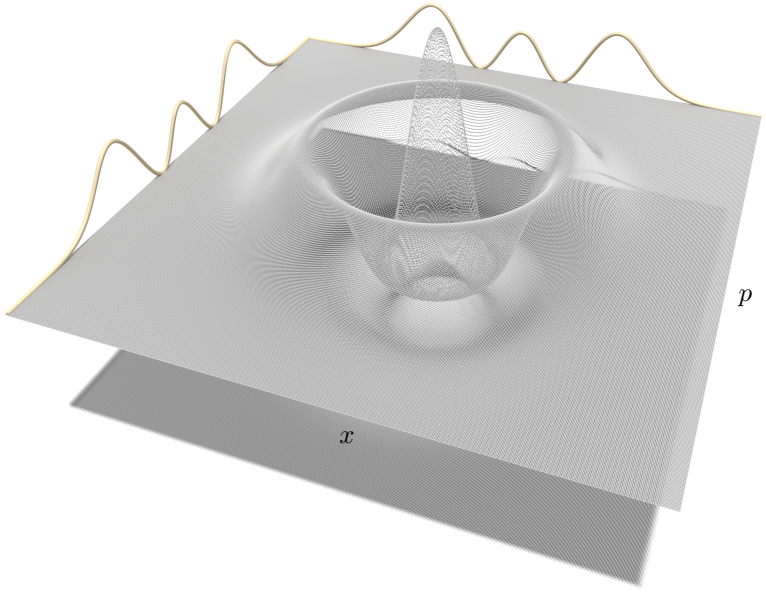
and the wave functions in x and p are the **Hermite–Gaussian functions**. Clearly, for $n > 0$, the Wigner functions must take on negative values to reproduce the zeros in the marginal distributions.

Below is a plot of the Wigner distribution and the marginals for the $n = 1$ harmonic-oscillator eigenstate,

[10] Wolfgang P. Schleich, *Quantum Optics in Phase Space* (Wiley, 2001).



and below is the corresponding plot for the $n = 2$ harmonic-oscillator eigenstate.



4.3.5 Weyl Correspondence and Operator Ordering

The Wigner representation casts quantum mechanics in terms of real-valued functions, rather than state vectors and operators. But then, how do we associate functions with operators, for example, if we want to compute expectation values in the Wigner formalism? While simple phase-space functions, such as x and p have obvious associations with the operators $\hat{x}$ and $\hat{p}$, more complicated functions such as x^2p^2 have ambiguous operator associations due to the ordering problem, where multiple possible operator orderings can correspond to the same classical function. (In this section we explicitly mark operators with hats to distinguish them from real numbers, since this is the whole point, with the exception of the density operator where the notation is already clear.) Indeed, it can be difficult to formulate a prescription for uniquely

associating phase-space functions with operators.[11]

4.3.5.1 Weyl's Rule

Consider the exponential phase-space function

$$e^{i(\pi_x x+\pi_p p)/\hbar}. \tag{4.91}$$

The notation here is intended to suggest that x and p are, in a sense, coordinates with ''conjugate'' variables π_x and π_p, respectively, because we will be constructing Fourier transforms in this way. We can choose to associate this function with the **characteristic operator**

$$\hat{M}(\pi_x,\pi_p) = e^{i(\pi_x\hat{x}+\pi_p\hat{p})/\hbar}. \tag{4.92}$$
(characteristic operator)

This association of the exponential function,

$$e^{i(\pi_x x+\pi_p p)/\hbar} \longrightarrow e^{i(\pi_x\hat{x}+\pi_p\hat{p})/\hbar}, \tag{4.93}$$
(Weyl association)

is sufficient to uniquely fix the operator association with *any* phase-space function.

Weyl's prescription[12] for the association of an arbitrary phase-space function $F(x,p)$ to an operator proceeds then as follows. The exponential function (4.91) implies the (inverse) Fourier transform relation

$$F(x,p) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} d\pi_x \int_{-\infty}^{\infty} d\pi_p\, \tilde{F}(\pi_x,\pi_p)\, e^{i(\pi_x x+\pi_p p)/\hbar}. \tag{4.94}$$

Converting the exponential factor to the characteristic operator, we find an operator related to the same Fourier-transform function $\tilde{F}(\pi_x,\pi_p)$:

$$\hat{F}(\hat{x},\hat{p}) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} d\pi_x \int_{-\infty}^{\infty} d\pi_p\, \tilde{F}(\pi_x,\pi_p)\, e^{i(\pi_x\hat{x}+\pi_p\hat{p})/\hbar}. \tag{4.95}$$

These relations uniquely define the association

$$F(x,p) \longrightarrow \hat{F}(\hat{x},\hat{p}). \tag{4.96}$$

For example, we can eliminate the intermediate function in Eq. (4.95),

$$\hat{F}(\hat{x},\hat{p}) = \frac{1}{(2\pi\hbar)^2}\int_{-\infty}^{\infty} d\pi_x \int_{-\infty}^{\infty} d\pi_p \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, F(x,p)\, e^{i[\pi_x(\hat{x}-x)+\pi_p(\hat{p}-p)]/\hbar}, \tag{4.97}$$
(Weyl's rule)

to obtain the operator explicitly in terms of the original function.

4.3.5.2 Expectation Values

Suppose we now compute the expectation value of Eq. (4.97) with respect to an arbitrary state ρ. Then we find

$$\left\langle \hat{F}(\hat{x},\hat{p}) \right\rangle = \frac{1}{(2\pi\hbar)^2}\int_{-\infty}^{\infty} d\pi_x \int_{-\infty}^{\infty} d\pi_p \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, F(x,p)\, M(\pi_x,\pi_p)\, e^{-i(\pi_x x+\pi_p p)/\hbar}, \tag{4.98}$$

where

$$M(\pi_x,\pi_p) := \left\langle \hat{M}(\pi_x,\pi_p) \right\rangle = \left\langle e^{i(\pi_x\hat{x}+\pi_p\hat{p})/\hbar} \right\rangle \tag{4.99}$$
(characteristic function)

[11] John Robert Shewell, ''On the Formation of Quantum-Mechanical Operators,'' *American Journal of Physics* **27**, 16 (1959) (doi: 10.1119/1.1934740).

[12] H. Weyl, ''Quantenmechanik und Gruppentheorie,'' *Zeitschrift für Physik* **46**, 1 (1927) (doi: 10.1007/BF02055756); Hermann Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, 1950).

is the **characteristic function**. Using a special case of the **Baker–Campbell–Hausdorff (BCH) expansion**,[13] which states that

$$\exp(A+B) = \exp(A)\exp(B)\exp\left(-\frac{1}{2}[A,B]\right) = \exp(B)\exp(A)\exp\left(\frac{1}{2}[A,B]\right) \tag{4.100}$$

if $[A,[A,B]] = [B,[A,B]] = 0$, we can see that

$$e^{i(\pi_x\hat{x}+\pi_p\hat{p})/\hbar} = e^{i\pi_p\hat{p}/2\hbar}e^{i\pi_x\hat{x}/\hbar}e^{i\pi_p\hat{p}/2\hbar}, \tag{4.101}$$

since $[x,p] = i\hbar$. Then we can write the characteristic function as

$$\begin{aligned} M(\pi_x,\pi_p) &= \mathrm{Tr}\left[e^{i\pi_p\hat{p}/2\hbar}e^{i\pi_x\hat{x}/\hbar}e^{i\pi_p\hat{p}/2\hbar}\rho\right] \\ &= \int_{-\infty}^{\infty} dx\, e^{i\pi_x x/\hbar}\langle x|e^{i\pi_p\hat{p}/2\hbar}\rho e^{i\pi_p\hat{p}/2\hbar}|x\rangle \\ &= \int_{-\infty}^{\infty} dx\, e^{i\pi_x x/\hbar}\langle x+\pi_p/2|\rho|x-\pi_p/2\rangle \end{aligned} \tag{4.102}$$

In the last step, we use the shifting property of the exponential factors:

$$\begin{aligned} e^{i\hat{p}x_0/\hbar}|x\rangle &= \int_{-\infty}^{\infty} dp'\, e^{i\hat{p}x_0/\hbar}|p'\rangle\langle p'|x\rangle \\ &= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dp'\, e^{-ip'(x-x_0)/\hbar}|p'\rangle \\ &= \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} dx'\int_{-\infty}^{\infty} dp'\, e^{-ip'(x-x_0)/\hbar}|x'\rangle\langle x'|p'\rangle \\ &= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx'\int_{-\infty}^{\infty} dp'\, e^{-ip'(x-x_0-x')/\hbar}|x'\rangle \\ &= \int_{-\infty}^{\infty} dx'\,\delta(x-x_0-x')|x'\rangle \\ &= |x-x_0\rangle. \end{aligned} \tag{4.103}$$

Eq. (4.102) has the form of an inverse Fourier transform, which we can happily invert, with the result

$$\langle x+\pi_p/2|\rho|x-\pi_p/2\rangle = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} d\pi_x\, M(\pi_x,\pi_p)\, e^{-i\pi_x x/\hbar}. \tag{4.104}$$

Now applying $(2\pi\hbar)^{-1}\int d\pi_p\, e^{-i\pi_p p/\hbar}$ to both sides, we find an expression for the Wigner function in terms of the characteristic function:

$$W(x,p) = \frac{1}{(2\pi\hbar)^2}\int_{-\infty}^{\infty} d\pi_x\int_{-\infty}^{\infty} d\pi_p\, M(\pi_x,\pi_p)\, e^{-i(\pi_x x+\pi_p p)/\hbar}. \tag{4.105}$$
(alternate definition)

That is, the Wigner function is, up to a constant factor, just the Fourier transform of the characteristic function. Essentially, we have just motivated the definition of the Wigner function as the Weyl correspondence of the density operator.

Using this expression to simplify Eq. (4.98), we find

$$\left\langle \hat{F}(\hat{x},\hat{p})\right\rangle = \int_{-\infty}^{\infty} dx\int_{-\infty}^{\infty} dp\, F(x,p)\, W(x,p). \tag{4.106}$$
(operator expectation value)

[13] R. M. Wilcox, "Exponential Operators and Parameter Differentiation in Quantum Physics," *Journal of Mathematical Physics* **8**, 962 (1967) (doi: 10.1063/1.1705306).

Thus we find another intuitively appealing result, that the expectation values of operators are given by an overlap integral of the corresponding phase-space function with the Wigner distribution. This relation further cements the analogy of the Wigner distribution with a joint probability density. Note that the expectation value here assumes the particular ordering implied by Weyl's rule. This turns out to be a symmetrized ordering, which we will explore more carefully below.

4.3.5.3 Weyl Correspondence: Inverse Form

In Eq. (4.97), we gave the operator $\hat{F}(\hat{x},\hat{p})$ in explicitly in terms of the function $F(x,p)$, but we did not give the inverse relation. We do so here, as a convenient prescription for obtaining the phase-space function from the operator. We start by writing the expectation value as

$$\begin{aligned}\left\langle \hat{F}(\hat{x},\hat{p})\right\rangle &= \mathrm{Tr}\left[\hat{F}(\hat{x},\hat{p})\rho\right] \\ &= \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\mathrm{Tr}\left[\hat{F}(\hat{x},\hat{p})|x'\rangle\langle x'|\rho|x\rangle\langle x|\right] \\ &= \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\langle x|\hat{F}(\hat{x},\hat{p})|x'\rangle\langle x'|\rho|x\rangle \\ &= \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dx'\,\langle x-x'/2|\hat{F}(\hat{x},\hat{p})|x+x'/2\rangle\langle x+x'/2|\rho|x-x'/2\rangle.\end{aligned} \tag{4.107}$$

In the last step, we performed the usual trick of letting $x' \longrightarrow x'+x$ and then $x \longrightarrow x-x'/2$. We can compare this expression with Eq. (4.106), which gives

$$\left\langle \hat{F}(\hat{x},\hat{p})\right\rangle = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\int_{-\infty}^{\infty}dx'\,F(x,p)\langle x+x'/2|\rho|x-x'/2\rangle e^{-ipx'/\hbar}. \tag{4.108}$$

If both relations are to hold for *any* density operator ρ, then we may identify the two integrands:

$$\langle x-x'/2|\hat{F}(\hat{x},\hat{p})|x+x'/2\rangle = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}dp\,F(x,p)e^{-ipx'/\hbar}. \tag{4.109}$$

Inverting this relation and letting $x' \longrightarrow -x'$, we find

$$F(x,p) = \int_{-\infty}^{\infty}dx'\,\langle x+x'/2|\hat{F}(\hat{x},\hat{p})|x-x'/2\rangle\, e^{-ipx'/\hbar}. \tag{4.110}$$
(Weyl correspondence)

Thus, we see how an arbitrary operator transforms to a phase-space function, and we see that this transformation also motivates the original form for the Wigner transform (4.59).

4.3.5.4 Weyl Ordering

As we noted above, the operator ordering implied by Weyl's rule is a symmetrized ordering. Again, the association (4.93) gives

$$e^{i(\pi_x x+\pi_p p)/\hbar} \longrightarrow e^{i\pi_x\hat{x}/2\hbar}e^{i\pi_p\hat{p}/\hbar}e^{i\pi_x\hat{x}/2\hbar} \tag{4.111}$$

after applying the BCH formula (4.100). Then noting that

$$e^{i(\pi_x x+\pi_p p)/\hbar} = \sum_{n=0}^{\infty}\frac{1}{n!}\left[\frac{i}{\hbar}(\pi_x x+\pi_p p)\right]^n = \sum_{n=0}^{\infty}\left(\frac{i}{\hbar}\right)^n\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}(\pi_x x)^k(\pi_p p)^{n-k}, \tag{4.112}$$

we can match the coefficients of $\pi_x^k\pi_p^{n-k}$ to find the correspondence

$$\frac{1}{n!}\binom{n}{k}x^k p^{n-k} \longrightarrow \sum_{m=0}^{k}\frac{1}{m!(n-k)!(k-m)!}\left(\frac{\hat{x}}{2}\right)^m\hat{p}^{n-k}\left(\frac{\hat{x}}{2}\right)^{k-m} \tag{4.113}$$

after expansion of the operator exponentials. Dividing through by the factorials on the left-hand side, this simplifies to

$$x^k p^{n-k} \longrightarrow \frac{1}{2^k} \sum_{m=0}^{k} \binom{k}{m} \hat{x}^m \hat{p}^{n-k} \hat{x}^{k-m}, \tag{4.114}$$

and letting $l = n - k$, we find the explicit Weyl ordering

$$x^k p^l \longrightarrow \frac{1}{2^k} \sum_{m=0}^{k} \binom{k}{m} \hat{x}^m \hat{p}^l \hat{x}^{k-m}. \tag{4.115}$$

(Weyl ordering rule)

Here we explicitly see the symmetric nature of the ordering. In particular, we have shown that the characteristic function gives all the symmetrically ordered moments

$$\left\langle (\hat{x}^k \hat{p}^l)_{\rm w} \right\rangle = (-i)^{k+l} \, \partial_{\pi_x}^k \partial_{\pi_p}^l M(\pi_x, \pi_p)\Big|_{\pi_x=0, \pi_p=0}, \tag{4.116}$$

(moment formula)

where $(\hat{x}^k \hat{p}^l)_{\rm w}$ denotes the symmetrized Weyl ordering of Eq. (4.115).

It is somewhat inconvenient to order any particular function by expanding it, applying the above rule term-by-term, and then resumming it. For general functions, it is easier to use **McCoy's formula**,[14] which says that if $\hat{F}_{\rm std}(\hat{x}, \hat{p})$ is an operator with "standard" ordering, having all $\hat{x}$'s written to the left and all $\hat{p}$'s written to the right, then the corresponding operator $\hat{F}_{\rm w}(\hat{x}, \hat{p})$ with Weyl-ordering operator is obtained by

$$\hat{F}_{\rm w}(\hat{x}, \hat{p}) = \exp\left(-\frac{i\hbar}{2} \frac{\partial^2}{\partial \hat{x} \partial \hat{p}} \right) \hat{F}_{\rm std}(\hat{x}, \hat{p}). \tag{4.117}$$

(McCoy's formula)

The orders of the factors must obviously be preserved in the differentiation.

4.3.6 Operator Products and Commutators

Consider the operator product

$$\hat{A} = \hat{B}\hat{C}. \tag{4.118}$$

How does the product go over to the Wigner representation? Using the Wigner correspondence in the form (4.110), we find that the operator product implies

$$A(x,p) = B(x,p) \exp\left[\frac{\hbar}{2i} \left(\overleftarrow{\partial_p} \overrightarrow{\partial_x} - \overleftarrow{\partial_x} \overrightarrow{\partial_p} \right) \right] C(x,p),$$

(Weyl product correspondence) (4.119)

where the arrows on the derivative operators indicate the direction of operation. Mostly, the Wigner function has given us relatively simple and intuitively appealing results. However, as we see now, the complexities we have hidden thus far start to become more obvious when looking at operator products.

The proof of this correspondence is left as an exercise (Problem 4.3), but the outline of the derivation is as follows. First, note that Eq. (4.95) gives the operator matrix elements as

$$\langle x' | \hat{F} | x'' \rangle = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} d\pi_x \, \tilde{F}(\pi_x, x'' - x') e^{i\pi_x (x' + x'')/2\hbar}. \tag{4.120}$$

This relation can then be used in the correspondence equation (4.110), and then the derivation carries through in essentially the same way as the others above.

[14] Neal H. McCoy, "On the Function in Quantum Mechanics which Corresponds to a Given Function in Classical Mechanics," *Proceedings of the National Academy of Sciences* (18), 674 (1932). See also John Robert Shewell, "On the Formation of Quantum-Mechanical Operators," *American Journal of Physics* **27**, 16 (1959) (doi: 10.1119/1.1934740).

An alternate form of the operator product is

$$A(x,p) = B\left(x - \frac{\hbar}{2i}\partial_p, p + \frac{\hbar}{2i}\partial_x\right) C(x,p) = C\left(x + \frac{\hbar}{2i}\partial_p, p - \frac{\hbar}{2i}\partial_x\right) B(x,p).$$

(Weyl product correspondence) (4.121)

These expressions follow from an argument similar to the one for the original form (4.119) (Problem 4.3).

These expressions for the product then give the following correspondence for the commutator:

$$[\hat{A}, \hat{B}] \longrightarrow \frac{2}{i} A(x,p) \sin\left[\frac{\hbar}{2}\left(\overleftarrow{\partial_p}\overrightarrow{\partial_x} - \overleftarrow{\partial_x}\overrightarrow{\partial_p}\right)\right] B(x,p).$$

(Weyl commutator correspondence) (4.122)

With the alternate forms (4.121), we can also write

$$[\hat{A}, \hat{B}] \longrightarrow \left[A\left(x - \frac{\hbar}{2i}\partial_p, p + \frac{\hbar}{2i}\partial_x\right) - A\left(x + \frac{\hbar}{2i}\partial_p, p - \frac{\hbar}{2i}\partial_x\right)\right] B(x,p)$$

(Weyl commutator correspondence) (4.123)

as an alternate correspondence. While the other properties of the Wigner function make it intuitively appealing, we can see that when operator products are involved, the situation becomes substantially more complicated.

4.3.7 Moyal Bracket

Finally, based on what we now know, it is straightforward to obtain the equation of motion for the Wigner function. Recalling the Schrödinger–von Neumann equation (4.8), we note by comparing the definition (4.59) to the Weyl correspondence (4.110), we see that the Weyl correspondence for the density operator reads

$$\frac{1}{2\pi\hbar}\rho \longrightarrow W(x,p).$$

(Weyl quantum-state correspondence) (4.124)

Using the commutator correspondence (4.122), we find that the Schrödinger–von Neumann equation (4.8) becomes

$$\begin{aligned}\partial_t W(x,p) &= -\frac{2}{\hbar} H(x,p) \sin\left[\frac{\hbar}{2}\left(\overleftarrow{\partial_p}\overrightarrow{\partial_x} - \overleftarrow{\partial_x}\overrightarrow{\partial_p}\right)\right] W(x,p) \\ &=: \{H, W\}_{\text{M}}.\end{aligned}$$

(Moyal bracket) (4.125)

The final abbreviation $\{A,B\}_{\text{M}}$ is the **Moyal bracket**,[15] so named to emphasize the connection to the classical **Poisson bracket**

$$\{A,B\}_{\text{P}} := (\partial_x A)(\partial_p B) - (\partial_p A)(\partial_x B).$$

(Poisson bracket) (4.126)

For a particle Hamiltonian in "standard form," $H = p^2/(2m) + V(x)$, the Moyal bracket can be written as the Poisson bracket plus quantum "correction" terms:

$$\partial_t W = \{H, W\}_{\text{P}} + \sum_{n=1}^{\infty} \frac{(-1)^n \hbar^{2n}}{2^{2n}(2n+1)!} (\partial_x^{2n+1} V)(\partial_p^{2n+1} W).$$

(Moyal bracket for particle Hamiltonian) (4.127)

This equation is especially suitable for comparing the quantum evolution with the evolution of a classical ("Liouville") distribution ρ_{L},

$$\partial_t \rho_{\text{L}}(x,p) = \{H, \rho_{\text{L}}\}_{\text{P}},$$

(Liouville equation) (4.128)

[15] J. E. Moyal, "Quantum Mechanics as a Statistical Theory," *Proceedings of the Cambridge Philosophical Society* **45**, 99 (1949).

which is described only by the Poisson bracket. This equation of motion follows from the classical expansion for a general phase-space function,

$$\begin{aligned}\frac{df(x,p,t)}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial p}\frac{dp}{dt} + \frac{\partial f}{\partial t} \\ &= \frac{\partial f}{\partial x}\frac{\partial H}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial H}{\partial x} + \frac{\partial f}{\partial t} \\ &= \{f, H\}_{\text{P}} + \frac{\partial f}{\partial t},\end{aligned} \tag{4.129}$$

along with Liouville's theorem, which says that $(d/dt)\rho_{\text{L}} = 0$, since $\rho_{\text{L}}(x(t), p(t), t)$ is an invariant along any classical trajectory. Thus we see that the usual analogy between quantum and classical mechanics[16] of

$$\frac{1}{i\hbar}[\hat{A}, \hat{B}] \longrightarrow \{A, B\}_{\text{P}}, \tag{4.130}$$

seen by comparing the Liouville equation (4.128) to the Heisenberg equation of motion (4.51), is made explicit via the Moyal bracket.

Notice that formally setting $\hbar = 0$ in (4.127) recovers the Liouville evolution (4.128), so that correspondence seems easy in this formulation; however, it must be emphasized that taking the limit $\hbar \to 0$ for a quantum system is not trivial and is not, in general, well defined due to the singular nature of the limit.

It is immediately clear from the form of the Moyal bracket (4.127) that quantum-classical correspondence is particularly simple for linear systems, such as the free particle and the harmonic oscillator, because the quantum-correction terms vanish. This yields identical quantum and classical evolution equations for the harmonic oscillator. This point was recognized early on by Schrödinger, when he constructed the coherent states of the harmonic oscillator that mimic the classical oscillating trajectories.[17] This is a critically important point, and so I'll repeat it: **for harmonic oscillators, the quantum and classical evolution equations are equivalent**. Thus, all quantum effects in the harmonic oscillator are *only* in the initial condition. It is only in *nonlinear* potentials that the dynamical evolution generates quantum effects.

4.3.8 Summary: Defining Properties

We conclude our discussion of the Wigner distribution with a summary of the properties that define it. It turns out that the following five properties are sufficient to uniquely define the Wigner function:[18]

1. $W(x,p)$ is a Hermitian, bilinear form of the state vector $|\psi\rangle$, so that $W(x,p) = \langle\psi|\hat{W}(x,p)|\psi\rangle$, where $\hat{W}$ is a Hermitian operator depending on x and p. This implies $W(x,p)$ is real.

2. $W(x,p)$ is normalized and produces the correct marginal probability densities of x and p.

3. The definition of $W(x,p)$ exhibits Galilean invariance: the replacement $\psi(x) \longrightarrow \psi(x + x_0)$ implies $W(x,p) \longrightarrow W(x + x_0, p)$ for translations, and $\psi(x) \longrightarrow e^{ip_0 x/\hbar}\psi(x)$ implies $W(x,p) \longrightarrow W(x, p + p_0)$ for boosts.

4. The definition of $W(x,p)$ is invariant under the reflections $x \longrightarrow -x$ and $t \longrightarrow -t$. Mathematically, this means that $\psi(x) \longrightarrow \psi(-x)$ implies $W(x,p) \longrightarrow W(-x,p)$ for space reflections, and $\psi(x) \longrightarrow \psi^*(x)$ implies $W(x,p) \longrightarrow W(x,-p)$ for time reflections.

5. The equation of motion for $W(x,p)$ is the classical one in the case of the free particle.

We have discussed most of these properties already, and the others are easy to see. The extra condition

[16] See P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th revised ed. (Oxford, 1967), § 21, p. 84.

[17] E. Schrödinger, "Der stetige Übergang von der Mikro- zur Makromechanik," *Naturwissenschaften* **14**, 664 (1926) (doi: 10.1007/BF01507634).

[18] M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, "Distribution Functions in Physics: Fundamentals," *Physics Reports* **106**, 121 (1984) (doi: 10.1016/0370-1573(84)90160-1), and references therein.

6. The squared modulus of the overlap of two states is given by $2\pi\hbar$ times the phase-space overlap integral of the corresponding Wigner functions, as in Eq. (4.70).

gives an alternate set of conditions sufficient to fix the form of the Wigner function, obtained by replacing condition 5 with this last condition.

4.3.9 Other Representations

Naturally, other choices regarding the desirable properties of quantum phase-space distributions lead to other distributions. We will consider a few distributions, some of which have some utility in quantum optics, but we will consider them only briefly.

4.3.9.1 Husimi or *Q* Function

Consider the following Wigner function for a *pure* Gaussian state, centered at $(0,0)$, from Eq. (4.73):

$$W_{V_x,V_p}(x,p) = \frac{1}{2\pi\sqrt{V_xV_p}} \exp\left[-\left(\frac{x^2}{2V_x} + \frac{p^2}{2V_p}\right)\right]. \tag{4.131}$$

Note that here we have assumed a covariance $C_{xp} = 0$ for simplicity, and thus $V_xV_p = \hbar^2/4$. The **Husimi Distribution**[19] for an arbitrary state ρ with Wigner distribution $W(x,p)$ is then given by the convolution of the Wigner function corresponding to ρ with the Gaussian Wigner function:

$$\begin{aligned} W_{\text{H}}(x,p) &:= \Big(W * W_{V_xV_p}\Big)(x,p) \\ &= \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dp'\; W(x',p')\, W_{V_x,V_p}(x-x', p-p'). \end{aligned} \tag{4.132}$$

(Husimi distribution)

From the overlap relation (4.70), we can see that the Husimi distribution has the following interpretation: $W_{\text{H}}(x',p')$ represents the projection (up to a factor of $2\pi\hbar$) of ρ into the Gaussian state represented by $W_{V_x,V_p}(x-x', p-p')$ [i.e., the Gaussian state centered at (x,p)]. Thus, the Husimi distribution represents a projection into an overcomplete basis of displaced Gaussian states.

Because of the above projection interpretation, it follows that the Husimi distribution is everywhere positive. It also is a ''smoothed'' version of the Wigner function, and can be useful for visualizing quantum states in phase space, especially in cases where interferences cause complicated oscillations in the Wigner distribution. The flip side of this is that the Husimi distribution tends to ''hide'' the quantum nature of certain states; this argument has been used both for and against it for comparing quantum states to classical distributions. The Husimi function also has neither the correct marginals nor a simple equation of motion.

In terms of ordering, the Husimi function corresponds to *anti-normal ordering*: that is, if $\hat{x}$ and $\hat{p}$ are decomposed into the creation and annihilation operators for the harmonic oscillator ($\hat{a}^\dagger$ and $\hat{a}$, respectively), then the ordering is such that all annihilation operators are to the left of the creation operators. In the case of a harmonic oscillator, and the variances V_x and V_p of the convolution kernel are chosen to match the ground state (i.e., $W_{V_x,V_p}(x'-x, p'-p)$ is a displaced ground state or a **coherent state**) the Husimi distribution reduces to the ***Q* function**. In this case, the distribution is connected with the operator association[20]

$$e^{z\alpha^* - z^*\alpha} \longrightarrow e^{-z^*\hat{a}} e^{z\hat{a}^\dagger}, \tag{4.133}$$

in place of Eq. (4.93) for the Wigner distribution. Here, α is the eigenvalue of $\hat{a}$, since it turns out that the coherent states are eigenstates of the annihilation operator.

[19] K. Husimi, *Proceedings of the Physico-Mathematical Society of Japan* **22**, 264 (1940).

[20] Hai-Woong Lee, ''Theory and Application of the Quantum Phase-Space Distribution Functions,'' *Physics Reports* **259**, 147 (1995) (doi: 10.1016/0370-1573(95)00007-4).

4.3.9.2 *P* Function

The **Glauber–Sudarshan *P* function**, on the other hand, corresponds to a normal ordering, where all the annihilation operators are written to the right. The formal association rule is thus

$$e^{z\alpha^* - z^*\alpha} \longrightarrow e^{z\hat{a}^\dagger} e^{-z^*\hat{a}}. \tag{4.134}$$

This correspondence is sufficient to define the function as well as the operator associations, as in Eqs. (4.94) and (4.95), though we must first equate $\exp(z\alpha^* - z^*\alpha) = \exp[i(\pi_x x + \pi_p p)/\hbar]$ to identify the z variable and thus perform the integration.

4.3.9.3 Standard-Ordered Distribution

The standard ordering gives the exponential association

$$e^{i(\pi_x x + \pi_p p)/\hbar} \longrightarrow e^{i\pi_x \hat{x}/\hbar} e^{i\pi_p \hat{p}/\hbar}, \tag{4.135}$$

which gives rise to the **standard-ordered distribution**. It turns out that the corresponding distribution can be written as[21]

$$W_{\text{S}}(x,p) := \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dx' e^{-ipx'/\hbar} \langle x + x'|\rho|x\rangle, \tag{4.136}$$

which follows from an argument similar to the one embodied by Eq. (4.102), in analogy with the definition (4.59) of the Wigner distribution.

4.3.9.4 Antistandard-Ordered Distribution

Finally, the antistandard ordering corresponds to

$$e^{i(\pi_x x + \pi_p p)/\hbar} \longrightarrow e^{i\pi_p \hat{p}/\hbar} e^{i\pi_x \hat{x}/\hbar}, \tag{4.137}$$

giving rise to the **antistandard-ordered distribution**, also known as the the **Kirkwood distribution** or the **Rihaczek distribution**. It turns out that this distribution can be written as[22]

$$W_{\text{A}}(x,p) := \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} dx' e^{-ipx'/\hbar} \langle x|\rho|x - x'\rangle. \tag{4.138}$$

This distribution is the complex conjugate of the standard-ordered distribution.

4.4 Multiple Degrees of Freedom

4.4.1 Merging Hilbert Spaces

Suppose two degrees of freedom are prepared in two quantum states completely independently of each other. This could happen, say, for two particles prepared in separate, distant galaxies. We will refer to the two degrees of freedom as ''particles,'' even though they could correspond to different degrees of freedom of the *same* system, such as the spin and center-of-mass position of an atom, or the spin and spatial profile of a photon.

Labeling the two particles as A and B, if the individual states of the particles are $|\psi\rangle_A$ and $|\psi\rangle_B$, then we can write the composite state as

$$|\psi\rangle = |\psi\rangle_A \otimes |\psi\rangle_B, \tag{4.139}$$

where $\otimes$ denotes the **tensor product** (or **direct product**). Often, this is product is written without an explicit tensor-product symbol:

$$|\psi\rangle_A \otimes |\psi\rangle_B \equiv |\psi\rangle_A |\psi\rangle_B \equiv |\psi_A\, \psi_B\rangle. \tag{4.140}$$

[21] Ibid.
[22] Ibid.

The particle labels can even be dropped, since the ordering determines which state applies to which particle.

We can also see the meaning of the tensor product in component form. Let each separate state be expressed in an orthonormal basis as

$$|\psi\rangle_A = \sum_\alpha c_\alpha^{(A)} |\alpha\rangle_A, \qquad |\psi\rangle_B = \sum_\beta c_\beta^{(B)} |\beta\rangle_B. \tag{4.141}$$

Then we can express the composite state as

$$|\psi\rangle = \sum_{\alpha\beta} c_{\alpha\beta} |\alpha_A\, \beta_B\rangle, \tag{4.142}$$

where

$$c_{\alpha\beta} = c_\alpha^{(A)} c_\beta^{(B)}. \tag{4.143}$$

Note that $c_{\alpha\beta}$ is still understood to be a *vector*-like object, with a single index. Thus, there is an implicit (bijective) mapping of the ordered index pair (α, β) to a single index, which we simply denote as $\alpha\beta$.

Similarly, we can write a density operator for two independent particles by the same tensor product:

$$\rho = \rho^{(A)} \otimes \rho^{(B)}. \tag{4.144}$$

We can also write this in component form for the density matrices as

$$\rho_{\alpha\mu\beta\nu} = \rho_{\alpha\beta}^{(A)} \rho_{\mu\nu}^{(B)}, \tag{4.145}$$

where again $\alpha\mu$ and $\beta\nu$ are to be taken as composite indices.

The same tensor-product notation applies to Hilbert spaces. That is, we can write

$$|\psi_A\, \psi_B\rangle \in \mathscr{H}_A \otimes \mathscr{H}_B \tag{4.146}$$

if $|\psi\rangle_A \in \mathscr{H}_A$ and $|\psi\rangle_B \in \mathscr{H}_B$.

4.4.2 Entanglement

The above composite states, described by tensor products of separated states, are called **separable states**. However, not all states are separable, and those that are not separable are called **entangled**. In some sense, entanglement is the ''most quantum'' of all quantum effects.

Thus, we can see that a composite state $|\psi\rangle$ is entangled if and only if it *cannot* be written in the separable form

$$|\psi\rangle = |\psi\rangle_A \otimes |\psi\rangle_B. \tag{4.147}$$

The definition for density operators is somewhat more general: a density operator for a composite system is separable if and only if it can be written in the form

$$\rho = \sum_\alpha P_\alpha \rho_\alpha^{(A)} \otimes \rho_\alpha^{(B)}. \tag{4.148}$$

Unfortunately, given an arbitrary mixed density operator, it is difficult to tell if it corresponds to an entangled state (in fact, this turns out to be an NP-hard problem).

The point is that two entangled systems do not have local states that can be treated independently. This is in conflict with the apparently reasonable assumption of **local realism**, which states that distant systems should have independent, observer-independent realities (in particular, they should not directly influence each other). Herein lies the importance of the famous Bell-type inequalities and the experimental verification that quantum mechanics violates the inequalities: local realism contradicts quantum mechanics. Specifically, quantum mechanics is inherently nonlocal because of entanglement.

4.4.2.1 Cloning

With the language of entanglement, it is relatively simple to demonstrate the **no-cloning theorem**,[23] which says that the state of a *single* quantum system cannot be copied to another particle. This turns out to be a simple consequence of unitary evolution.

Let's examine just a simple case. Suppose that cloning is possible on a two-state system from particle A to particle B. Particle B must be in a particular state to begin with, and without loss of generality we may take this to be the "0" state. Then to copy the eigenstates of A, we see that there must be a unitary transformation U such that

$$U|0\rangle_A|0\rangle_B = |0\rangle_A|0\rangle_B, \qquad U|1\rangle_A|0\rangle_B = |1\rangle_A|1\rangle_B. \tag{4.149}$$

However, if particle A is in the superposition state

$$|\psi\rangle_A = \frac{1}{\sqrt{2}}\Big(|0\rangle_A + |1\rangle_A\Big), \tag{4.150}$$

then we see that the cloning operator gives

$$U|\psi\rangle_A|0\rangle_B = \frac{1}{\sqrt{2}}\Big(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B\Big), \tag{4.151}$$

which is an entangled state (a **Bell state**). However, what we wanted for cloning to work properly is the *separable* state

$$U|\psi\rangle_A|0\rangle_B = \frac{1}{2}\Big(|0\rangle_A + |1\rangle_A\Big)\Big(|0\rangle_B + |1\rangle_B\Big). \tag{4.152}$$

We can see that the problem in this particular example is that U acts nonlocally, and thus induces entanglement between the two particles. In fact, the **controlled-NOT** (CNOT) gate is a quantum operation that effects the transformations in Eqs. (4.149) (if A and B are in eigenstates, the CNOT flips the state of system B if and only if system A is in state 1).

Of course, it *is* possible to clone a state if you already know everything about it (i.e., you have classical knowledge of the state), or if you have an infinite ensemble of copies. (Copying a state is possible to within some fidelity tolerance for a finite ensemble of copies.) In this case, enough measurements may be made to reconstruct the state of the system arbitrarily well, and of course this procedure does not correspond to a unitary transformation. The problem with the *single* system is that in general, a measurement of the system destroys its state, and a single measurement is not enough to determine the state of the system. Of course, there is no problem with the cloning of the *basis* states, as in Eqs. (4.149); the problem is in cloning general states that are not orthogonal to the basis states. In particular this means that with a bit of *extra* information beyond what is contained in the quantum state (e.g., the state of particle A is either $|0\rangle_A$ or $|1\rangle_A$, but not any coherent superposition of the two), cloning may in fact be possible.

4.4.3 Peres–Horodecki Criterion

Given an arbitrary density operator, how do we tell if it corresponds to an entangled or separable state? As we mentioned above, this is a very difficult problem in general. Nevertheless, we can briefly discuss one important criterion for separability. The **Peres–Horodecki** or **positive partial transpose** (PPT) criterion[24] starts with the following observation. Suppose that we have a separable, composite state for two systems of the form

$$\rho = \rho^{(A)} \otimes \rho^{(B)}. \tag{4.153}$$

[23] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature* **299**, 802 (1982); D. Dieks, "Communication by EPR devices," *Physics Letters A* **92**, 271 (1982).

[24] Asher Peres, "Separability Criterion for Density Matrices," *Physical Review Letters* **77**, 1413 (1996). Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki, "Separability of Mixed States: Necessary and Sufficient Conditions," *Physics Letters A* **223**, 1 (1996).

Now suppose that we take the transpose of one subsystem, say B, to obtain a new "density operator:"

$$\tilde{\rho} = \rho^{(A)} \otimes \left(\rho^{(B)}\right)^{\mathsf{T}}. \tag{4.154}$$

Note here that for the transposition operation, $\left(\rho^{(B)}\right)^{\mathsf{T}} \equiv \left(\rho^{(B)}\right)^{*}$. The question now is, is $\tilde{\rho}$ still a valid density operator? In general, a Hermitian operator ρ is a valid density operator if it has unit trace and is **positive semidefinite**; that is, if all its eigenvalues are nonnegative (equivalently, every diagonal matrix element of the density operator is nonnegative in every basis). In this sense, it represents a sensible quantum-mechanical probability distribution. This property often goes by the name of **rho-positivity**. The transpose operation does not affect the eigenvalue spectrum, so the transpose of $\rho^{(B)}$ is still a valid density operator, and it follows that $\tilde{\rho}$ is also a valid density operator. Clearly, this argument also holds for the more general separable state (4.148). Thus, we have established a *necessary* condition for the composite density operator to be separable.

In component form, we can again write the density operator as

$$\rho_{\alpha\mu\beta\nu} = \rho^{(A)}_{\alpha\beta}\,\rho^{(B)}_{\mu\nu}, \tag{4.155}$$

in which case the PPT is given by interchanging μ and ν:

$$\tilde{\rho}_{\alpha\mu\beta\nu} = \rho^{(A)}_{\alpha\beta}\,\rho^{(B)}_{\nu\mu}. \tag{4.156}$$

Thus, for an arbitrary density operator, the PPT test consists of computing the PPT of the operator, and then testing $\tilde{\rho}$ for positivity. The positivity of the transposed density operator,

$$\tilde{\rho}_{\alpha\mu\beta\nu} = \rho_{\alpha\nu\beta\mu} \geq 0, \tag{4.157}$$

(positive partial transpose criterion)

is a necessary (and sufficient in some cases, as we will discuss below) criterion for separability.

Fine, so let's take it for a test drive. Consider the entangled Schrödinger-cat state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|0_A 0_B\rangle + |1_A 1_B\rangle\Big). \tag{4.158}$$

The density operator is

$$\rho = \frac{1}{2}\Big(|00\rangle\langle 00| + |11\rangle\langle 11| + |00\rangle\langle 11| + |11\rangle\langle 00|\Big), \tag{4.159}$$

so that the PPT operation on subsystem B gives

$$\tilde{\rho} = \frac{1}{2}\Big(|00\rangle\langle 00| + |11\rangle\langle 11| + |01\rangle\langle 10| + |10\rangle\langle 01|\Big). \tag{4.160}$$

With the index ordering $(00, 01, 10, 11)$, this density operator corresponds to the density matrix

$$(\tilde{\rho}_{\alpha\beta}) = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.161}$$

which has eigenvalues $1/2$ and $-1/2$ with multiplicities 3 and 1, respectively. (We can also see that there is a problem since it is easy to see that the determinant is $-1/2^4$, and so at least one eigenvalue must be negative.) Hence this is not a proper density matrix, and the state is not separable according to the PPT criterion.

Horodecki[3] have shown that the PPT criterion is also *sufficient* for density operators in $\mathscr{H}_2 \otimes \mathscr{H}_2$ and $\mathscr{H}_2 \otimes \mathscr{H}_3$. That is sufficiency holds if the subsystems are both qubits (two-state quantum systems) or there is one qubit and one qutrit (three-state quantum system). For larger Hilbert spaces, the PPT criterion breaks down as a sufficient condition, but remains a useful tool.

More general and powerful criteria than the PPT can be devised,[25] though of course the general problem of distinguishing separable from entangled states is still computationally inefficient.

[25] Andrew C. Doherty, Pablo A. Parrilo, Federico M. Spedalieri, "Distinguishing separable and entangled states," *Physical Review Letters* **88**, 187904 (2002).

4.4.3.1 Wigner Representation

The PPT condition has an interesting interpretation in the Wigner representation. As we noted above, the transpose of the state is merely the complex conjugation operation. But as we discussed for the Wigner function, complex conjugation corresponds to time-reversal, and thus under the transformation $\psi(x) \longrightarrow \psi^*(x)$, the Wigner function undergoes the corresponding transformation $W(x,p) \longrightarrow W(x,-p)$.

Now consider the Wigner function generalized to N degrees of freedom:

$$W(x_\alpha, p_\alpha) := \frac{1}{(2\pi\hbar)^N} \int_{-\infty}^{\infty} dx'_1 \cdots \int_{-\infty}^{\infty} dx'_N \, e^{-ip_\beta x'_\beta/\hbar} \times \langle x_1 + x'_1/2, \ldots, x_N + x'_N/2 | \rho | x_1 + x'_1/2, \ldots, x_N + x'_N/2 \rangle. \tag{4.162}$$

Thus, in the Wigner representation the PPT criterion reads as follows. For the bipartite Wigner function $W(x_1, x_2, p_1, p_2)$, a necessary condition for separability is that

$$\tilde{W}(x_1, x_2, p_1, p_2) := W(x_1, x_2, p_1, -p_2) \tag{4.163}$$
(positive partial transpose)

be a valid Wigner function. The inversion of one momentum here *looks* like a physical transformation—it *is* for a system of one degree of freedom—but for a bipartite system it is *not*, because it maps some physical states to unphysical ones. This test is not necessarily easier than the original one for the density operator. However, Simon[26] showed that this form of the Peres–Horodecki criterion is also *sufficient* for Gaussian bipartite states.

We can see from the definition (4.162) of the Wigner function that the area theorem generalizes to

$$\left[\int d^N x \int d^N p \; W^2(x_\alpha, p_\alpha) \right]^{-1} \geq (2\pi\hbar)^N. \tag{4.164}$$
(area theorem)

We can write the Gaussian Wigner function as

$$W(z_\alpha) = \frac{1}{(2\pi)^N \sqrt{\det(S_{\alpha\beta})}} \exp\left[-\frac{1}{2} z_\alpha \left(S^{-1}\right)_{\alpha\beta} z_\beta \right], \tag{4.165}$$
(Gaussian state)

where we are using the generalized coordinate ordering

$$z_\alpha := (x_1, \ldots, x_N, p_1, \ldots, p_N), \tag{4.166}$$

and the covariance matrix is thus

$$S_{\alpha\beta} := \langle z_\alpha z_\beta \rangle = \frac{1}{2} \langle [\hat{z}_\alpha, \hat{z}_\beta]_+ \rangle \tag{4.167}$$

in the case where $\langle z_\alpha \rangle = 0$ (which is not a restrictive assumption). Here, $[a,b]_+ := ab + ba$ denotes the anticommutator bracket. Thus, the area theorem again implies the generalized "uncertainty relation" for Gaussian states

$$\det(S_{\alpha\beta}) \geq \left(\frac{\hbar^2}{4} \right)^N, \tag{4.168}$$
(area theorem)

where again equality holds only for pure states. We can use this as one necessary criterion for the validity of a Gaussian Wigner function. However, the PPT operation corresponds to a sign change in the covariance matrix of the form

$$\begin{bmatrix} + & + & + & + \\ + & + & + & + \\ + & + & + & + \\ + & + & + & + \end{bmatrix} \longrightarrow \begin{bmatrix} + & + & + & - \\ + & + & + & - \\ + & + & + & - \\ - & - & - & + \end{bmatrix}, \tag{4.169}$$

and thus we can see that $\det(S_{\alpha\beta})$ is actually *invariant* under PPT. Thus, the condition (4.168) is not strong enough to see why something goes wrong with the Wigner function under PPT.

[26] R. Simon, "Peres–Horodecki Separability Criterion for Continuous Variable Systems," *Physical Review Letters* **84**, 2726 (2000) (doi: 10.1103/PhysRevA.84.2726).

4.4.3.2 Generalized Uncertainty Relation

To express a better uncertainty relation, we first note that the commutators for the canonical coordinates can be written in the compact form

$$[z_\alpha, z_\beta] = i\hbar\Omega_{\alpha\beta}, \tag{4.170}$$
(commutation relation)

where $\Omega_{\alpha\beta}$ is the **canonical cosymplectic two-form**, defined by

$$(\Omega_{\alpha\beta}) := \begin{pmatrix} 0_n & \mathcal{I}_n \\ -\mathcal{I}_n & 0_n \end{pmatrix}, \tag{4.171}$$
(canonical cosymplectic form)

with $\mathcal{I}_n$ denoting the $n \times n$ identity matrix and 0_n the $n \times n$ null matrix. [Note that the cosymplectic form satisfies $-(\Omega_{\alpha\beta}) = (\Omega_{\alpha\beta})^\mathsf{T} = (\Omega_{\alpha\beta})^{-1}$.] The cosymplectic form effectively defines the structure of Hamiltonian mechanics. In particular, the classical Poisson bracket reads

$$\{f, g\}_\text{P} := (\partial_x f)(\partial_p g) - (\partial_p f)(\partial_x g) \quad \Longrightarrow \quad \{f, g\}_\text{P} = \frac{\partial f}{\partial z_\alpha}\Omega_{\alpha\beta}\frac{\partial g}{\partial z_\beta}, \tag{4.172}$$

and the classical Hamilton equations are (Section 8.2.2.2)

$$\partial_t p_\alpha = \partial_{x_\alpha} H, \;\; \partial_t x_\alpha = -\partial_{p_\alpha} H \quad \Longrightarrow \quad \partial_t z_\alpha = \Omega_{\alpha\beta}\frac{\partial H}{\partial z_\beta} = \{z_\alpha, H\}_\text{P}. \tag{4.173}$$

Essentially, the cosymplectic form mixes positions and momenta, but only if they belong to the same degree of freedom. It also treats them on equal footing, but an exchange of the positions with the momenta is accompanied by a minus sign.

The point of all this is that we can use the cosymplectic form to write down a generalized uncertainty relation:[27]

$$(S_{\alpha\beta}) + i\frac{\hbar}{2}(\Omega_{\alpha\beta}) \geq 0. \tag{4.174}$$
(generalized uncertainty relation)

Here, $(A_{\alpha\beta}) \geq 0$ means that the Hermitian operator $(A_{\alpha\beta})$ is positive semidefinite (i.e., it has no negative eigenvalues).

So what does *that* mean? Well, we can try this out for one degree of freedom, in which case we have

$$\begin{bmatrix} V_x & C_{xp} + i\hbar/2 \\ C_{xp} - i\hbar/2 & V_p \end{bmatrix} \geq 0. \tag{4.175}$$

Recalling positive semidefiniteness (positivity) is also required of the density operator, we note that this means that all the eigenvalues of the above matrix must be nonnegative. Diagonalizing the above matrix, we find that

$$\frac{1}{2}\left(V_x + V_p \pm \sqrt{(V_x - V_p)^2 + 4C_{xp}^2 + \hbar^2}\right) \geq 0. \tag{4.176}$$

Only the case of the negative sign is nontrivial, and this simplifies to

$$V_x V_p - C_{xp}^2 \geq \frac{\hbar^2}{4}, \tag{4.177}$$

which is precisely the uncertainty condition that we saw before in Eq. (4.72) for the Gaussian state [i.e., the one-dimensional version of Eq. (4.168)]. However, we're trying to establish this as a *general* criterion, not restricted to Gaussian states.

To establish this result in the case of a single degree of freedom, note that this result (4.177) reduced to the usual uncertainty principle for a diagonal covariance matrix ($C_{xp} = 0$). Now note that *any* covariance

[27] R. Simon, N. Mukunda, and Biswadeb Dutta, "Quantum-noise matrix for multimode systems: U(n) invariance, squeezing, and normal forms," *Physical Review A* **49** 1567 (1994) (doi: 10.1103/PhysRevA.49.1567).

matrix may be obtained from a diagonal one by a suitable initial choice of diagonal variance matrix and a coordinate rotation in phase space. A coordinate rotation preserves areas in phase space, so in a straightforward way, this establishes the result (4.177) for an arbitrary covariance matrix for a one-degree-of-freedom system. The coordinate rotation here is a special case of a *symplectic* transformation, which we will now examine more closely.

To handle the case of multiple degrees of freedom, we need to be more sophisticated about our coordinate transformations. In particular, let's make a time-independent coordinate transformation from z_α to new coordinates $\tilde{z}_\alpha = \tilde{z}_\alpha(z_\beta)$, which are continuous, differentiable functions of the old coordinates. The Hamiltonian is a scalar, so $H(z_\alpha) = \tilde{H}(\tilde{z}_\alpha)$, and so in the new coordinate system, we can write the equations of motion as

$$\partial_t \tilde{z}_\mu = \frac{\partial \tilde{z}_\mu}{\partial z_\alpha} \partial_t z_\alpha = \frac{\partial \tilde{z}_\mu}{\partial z_\alpha} \Omega_{\alpha\beta} \frac{\partial H}{\partial z_\beta} = \left[\frac{\partial \tilde{z}_\mu}{\partial z_\alpha} \Omega_{\alpha\beta} \frac{\partial \tilde{z}_\nu}{\partial z_\beta} \right] \frac{\partial H}{\partial \tilde{z}_\nu}, \tag{4.178}$$

where we used Hamilton's equations in the old coordinate system. Thus we can define

$$\tilde{\Omega}_{\mu\nu} := \frac{\partial \tilde{z}_\mu}{\partial z_\alpha} \Omega_{\alpha\beta} \frac{\partial \tilde{z}_\nu}{\partial z_\beta}, \tag{4.179}$$

which is the cosymplectic form for the new coordinate system. (Recall that $\Omega_{\alpha\beta}$ is the cosymplectic form in *canonical coordinates*.) Hamilton's equations in the new coordinates can thus be written

$$\partial_t \tilde{z}_\mu = \tilde{\Omega}_{\mu\nu} \frac{\partial H}{\partial \tilde{z}_\nu} = \{\tilde{z}_\mu, H\}_{\tilde{\mathrm{P}}}, \tag{4.180}$$

where the new Poisson bracket is

$$\{f, g\}_{\tilde{\mathrm{P}}} = \frac{\partial f}{\partial \tilde{z}_\mu} \tilde{\Omega}_{\mu\nu} \frac{\partial g}{\partial \tilde{z}_\nu}. \tag{4.181}$$

Now we can define a **canonical transformation**, which is a coordinate transformation that leaves the cosymplectic form unchanged. In this case, the new coordinates $\tilde{z}_\alpha$ are canonical. Thus, the transformation from z_α to $\tilde{z}_\alpha$ is canonical if and only if

$$\tilde{\Omega}_{\mu\nu} = \Omega_{\mu\nu} \quad \Longleftrightarrow \quad \Omega_{\mu\nu} = \frac{\partial \tilde{z}_\mu}{\partial z_\alpha} \Omega_{\alpha\beta} \frac{\partial \tilde{z}_\nu}{\partial z_\beta}. \tag{4.182}$$

(canonical transformation)

Again, this is because $\tilde{\Omega}_{\alpha\beta}$ must have the special form for the new coordinates to be canonical. In particular, note that evolution over a finite time interval T represents a coordinate change $\tilde{z}_\alpha(t) = z_\alpha(t+T)$ that satisfies these properties, and thus time evolution represents a canonical transformation.

Consider now a *linear*, canonical coordinate change, or possibly a nonlinear canonical coordinate change only in the neighborhood of a particular point z^α, so that we may linearize the transformation. Then we may represent the transformation by a matrix,

$$A_{\mu\nu} := \frac{\partial \tilde{z}^\mu}{\partial z^\nu}, \tag{4.183}$$

where $A_{\mu\nu}$ is independent of the coordinates and $\tilde{z}_\alpha = A_{\alpha\beta} z_\beta$. Being a canonical transformation, the linear mapping $A_{\mu\nu}$ satisfies

$$A_{\mu\alpha} \Omega_{\alpha\beta} A_{\nu\beta} = \Omega_{\mu\nu}, \tag{4.184}$$

or in other words $\mathbf{A}$ preserves the canonical cosymplectic form under a similarity transformation:

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \boldsymbol{\Omega}. \tag{4.185}$$

(symplectic matrix)

Any matrix that satisfies this condition is said to be a **symplectic matrix**.[28]

[28] The set of all $2n \times 2n$ real symplectic matrices forms a group, called the **symplectic group**, denoted $\mathrm{Sp}(2n, \mathbb{R})$. Hermann Weyl coined this term; he originally wanted to call this the "complex group," but wanted to avoid confusion with complex numbers. So he chose the term "symplectic" as a (rough) Greek translation of the word "complex." See Hermann Weyl, *The Classical Groups: Their Invariants and Representations* (Princeton, 1939), p. 165.

How does the covariance matrix transform? In general, we can expand the new coordinates as

$$\tilde{z}_\alpha = \left.\frac{\partial \tilde{z}_\alpha}{\partial z_\beta}\right|_{z_\alpha=0} z_\beta + O(z^2). \tag{4.186}$$

In new coordinates,

$$\tilde{S}_{\alpha\beta} = \langle \tilde{z}_\alpha \tilde{z}_\beta \rangle = \left.\frac{\partial \tilde{z}_\alpha}{\partial z_\mu}\right|_{z_\mu=0} \langle z_\mu z_\nu \rangle \left.\frac{\partial \tilde{z}_\beta}{\partial z_\nu}\right|_{z_\nu=0} + O(z^3). \tag{4.187}$$

Thus, we see that the covariance matrix only goes simply over to a covariance matrix under a linear transformation, in which case the coordinate transformation reduces to a symplectic matrix $A_{\alpha\beta}$, and we can write

$$\tilde{S}_{\alpha\beta} = A_{\alpha\mu} S_{\mu\nu} A_{\beta\nu}. \tag{4.188}$$

Thus, we can see that

$$A_{\alpha\mu}\Big(S_{\mu\nu} + i\frac{\hbar}{2}\Omega_{\mu\nu}\Big)A_{\beta\nu} = A_{\alpha\mu}S_{\mu\nu}A_{\beta\nu} + i\frac{\hbar}{2}\Omega_{\mu\nu} = \tilde{S}_{\mu\nu} + i\frac{\hbar}{2}\Omega_{\mu\nu}, \tag{4.189}$$

and thus the uncertainty relation (4.174) is invariant under linear canonical transformations,

$$(\tilde{S}_{\alpha\beta}) + i\frac{\hbar}{2}(\Omega_{\alpha\beta}) \geq 0, \tag{4.190}$$

as we expect. We have used the fact that the statement $\mathbf{M} \geq 0$ is equivalent to the statement $\tilde{\mathbf{M}} \geq 0$ if $\mathbf{M}$ and $\tilde{\mathbf{M}}$ are related by an invertible transformation (you should try proving this, it isn't difficult). This is certainly true here of the symplectic matrices, since we generally assume nonsingular coordinate transformations.

Okay, now let's return to the uncertainty relation for N degrees of freedom. We'll start again by considering only a *diagonal* covariance matrix. Then the matrix

$$(S_{\alpha\beta}) + i\frac{\hbar}{2}(\Omega_{\alpha\beta}) \tag{4.191}$$

only has elements along three diagonals: the cosymplectic form again couples positions to momenta, but *only* if they belong to the same degree of freedom. Thus, this matrix is easy to diagonalize: the matrix essentially decomposes into a set of N 2×2 blocks of the form

$$\begin{bmatrix} V_{x_\alpha} & i\hbar/2 \\ -i\hbar/2 & V_{p_\alpha} \end{bmatrix}, \tag{4.192}$$

just as we had in the case of one degree of freedom. Then, by the same argument, we can see that the uncertainty relation in this case reduces to the set of uncertainty relations

$$V_{x_\alpha} V_{p_\alpha} \geq \frac{\hbar^2}{4}, \tag{4.193}$$

which we know to be true.

For the *general* case, we then rely on the fact that *any* covariance matrix can be reduced to diagonal form by a linear, symplectic transformation (i.e., via a symplectic matrix). This seems reasonable, since as long as we are doing an effective linearization by only considering the covariance matrix, there should exist a linear, canonical transformation that takes any covariance matrix to any other (with the same "volume" or purity). However, this result can be formalized in Williamson's theorem[29], which guarantees that we can always use a symplectic transformation to obtain the diagonal form. We showed that the uncertainty relation is invariant under linear, canonical transformations, and thus the uncertainty relation (4.174) holds in general.

[29] Simon *et al.*, ibid.

4.4.3.3 Sufficiency for Gaussian States

Now we can get back to the original question, what goes wrong with the Wigner function under the partial transpose operation for entangled states? Well, we need to do a little more work with the uncertainty relation (4.174) to see what it means. Let $M_{\alpha\beta}$ be positive semidefinite. This means that

$$c_\alpha^* M_{\alpha\beta} c_\alpha \geq 0 \tag{4.194}$$

for any complex vector c_α, and note that we may take $M_{\alpha\beta}$ to be Hermitian. Now let

$$c_\alpha = a_\alpha + ib_\alpha, \tag{4.195}$$

where a_α and b_α are real vectors. Then the positivity condition reads

$$a_\alpha M_{\alpha\beta} a_\beta + b_\alpha M_{\alpha\beta} b_\beta + i a_\alpha M_{\alpha\beta} b_\beta - i b_\alpha M_{\alpha\beta} a_\beta \geq 0. \tag{4.196}$$

Since $M_{\alpha\beta}$ is Hermitian,

$$b_\alpha M_{\alpha\beta} a_\beta = a_\alpha M_{\alpha\beta}^* b_\beta. \tag{4.197}$$

Then with

$$M_{\alpha\beta} - M_{\alpha\beta}^* = i\mathrm{Im}[M_{\alpha\beta}], \tag{4.198}$$

the positivity condition is

$$a_\alpha M_{\alpha\beta} a_\beta + b_\alpha M_{\alpha\beta} b_\beta - 2a_\alpha \mathrm{Im}[M_{\alpha\beta}] b_\beta \geq 0. \tag{4.199}$$

Letting $M_{\alpha\beta} \longrightarrow S_{\alpha\beta} + i(\hbar/2)\Omega_{\alpha\beta}$ and using $a_\alpha \Omega_{\alpha\beta} a_\beta = 0$ (because $\Omega_{\alpha\beta} = -\Omega_{\beta\alpha}$), we see that the uncertainty relation becomes

$$a_\alpha S_{\alpha\beta} a_\beta + b_\alpha S_{\alpha\beta} b_\beta \geq \hbar a_\alpha \Omega_{\alpha\beta} b_\beta. \tag{4.200}$$

The left-hand side is invariant under the exchange $a_\alpha \longleftrightarrow b_\alpha$, but the right-hand side changes sign, giving a second condition. Only one of these two conditions is nontrivial, and so it is appropriate to rewrite the uncertainty condition as

$$a_\alpha S_{\alpha\beta} a_\beta + b_\alpha S_{\alpha\beta} b_\beta \geq \hbar \left| a_\alpha \Omega_{\alpha\beta} b_\beta \right| \tag{4.201}$$

to emphasize the nontrivial condition.

In the case of two degrees of freedom,

$$\Omega_{\alpha\beta} b_\beta = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} b_3 \\ b_4 \\ -b_1 \\ -b_2 \end{bmatrix}, \tag{4.202}$$

and so

$$a_\alpha \Omega_{\alpha\beta} b_\beta = (a_1 b_3 - b_1 a_3) + (a_2 b_4 - b_2 a_4). \tag{4.203}$$

Thus, the uncertainty condition becomes

$$a_\alpha S_{\alpha\beta} a_\beta + b_\alpha S_{\alpha\beta} b_\beta \geq \hbar \left| (a_1 b_3 - b_1 a_3) + (a_2 b_4 - b_2 a_4) \right|. \tag{4.204}$$

Recalling that the partial transpose corresponds to $p_2 \longrightarrow -p_2$, we see that the partial transpose is induced by the operator $\Lambda = \mathrm{diag}(1, 1, 1, -1)$, so that we can write the transposition as

$$S_{\alpha\beta} \longrightarrow S'_{\alpha\beta} = \Lambda_{\alpha\mu} S_{\mu\nu} \Lambda_{\beta\nu}, \qquad z_\alpha \longrightarrow z'_\alpha = \Lambda_{\alpha\beta} z_\beta. \tag{4.205}$$

Note that $\Lambda_{\alpha\beta}$ does not correspond to a unitary transformation, nor does it represent a canonical transformation, as we can see from the transformation of the cosymplectic form:

$$\Omega_{\alpha\beta} \longrightarrow \Omega'_{\alpha\beta} = \Lambda_{\alpha\mu} \Omega_{\mu\nu} \Lambda_{\beta\nu} = \begin{bmatrix} \mathbf{0} & \mathbf{J} \\ -\mathbf{J} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{J} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{4.206}$$

This clearly does not have the same form as the original cosymplectic form. Again, the Peres–Horodecki criterion says that under partial transposition, the resulting Wigner function for a separable state must still be a valid state. Thus, we have the new uncertainty relation that must be satisfied for separable states:

$$(S'_{\alpha\beta}) + i\frac{\hbar}{2}(\Omega_{\alpha\beta}) \geq 0. \tag{4.207}$$

Since $(\Lambda_{\alpha\beta}) = (\Lambda_{\alpha\beta})^{-1}$, this condition is equivalent to

$$(S_{\alpha\beta}) + i\frac{\hbar}{2}(\Omega'_{\alpha\beta}) \geq 0. \tag{4.208}$$

Repeating the above derivation, we see that

$$\Omega'_{\alpha\beta}b_\beta = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} b_3 \\ -b_4 \\ -b_1 \\ b_2 \end{bmatrix}, \tag{4.209}$$

and thus

$$a_\alpha\Omega'_{\alpha\beta}b_\beta = (a_1b_3 - b_1a_3) - (a_2b_4 - b_2a_4). \tag{4.210}$$

This is the same as the uncertainty relation as before save for a minus sign. Noting that $w \geq |u+v|$ and $w \geq |u-v|$ implies that $w \geq |u| + |v|$ for real numbers u, v, and w, we see from Eq. (4.204) that the uncertainty relation for *separable* states reads

$$a_\alpha S_{\alpha\beta}a_\beta + b_\alpha S_{\alpha\beta}b_\beta \geq \hbar\,|a_1b_3 - b_1a_3| + \hbar\,|a_2b_4 - b_2a_4|\,. \tag{4.211}$$

We can see that this is a stronger condition than the uncertainty relation (4.204).

Thus we see that separable states obey a stricter uncertainty law than generic states. Now let's consider an example to try to get a feeling for this. Consider the case

$$(a_\alpha) = \frac{1}{x_0}\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \qquad (b_\alpha) = \frac{1}{p_0}\begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \tag{4.212}$$

where x_0 and p_0 are length and momentum scales, respectively, that we introduce to make the units come out right. Then

$$(a_1b_3 - b_1a_3) = \frac{1}{x_0p_0}, \qquad (a_2b_4 - b_2a_4) = -\frac{1}{x_0p_0}, \tag{4.213}$$

and the terms on the left-hand side of the uncertainty relation are

$$\begin{aligned} a_\alpha S_{\alpha\beta}a_\beta &= \frac{1}{x_0^2}\left(V_{x_1} + V_{x_2} + 2C_{x_1x_2}\right) = \frac{1}{x_0^2}\left\langle (x_1+x_2)^2\right\rangle \\ b_\alpha S_{\alpha\beta}b_\beta &= \frac{1}{p_0^2}\left(V_{p_1} + V_{p_2} - 2C_{p_1p_2}\right) = \frac{1}{p_0^2}\left\langle (p_1-p_2)^2\right\rangle. \end{aligned} \tag{4.214}$$

In this case, we can see that the uncertainty relation

$$\frac{1}{x_0^2}\left\langle (x_1+x_2)^2\right\rangle + \frac{1}{p_0^2}\left\langle (p_1-p_2)^2\right\rangle \geq 0, \tag{4.215}$$

which always trivially holds. The separability condition, on the other hand, gives

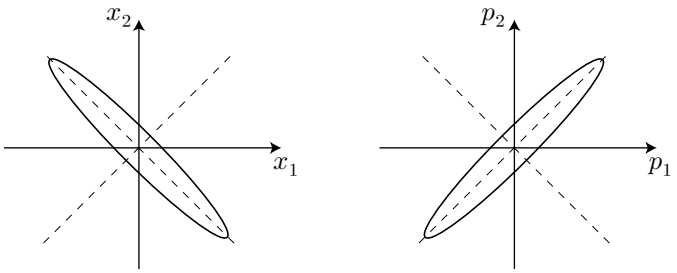
$$\frac{1}{x_0^2}\left\langle (x_1+x_2)^2\right\rangle + \frac{1}{p_0^2}\left\langle (p_1-p_2)^2\right\rangle \geq \frac{2\hbar}{x_0p_0}. \tag{4.216}$$

Introducing scaled variables $x'_\alpha := x_\alpha/x_0$ and $p'_\alpha := p_\alpha/p_0$, and choosing $x_0 p_0 = \hbar$, these conditions become

$$\begin{aligned} \left\langle (x'_1 + x'_2)^2 \right\rangle + \left\langle (p'_1 - p'_2)^2 \right\rangle &\geq 0 \quad \text{(uncertainty)} \\ \left\langle (x'_1 + x'_2)^2 \right\rangle + \left\langle (p'_1 - p'_2)^2 \right\rangle &\geq 2 \quad \text{(separability)}. \end{aligned} \tag{4.217}$$

It is certainly possible to find a Gaussian state to violate this second condition. Such a state is a *two-mode squeezed state*, and represents entanglement between the particles.

In effect the entanglement arises via *correlations* between x_1 and x_2, and between p_1 and p_2. That is, even if p_1 and p_2 have large variances, it is possible for the *difference* to be well-defined. To compensate, the *sum* must be ill-defined. Thus, the two-mode squeezed state with $x_1 + x_2$ and $p_1 - p_2$ squeezed would look something like this:



Note that the choice of the particular combination $(x_1 + x_2)$ and $(p_1 - p_2)$ is not accidental, as these operators form a pair of commuting observables:

$$[x_1 + x_2, p_1 - p_2] = [x_1, p_1] - [x_2, p_2] = 0. \tag{4.218}$$

For example, if we had instead chosen the combination $(p_1 + p_2)$, this would not have worked out, since the variables no longer commute:

$$[x_1 + x_2, p_1 + p_2] = [x_1, p_1] + [x_2, p_2] = 2i\hbar. \tag{4.219}$$

In this case,

$$(b_\alpha) = \frac{1}{p_0} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \tag{4.220}$$

and so $(a_2 b_4 - b_2 a_4) = 1/x_0 p_0$ and $b_\alpha S_{\alpha\beta} b_\beta = \left\langle (p_1 + p_2)^2 \right\rangle / p_0^2$. However, the uncertainty relation and separability conditions are equivalent in this case, reading (in the same scaled coordinates)

$$\left\langle (x'_1 + x'_2)^2 \right\rangle + \left\langle (p'_1 + p'_2)^2 \right\rangle \geq 2 \quad \text{(separability } \textit{and} \text{ uncertainty)}. \tag{4.221}$$

For the noncommuting observables, as usual, correlation in one set leads to decorrelation of the others, in such a way that squeezing does not work. Now we can see what goes wrong in the partial transpose: it is possible to construct a state squeezed in $(x_1 + x_2)$ and $(p_1 - p_2)$, and under the transpose, we obtain a state that violates the uncertainty relation for $(x_1 + x_2)$ and $(p_1 + p_2)$. Thus, we see explicitly that separability implies uncertainty relations *even among commuting observables*.

For Gaussian states, the variances represent the end of the story: they fully characterize the Gaussian. Indeed, it has been shown[30] that the above separability criteria are sufficient for separability for two-mode continuous Gaussian systems. (We knew all along that the separability criterion was *necessary*.)

[30] R. Simon, ibid. For a related analysis, see Lu-Ming Duan, G. Giedke, J. I. Cirac, and P. Zoller, "Inseparability Criterion for Continuous Variable Systems," *Physical Review Letters* **84**, 2722 (2000).

4.4.4 Indistinguishability

Bosonic and fermionic quantum particles that are identical are furthermore *indistinguishable*, even in principle. This induces a structure that *looks* a lot like entanglement, but isn't really the same.[31] Consider the two-particle state

$$|\psi\rangle = |(\psi_1)_A\,(\psi_2)_B\rangle. \tag{4.222}$$

That is, particle A is in state $|\psi_1\rangle$, and particle B is in state $|\psi_2\rangle$. This state is appropriate for *distinguishable* particles. But if the two particles are *indistinguishable*, the state must be invariant under exchange of the particle labels,

$$|\psi\rangle = |(\psi_2)_A\,(\psi_1)_B\rangle, \tag{4.223}$$

which for this state would imply that $|\psi_1\rangle = |\psi_2\rangle$. It is hardly satisfactory for every indistinguishable particle to be in the same state, so we can introduce an explicit symmetrization (antisymmetrization) as follows:

$$|\psi\rangle_\pm = \frac{1}{\sqrt{2}}\Big(|(\psi_1)_A\,(\psi_2)_B\rangle \pm |(\psi_2)_A\,(\psi_1)_B\rangle\Big), \tag{4.224}$$

In the case of the minus sign, the particle exchange is accompanied by a factor (-1), amounting to just an overall phase, which is certainly permissible.

What happens when we try to superpose these states? We can see that

$$\frac{1}{\sqrt{2}}\Big(|\psi\rangle_+ + |\psi\rangle_-\Big) = |(\psi_1)_A\,(\psi_2)_B\rangle, \tag{4.225}$$

and so we end up in a state that's no good. Evidently, we can't superpose states corresponding to different symmetrizations. Thus, we must postulate that once a pair of particles obey a certain symmetrization rule, they must *always* do so. (This is equivalent to the statement that the operator corresponding to the exchange operation commutes with the Hamiltonian.) Of course, the particles corresponding to the + sign in (4.224) are **bosons**, and those corresponding to the − sign are **fermions**. Again, Eq. (4.224) is suggestive of an entangled state, but only in a trivial sense, since it is nonsensical to speak of separate identities for the two particles. It is completely, fundamentally impossible to tell them apart. Another way to see this that even with an *unsymmetrized* state vector, we can always impose the symmetrization via the representation:

$$\frac{1}{\sqrt{2}}\Big(\langle(x_1)_A\,(x_2)_B| \pm \langle(x_2)_A\,(x_1)_B|\Big)|\psi\rangle = \frac{1}{\sqrt{2}}\left[\psi(x_1,x_2) \pm \psi(x_2,x_1)\right]. \tag{4.226}$$

Here, x_1 and x_2 represent two different position coordinates, it is the ordering of the arguments that determines which particle is associated with which position.

4.4.4.1 Exchange "Force"

One consequence of indistinguishability is an interference effect that looks something like an effective force between indistinguishable particles. This effect, the *exchange force*, is particularly important in understanding atomic and molecular structure, condensed matter systems, quantum degenerate gases, and astrophysics (where ''degeneracy pressure'' prevents white dwarfs and neutron stars from collapsing).

Consider two particles A and B in respective states $|\psi_1\rangle$ and $|\psi_2\rangle$, which we assume to be orthonormal. For distinguishable particles, the composite state is

$$|\psi\rangle = |(\psi_1)_A\,(\psi_2)_B\rangle. \tag{4.227}$$

The joint spatial probability density is

$$P(A \text{ at } x_A, B \text{ at } x_B) = |\langle x_A\, x_B|\psi\rangle|^2 = |\psi_1(x_A)|^2|\psi_2(x_B)|^2 \quad \text{(distinguishable)}, \tag{4.228}$$

[31] The presentation here follows parts of David J. Griffiths, *Introduction to Quantum Mechanics* (Prentice–Hall, 1995), Chapter 5, p. 177; Asher Peres, *Quantum Theory: Concepts and Methods* (Kluwer, 1995), Section 5-4, p. 126; and lecture notes by M. Baranger and J. Negele. See also Daniel F. Styer, ''Common misconceptions regarding quantum mechanics,'' *American Journal of Physics* **64**, 31 (1996).

where $\psi_1(x_A) = \langle x_A|(\psi_1)_A\rangle$ and $\psi_2(x_B) = \langle x_B|(\psi_2)_B\rangle$. If we choose not to distinguish between the particles, we can compute the probability of finding one at x and the other at x',

$$\begin{aligned} P(\text{1 particle at } x\text{, 1 particle at } x') &= P(A \text{ at } x,\, B \text{ at } x') + P(A \text{ at } x',\, B \text{ at } x) \\ &= |\psi_1(x)|^2|\psi_2(x')|^2 + |\psi_1(x')|^2|\psi_2(x)|^2 \quad \text{(distinguishable)}. \end{aligned} \tag{4.229}$$

This expression facilitates comparison with the indistinguishable case.

Now consider indistinguishable particles, either bosons or fermions

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|(\psi_1)_A\,(\psi_2)_B\rangle \pm |(\psi_2)_A\,(\psi_1)_B\rangle\Big), \tag{4.230}$$

In this case the joint spatial probability density becomes

$$\begin{aligned} P(A \text{ at } x_A, B \text{ at } x_B) &= |\langle x_A\, x_B|\psi\rangle|^2 \\ &= \frac{1}{2}\Big(|\psi_1(x_A)|^2|\psi_2(x_B)|^2 + |\psi_2(x_A)|^2|\psi_1(x_B)|^2 \\ &\qquad \pm 2\mathrm{Re}[\psi_1^*(x_A)\psi_2^*(x_B)\psi_2(x_A)\psi_1(x_B)]\Big) \quad \text{(bosons/fermions)}. \end{aligned} \tag{4.231}$$

Again, we must drop the particle labels, so

$$\begin{aligned} P(\text{1 particle at } x, \text{1 particle at } x') = \Big(&|\psi_1(x)|^2|\psi_2(x')|^2 + |\psi_2(x)|^2|\psi_1(x')|^2 \\ &\pm 2\mathrm{Re}[\psi_1^*(x)\psi_2^*(x')\psi_2(x)\psi_1(x')]\Big) \quad \text{(bosons/fermions)}. \end{aligned} \tag{4.232}$$

The final interference term is the **exchange term**. Note that it is nonvanishing only if the two wave functions $\psi_1(x)$ and $\psi_2(x)$ overlap.

To see the effect of the exchange term, consider the probability density when $x = x'$. For distinguishable particles, the probability density is simply

$$P(\text{both at } x) = 2|\psi_1(x)|^2|\psi_2(x)|^2, \tag{4.233}$$
(distinguishable particles)

while in the indistinguishable case,

$$\begin{aligned} P(\text{both at } x) &= 2|\psi_1(x)|^2|\psi_2(x)|^2 \pm 2\mathrm{Re}[|\psi_1(x)|^2|\psi_2(x)|^2] \\ &= \begin{cases} 4|\psi_1(x)|^2|\psi_2(x)|^2 & \text{(bosons)} \\ 0 & \text{(fermions)} \end{cases} \end{aligned}$$
(indistinguishable particles) (4.234)

Thus, we see that the probability density for being at the same position *doubles* with respect to the distinguishable case for bosons, and vanishes for fermions. Thus, it is common to speak of the attractive ''**exchange force**,'' or **exchange interaction** between bosons, and the repulsive ''force'' between fermions. This effect is not a force, however; it is simply an interference effect due to the symmetry properties under exchange. In particular, if two noninteracting particles ''collide'' and then separate, there is no scattering or net phase shift after the crossing of the particles due to the exchange interaction. However, these *would* occur if the exchange interaction really could be modeled as a force (instead, the exchange interaction affects the details of the particles spatial distributions only while they are overlapping).[32]

[32] W. J. Mullin and G. Blaylock, ''Quantum statistics: Is there an effective fermion repulsion or boson attraction?'' *American Journal of Physics* **71**, 1223 (2003) (doi: 10.1119/1.1590658).

4.4.5 Open Systems: Church of the Larger Hilbert Space

One important use of the density operator is in describing *open* quantum systems—systems interacting with auxiliary systems (*environments* or *reservoirs*) that we don't have access to. We will treat open quantum systems in great detail, but for now let's examine a simple model for why the density operator is useful.

Consider the entangled state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|0_A\rangle|0_B\rangle + |1_A\rangle|1_B\rangle\Big) \tag{4.235}$$

between particles (qubits) A and B. Suppose that we have access to particle A, but particle B is locked up in a box, so that we don't know anything about it. The density operator for the composite system is

$$\rho = |\psi\rangle\langle\psi| = \frac{1}{2}\Big(|0_A\rangle|0_B\rangle\langle 0_A|\langle 0_B| + |1_A\rangle|1_B\rangle\langle 1_A|\langle 1_B| + |1_A\rangle|1_B\rangle\langle 0_A|\langle 0_B| + |0_A\rangle|0_B\rangle\langle 1_A|\langle 1_B|\Big) \tag{4.236}$$

We can define the **reduced density operator** that describes *only* particle A by performing a partial trace over the state of particle B:

$$\rho_A = \mathrm{Tr}_B[\rho] := \sum_\alpha \langle\alpha_B|\rho|\alpha_B\rangle = \frac{1}{2}\Big(|0_A\rangle\langle 0_A| + |1_A\rangle\langle 1_A|\Big).$$

(reduced density operator) (4.237)

Thus, we can see that the reduced state of particle A corresponds to a *completely incoherent* superposition of the two states, even though the composite system carried a completely coherent superposition.

This is a simple model for the process of **decoherence**. A quantum system can start in a local state of coherent superposition. But if it interacts with the environment, the coupling causes entanglement between the system and environment, since the interaction is nonlocal. Because we don't have access to the state of the environment, we must trace over it, which reduces the purity of the reduced density operator. Note that we can't keep track of the environment *even in principle*, since it generally has many degrees of freedom. As the interaction continues, the entanglement progresses, driving the reduced density operator towards a completely incoherent superposition. This is, at a simple level, why classical (macroscopic) things behave classically: coupling to the environment destroys quantum coherence.

Conversely, whenever we have a system described by a mixed density operator,

$$\rho = \sum_\alpha P_\alpha |\psi_\alpha\rangle\langle\psi_\alpha|, \tag{4.238}$$

we can always think of it as part of a larger system. We can see this as follows. We will introduce a fictitious environment with orthonormal basis states $|\alpha_\mathrm{E}\rangle$. Then we can write the state vector for the composite system as

$$|\psi_\mathrm{total}\rangle = \sum_\alpha \sqrt{P_\alpha}|\psi_\alpha\rangle|\alpha_\mathrm{E}\rangle. \tag{4.239}$$

(purification of mixed state)

When we compute the total density operator for the composite pure state and trace over the environment, we recover the original density operator (4.238) as the reduced density operator of the larger state. This procedure of switching to a larger pure state is referred to as **purification** or "the doctrine of the Church of the larger Hilbert space."[33] The space of these extra environmental degrees of freedom is often referred to as the **ancilla**.[34] Often, this is a useful picture for thinking about mixed quantum states, especially in quantum-information problems.

[33] Terminology introduced by John Smolin; see Daniel Gottesman and Hoi-Kwong Lo, "From Quantum Cheating to Quantum Security," *Physics Today* **53**, no. 11, 22 (2000) (doi: 10.1063/1.1333282).

[34] Although it just doesn't sound right to my ear, "ancilla" is a singular noun that comes from latin, where it means "hand-maiden," but in contemporary English can mean something that helps to achieve something difficult. (The plural form can be "ancillae" or "ancillas.")

4.5 Master Equation

A main and obvious advantage of the density-operator formalism is that it provides a method for handling *nonunitary* evolution of the quantum state. This generally occurs in the treatment of *open* quantum systems: quantum systems coupled to external systems that we do not directly track.[35] We saw the simplest example of this in the previous section: maximal entanglement of two qubits enforces minimum purity in the reduced state of one qubit. Now we will take this idea and study more generally how *weak* entanglement with an external system leads to a general, nonunitary evolution equation for the density operator.

We will thus study the evolution of a quantum system, described by Hamiltonian $H_{\rm S}$, interacting with a "reservoir" (or "heat bath" or "environment"), described by Hamiltonian $H_{\rm R}$. We will assume the system–reservoir interaction, described by $H_{\rm SR}$, to be weak, causing slow evolution on the uncoupled time scales of the system and reservoir separately. The evolution of the *total* system is unitary, given by

$$\partial_t \rho_{\rm SR} = -\frac{i}{\hbar}[H, \rho_{\rm SR}], \tag{4.240}$$

where $\rho_{\rm SR}$ is the combined state of the system and reservoir, and the total Hamiltonian is

$$H = H_{\rm S} + H_{\rm R} + H_{\rm SR}. \tag{4.241}$$

Our goal is to derive an equation of motion for the state of the system alone, given by a partial trace over the reservoir degrees of freedom:

$$\rho := {\rm Tr}_{\rm R}[\rho_{\rm SR}]. \tag{4.242}$$

Note that so long as we are interested in operators that act solely on the system's Hilbert space, this reduced density operator is sufficient to compute any appropriate expectation values.

We will derive the master equation with a number of approximations and idealizations, mostly related to the reservoir having many degrees of freedom. The approximations here typically work extremely well in quantum optics, though not necessarily in other areas such as condensed-matter physics where, for example, the weak-coupling idealization may break down. Examples of reservoirs include the quantum electromagnetic field (in a vacuum or thermal state), or the internal degrees of freedom of a composite object.

4.5.1 Interaction Representation

The first step is to switch to the interaction representation, in effect hiding the fast dynamics of the uncoupled system and reservoir, and focusing on the slow dynamics induced by $H_{\rm SR}$. We do this as in Section 4.2.3 via the transformations

$$\begin{aligned} \tilde{\rho}_{\rm SR}(t) &= e^{i(H_{\rm S}+H_{\rm R})t/\hbar}\rho_{\rm SR}(t)e^{-i(H_{\rm S}+H_{\rm R})t/\hbar} \\ \tilde{H}_{\rm SR}(t) &= e^{i(H_{\rm S}+H_{\rm R})t/\hbar}H_{\rm SR}e^{-i(H_{\rm S}+H_{\rm R})t/\hbar}, \end{aligned} \tag{4.243}$$

so that the formerly time-independent interaction becomes explicitly time-dependent. The equation of motion then becomes

$$\partial_t \tilde{\rho}_{\rm SR}(t) = -\frac{i}{\hbar}[\tilde{H}_{\rm SR}(t), \tilde{\rho}_{\rm SR}(t)]. \tag{4.244}$$

Integrating this from t to $t + \Delta t$,

$$\tilde{\rho}_{\rm SR}(t+\Delta t) = \tilde{\rho}_{\rm SR}(t) - \frac{i}{\hbar}\int_t^{t+\Delta t} dt' \, [\tilde{H}_{\rm SR}(t'), \tilde{\rho}_{\rm SR}(t')]. \tag{4.245}$$

Iterating this equation by using it as an expression for $\tilde{\rho}_{\rm SR}(t')$,

$$\tilde{\rho}_{\rm SR}(t+\Delta t) - \tilde{\rho}_{\rm SR}(t) = -\frac{i}{\hbar}\int_t^{t+\Delta t} dt' \, [\tilde{H}_{\rm SR}(t'), \tilde{\rho}_{\rm SR}(t)] - \frac{1}{\hbar^2}\int_t^{t+\Delta t} dt' \int_t^{t'} dt'' \, [\tilde{H}_{\rm SR}(t'), [\tilde{H}_{\rm SR}(t''), \tilde{\rho}_{\rm SR}(t'')]]. \tag{4.246}$$

[35] For further reading, see William H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, 1973), Chapter 6; Claude Cohen–Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Atom–Photon Interactions: Basic Processes and Applications* (Wiley, 1992), Chapter IV; and Howard Carmichael, *An Open Systems Approach to Quantum Optics* (Springer, 1993), Chapter 1.

Now in taking the trace over the reservoir, we will assume that the first term on the right-hand side vanishes. More specifically, we assume

$$\mathrm{Tr}_{\mathrm{R}}[\tilde{H}_{\mathrm{SR}}(t')\tilde{\rho}_{\mathrm{SR}}(t)] = 0. \tag{4.247}$$

This follows by assuming that the total system–reservoir state always approximately factorizes

$$\tilde{\rho}_{\mathrm{SR}}(t) \approx \tilde{\rho}(t) \otimes \tilde{\rho}_{\mathrm{R}}, \tag{4.248}$$

where $\tilde{\rho}_{\mathrm{R}}$ is the *stationary* state of the reservoir. This amounts to assuming that the reservoir is large and complex, and weak coupling of the system to the reservoir, so that the perturbation to the reservoir by the system is small. In this case, the time interval $\Delta t \gg \tau_{\mathrm{c}}$, where τ_{c} is the correlation time of the reservoir—the time for reservoir and system–reservoir correlations to decay away. This also amounts to a **coarse-graining approximation**, which means that we are smoothing out any fast dynamics on time scales of the order of τ_{c} or shorter. Thus, any correlations that have arisen in past time intervals have decayed away. Of course, *new* correlations arise due to the coupling in the present time interval, which will give rise to nonunitary terms in the evolution equation for the reduced state. Then the assumption (4.247) amounts to

$$\mathrm{Tr}_{\mathrm{R}}[\tilde{H}_{\mathrm{SR}}(t')\tilde{\rho}_{\mathrm{R}}] = 0. \tag{4.249}$$

This assumption means essentially that there is no dc component to the system–reservoir coupling—that is, the system–reservoir coupling consists of *fluctuations* about a zero mean. This can always be arranged by absorbing any nonzero mean into the system Hamiltonian.

4.5.2 Born–Markov Approximation

Since the first term vanishes under the partial trace, with the trace Eq. (4.246) becomes

$$\Delta\tilde{\rho}(t) \approx -\frac{1}{\hbar^2}\int_t^{t+\Delta t} dt' \int_t^{t'} dt'' \,\mathrm{Tr}_{\mathrm{R}}[\tilde{H}_{\mathrm{SR}}(t'), [\tilde{H}_{\mathrm{SR}}(t''), \tilde{\rho}_{\mathrm{SR}}(t'')]], \tag{4.250}$$

with $\Delta\tilde{\rho}(t) := \tilde{\rho}(t+\Delta t) - \tilde{\rho}(t)$. Now we will make the **Born–Markov approximation** by setting

$$\tilde{\rho}_{\mathrm{SR}}(t'') \approx \tilde{\rho}(t) \otimes \tilde{\rho}_{\mathrm{R}}. \tag{4.251}$$

In fact there is a pair of approximations at work here. The **Born approximation** amounts to assuming the factorization in (4.248), which we have justified in terms of a large, complex reservoir with a short coherence time. The **Markov approximation** amounts to setting $\rho(t'')$ to $\rho(t)$ in (4.251), which will result in an evolution equation that only depends on $\rho(t)$, and not the past history of the density operator. We can justify this approximation by noting that Δt is small and H_{SR} induces a weak perturbation, so that $\rho(t'') = \rho(t) + O(\Delta t)$. Then this amounts to a lowest-order expansion in Δt of the right-hand side of Eq. (4.250), which is appropriate in view of the limit $\Delta t \longrightarrow 0$ to obtain a differential equation (though in a coarse-grained sense, since strictly speaking we always require $\Delta t \gg \tau_{\mathrm{c}}$).

Next we change integration variables by setting

$$\tau := t' - t'', \tag{4.252}$$

so that the integration becomes

$$\begin{aligned}\int_t^{t+\Delta t} dt' \int_t^{t'} dt'' &= \int_0^{\Delta t} d\tau \int_{t+\tau}^{t+\Delta t} dt' \\ &\approx \int_0^{\infty} d\tau \int_t^{t+\Delta t} dt'.\end{aligned} \tag{4.253}$$

In writing down the final, approximate form for the integrals, we have used the fact that the integrand involves an expectation value of the interaction Hamiltonian taken at times that differ by τ, as we will explore further shortly. That is, the integrand involves reservoir correlation functions, which decay away on the time scale τ_{c}.

4.5.3 Interaction

Now we make a reasonably general assumption regarding the interaction Hamiltonian; namely, that it can be written as a sum of products over system and reservoir operators:

$$H_{\rm SR} = \hbar S_\alpha R_\alpha. \tag{4.254}$$

(Recall that repeated indices imply summation.) The interpretation here is that if S_α is a Hermitian operator, then it represents an observable that is being effectively (or actually) monitored via coupling to the environment. For example, a position measurement is represented by an interaction of the form $H_{\rm SR} = xR$ (Chapter 19). Alternately, the operators need not be Hermitian. For example, an interaction of the form $H_{\rm SR} = SR^\dagger + S^\dagger R$ represents the exchange of quanta (e.g., of energy) between the system and reservoir, and would thus represent dissipation or loss of energy to the reservoir. Such interactions occur in spontaneous emission (Chapter 11) and cavity decay (Chapter 12).

With the interaction of the form (4.254) and the change of integration in Eqs. (4.253), the change (4.250) in the quantum state becomes

$$\begin{aligned}\Delta\tilde\rho(t) \approx -\int_0^\infty d\tau \int_t^{t+\Delta t} dt' \Big\{ &\Big[\tilde S_\alpha(t')\tilde S_\beta(t'-\tau)\tilde\rho(t) - \tilde S_\beta(t'-\tau)\tilde\rho(t)\tilde S_\alpha(t')\Big] G_{\alpha\beta}(\tau) \\ &+ \Big[\tilde\rho(t)\tilde S_\beta(t'-\tau)\tilde S_\alpha(t') - \tilde S_\alpha(t')\tilde\rho(t)\tilde S_\beta(t'-\tau)\Big] G_{\beta\alpha}(-\tau)\Big\},\end{aligned} \tag{4.255}$$

where we have defined the reservoir correlation functions

$$G_{\alpha\beta}(\tau) := {\rm Tr}_{\rm R}\Big[\tilde R_\alpha(t')\tilde R_\beta(t'-\tau)\,\tilde\rho_{\rm R}\Big] = \Big\langle \tilde R_\alpha(t')\tilde R_\beta(t'-\tau)\Big\rangle_{\rm R} = \Big\langle \tilde R_\alpha(\tau)\tilde R_\beta(0)\Big\rangle_{\rm R}, \tag{4.256}$$

which depend only on the time difference because the reservoir is in a stationary state. Now we make the further assumption

$$\tilde S_\alpha(t) = e^{iH_{\rm S}t/\hbar} S_\alpha e^{-iH_{\rm S}t/\hbar} = S_\alpha e^{i\omega_\alpha t} \tag{4.257}$$

about the interaction-picture system operators. This is not necessarily a restrictive assumption, since multiple frequencies for a given system operator may be separated in the sum in (4.254). Then Eq. (4.255) becomes

$$\begin{aligned}\Delta\tilde\rho(t) \approx -\int_0^\infty d\tau \int_t^{t+\Delta t} dt' \Big\{ &\Big[S_\alpha S_\beta\tilde\rho(t) - S_\beta\tilde\rho(t)S_\alpha\Big] G_{\alpha\beta}(\tau) \\ &+ \Big[\tilde\rho(t)S_\beta S_\alpha - S_\alpha\tilde\rho(t)S_\beta\Big] G_{\beta\alpha}(-\tau)\Big\} e^{i\omega_\alpha t'} e^{i\omega_\beta(t'-\tau)}.\end{aligned} \tag{4.258}$$

Now defining

$$\begin{aligned} I(\omega_\alpha+\omega_\beta) &:= \int_t^{t+\Delta t} dt'\, e^{i(\omega_\alpha+\omega_\beta)t'} \\ w^+_{\alpha\beta} &:= \int_0^\infty d\tau\, e^{-i\omega_\beta\tau} G_{\alpha\beta}(\tau) \\ w^-_{\beta\alpha} &:= \int_0^\infty d\tau\, e^{-i\omega_\beta\tau} G_{\beta\alpha}(-\tau), \end{aligned} \tag{4.259}$$

we can write

$$\Delta\tilde\rho(t) \approx -\Big\{\Big[S_\alpha S_\beta\tilde\rho(t) - S_\beta\tilde\rho(t)S_\alpha\Big] w^+_{\alpha\beta} + \Big[\tilde\rho(t)S_\beta S_\alpha - S_\alpha\tilde\rho(t)S_\beta\Big] w^-_{\beta\alpha}\Big\} I(\omega_\alpha+\omega_\beta). \tag{4.260}$$

Under the assumption of fast (uncoupled) system and reservoir dynamics,

$$\Delta t \gg (\omega_\alpha+\omega_\beta)^{-1}, \tag{4.261}$$

the integral $I(\omega_\alpha+\omega_\beta)$ averages to zero unless $\omega_\alpha+\omega_\beta = 0$. Thus we may replace the integral with a Kronecker delta,

$$I(\omega_\alpha+\omega_\beta) = \Delta t\,\delta(\omega_\alpha, -\omega_\beta). \tag{4.262}$$

Now formally taking the limit of small Δt,

$$\partial_t \tilde{\rho}(t) \approx \frac{\Delta\tilde{\rho}(t)}{\Delta t} = -\delta(\omega_\alpha, -\omega_\beta)\Big\{\Big[S_\alpha S_\beta \tilde{\rho}(t) - S_\beta \tilde{\rho}(t) S_\alpha\Big] w^+_{\alpha\beta} + \Big[\tilde{\rho}(t) S_\beta S_\alpha - S_\alpha \tilde{\rho}(t) S_\beta\Big] w^-_{\beta\alpha}\Big\}, \tag{4.263}$$

where again we must keep in mind that this differential equation is coarse-grained in the sense of not representing dynamics on time scales as short as τ_c or $(\omega_\alpha + \omega_\beta)^{-1}$ for different frequencies. Now transforming out of the interaction representation, using the assumption (4.257) and $\omega_\alpha = -\omega_\beta$,

$$\partial_t \rho(t) = -\frac{i}{\hbar}[H_S, \rho(t)] - \delta(\omega_\alpha, -\omega_\beta)\Big\{\Big[S_\alpha S_\beta \rho(t) - S_\beta \rho(t) S_\alpha\Big] w^+_{\alpha\beta} + \Big[\rho(t) S_\beta S_\alpha - S_\alpha \rho(t) S_\beta\Big] w^-_{\beta\alpha}\Big\}. \tag{4.264}$$

Now we use the fact that H_{SR} is Hermitian, so terms of the form SR in (4.254) that are not Hermitian must be accompanied by their adjoint terms $S^\dagger R^\dagger$. Clearly, terms where $S_\alpha = S^\dagger_\beta$ satisfy $\delta(\omega_\alpha, -\omega_\beta) = 1$, so we can explicitly combine these pairs of terms to write the master equation in terms of only a single sum:

$$\partial_t \rho(t) = -\frac{i}{\hbar}[H_S, \rho(t)] + \sum_\alpha \Big\{\Big[S_\alpha \rho(t) S^\dagger_\alpha - S^\dagger_\alpha S_\alpha \rho(t)\Big] w^+_\alpha + \Big[S_\alpha \rho(t) S^\dagger_\alpha - \rho(t) S^\dagger_\alpha S_\alpha\Big] w^-_\alpha\Big\}. \tag{4.265}$$

Of course, terms of the same form carry through when S_α is Hermitian. In the expression above we have also defined the reduced integrals

$$\begin{aligned} w^+_\alpha &:= \int_0^\infty d\tau\, e^{-i\omega_\alpha \tau} \Big\langle \tilde{R}^\dagger_\alpha(\tau) \tilde{R}_\alpha(0)\Big\rangle_R \\ w^-_\alpha &:= \int_0^\infty d\tau\, e^{i\omega_\alpha \tau} \Big\langle \tilde{R}^\dagger_\alpha(0) \tilde{R}_\alpha(\tau)\Big\rangle_R = [w^+_\alpha]^*. \end{aligned} \tag{4.266}$$

Note that other cross-terms could in principle occur in Eq. (4.264) that satisfy $\omega_\alpha = -\omega_\beta$, which we appear to be missing here. However, if we end up with terms like $S_1 \rho S_2^\dagger$, this can always be absorbed into terms of the form $(S_1 + S_2)\rho(S_1 + S_2)^\dagger$, representing interferences in the couplings represented by $S_{1,2}$. The cross terms are weighted by a cross-correlation function between R_1 and R_2, representing the cross terms of the coherence. In the absence of cross coherence, only terms of the form $S_1 \rho S_1^\dagger$ and $S_2 \rho S_2^\dagger$ should appear. Weighted combinations of these terms with $(S_1 + S_2)\rho(S_1 + S_2)^\dagger$ terms can account for any degree of coherence. (See Section 6.2.4.1 for a discussion of interference contributions of this form in the context of quantum beats in three-level atoms.)

Now separating out the real and imaginary parts of the integrals (4.266) in (4.265),

$$\partial_t \rho(t) = -\frac{i}{\hbar}[H_S, \rho(t)] + \sum_\alpha 2\mathrm{Re}[w^+_\alpha]\Big\{S_\alpha \rho(t) S^\dagger_\alpha - \frac{1}{2}\Big[S^\dagger_\alpha S_\alpha \rho(t) + \rho(t) S^\dagger_\alpha S_\alpha\Big]\Big\} - i\sum_\alpha \mathrm{Im}[w^+_\alpha]\Big[S^\dagger_\alpha S_\alpha, \rho(t)\Big]. \tag{4.267}$$

Note that the last term has the form of Hamiltonian evolution, while the second term does not; these represent energy shifts and dissipation/diffusion effects, respectively, due to the interaction with the reservoir. Now separating out the real and imaginary parts of the integrals, we have the final result

$$\partial_t \rho(t) = -\frac{i}{\hbar}[H_S + H_{\mathrm{eff}}, \rho(t)] + \sum_\alpha k_\alpha D[S_\alpha]\rho(t),$$

(Born–Markov master equation) (4.268)

where the effective Hamiltonian for the reservoir interaction, leading to a "generalized Lamb shift," is

$$H_{\mathrm{eff}} := \hbar \sum_\alpha \mathrm{Im}[w^+_\alpha] S^\dagger_\alpha S_\alpha,$$

(effective Hamiltonian for generalized Lamb shift) (4.269)

and we have defined the **Lindblad superoperator**

$$D[c]\rho := c\rho c^\dagger - \frac{1}{2}\left[c^\dagger c\rho + \rho c^\dagger c\right]. \tag{4.270}$$
(Lindblad superoperator)

with coefficient

$$k_\alpha := 2\mathrm{Re}[w_\alpha^+]. \tag{4.271}$$
(dissipation/diffusion coupling coefficient)

We have thus arrived at the general **Lindblad form** of the master equation in the Born–Markov approximation, which we return to and justify in the context of measurement in Section 19.1. Again, the system operators S_α represent the coupling channel of the system to the reservoir, and thus the channel by which the system may be observed. Thus, for example, if $S_\alpha \longrightarrow x$, then we have the master equation for a position measurement, whereas if $S_\alpha \longrightarrow a$, where a is the annihilation operator for the harmonic oscillator, then we have the master equation for energy loss (and thus damping) of a quantum harmonic oscillator.

4.6 Exercises

Problem 4.1

(a) Using the expression for the Wigner function in terms of a pure state, argue that the Wigner function corresponds to an overlap integral of two wave functions, and thus derive the upper bound

$$|W(x,p)| \leq \frac{1}{\pi\hbar} \tag{4.272}$$

for the magnitude.

(b) Derive a similar upper bound for the Husimi distribution $W_{\rm H}(x,p)$.

Problem 4.2

For a harmonic oscillator of frequency ω and mass m, the density operator for a thermal state of temperature T is given by the Boltzmann-type sum

$$\rho = \left(1 - e^{-\hbar\omega/k_{\rm B}T}\right) \sum_{n=0}^{\infty} e^{-n\hbar\omega/k_{\rm B}T} |n\rangle\langle n|. \tag{4.273}$$

Carry out the appropriate summation over the Wigner functions for the harmonic-oscillator eigenstates to show that the thermal state is Gaussian.[36] What are the variances of the thermal state? Show that your variance expressions are sensible in the limits of low and high temperature.

It may help to know that the summation formula

$$\sum_{j=0}^{\infty} e^{-jx} \frac{(n+j)!}{j!} = n!(1-e^{-x})^{-(1+n)}, \tag{4.274}$$

valid for nonnegative n, is not difficult to prove by induction. (Translation: if you use it, you should prove it by induction.)

Problem 4.3

Derive the Weyl correspondences

$$A(x,p) = B(x,p) \exp\left[\frac{\hbar}{2i}\left(\overleftarrow{\partial_p}\overrightarrow{\partial_x} - \overleftarrow{\partial_x}\overrightarrow{\partial_p}\right)\right] C(x,p) \tag{4.275}$$

and

$$A(x,p) = B\left(x - \frac{\hbar}{2i}\partial_p, p + \frac{\hbar}{2i}\partial_x\right) C(x,p) = C\left(x + \frac{\hbar}{2i}\partial_p, p - \frac{\hbar}{2i}\partial_x\right) B(x,p) \tag{4.276}$$

for the operator product $\hat{A} = \hat{B}\hat{C}$.

Problem 4.4

Let $\mathbf{A}$ be a symplectic matrix, so that $\mathbf{A}$ satisfies $\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^{\mathsf{T}} = \boldsymbol{\Omega}$.

(a) Show that $\mathbf{A}^{-1} = -\boldsymbol{\Omega}\mathbf{A}^{\mathsf{T}}\boldsymbol{\Omega}$.

(b) Show that $\mathbf{A}^{\mathsf{T}}$ is symplectic.

Problem 4.5

Let $\mathbf{A}$ be a symplectic matrix.

(a) Show that if λ is an eigenvalue of $\mathbf{A}$, then $1/\lambda$ is also an eigenvalue of $\mathbf{A}$.

(b) What is the determinant of $\mathbf{A}$?

(c) Give a physical interpretation of the eigenvalues and the determinant of $\mathbf{A}$.

[36] Note that in the Church of the Larger Hilbert Space, the thermal state corresponds to a two-mode (Gaussian) squeezed state—that is, when you trace over one of the modes, you can choose the variances such that you recover the thermal state for the remaining mode.

Problem 4.6

Show that under Hamiltonian evolution,

$$\partial_t \rho = -\frac{i}{\hbar}[H, \rho], \tag{4.277}$$

the purity $\mathrm{Tr}[\rho^2]$ is a constant of the motion.

Problem 4.7

(a) Show that for a single particle in the sinusoidal (pendulum) potential

$$V(x) = -\alpha \cos(kx), \tag{4.278}$$

the equation of motion for the Wigner function may be written

$$\partial_t W(x,p) = -\frac{p}{m}\partial_x W(x,p) + \frac{\alpha}{\hbar}\sin(kx)\left[W\left(x, p+\frac{\hbar k}{2}\right) - W\left(x, p-\frac{\hbar k}{2}\right)\right]. \tag{4.279}$$

(b) We can make this equation of motion look "more classical" by defining the **Wigner effective potential** by writing the Liouville-type equation

$$\partial_t W = -\frac{p}{m}\partial_x W + \partial_x V_{\mathrm{eff}} \partial_p W. \tag{4.280}$$

Write down an expression for $\partial_x V_{\mathrm{eff}}$, and then show that for the minimum-uncertainty Gaussian state with variance V_x and no covariance,

$$W(x,p) = \frac{1}{\pi\hbar}\exp\left[-\frac{(x-\langle x\rangle)^2}{2V_x} - \frac{2V_x(p-\langle p\rangle)^2}{\hbar^2}\right], \tag{4.281}$$

the effective potential can be written

$$V_{\mathrm{eff}} = -\alpha\cos(kx)\exp\left(-\frac{k^2 V_x}{2}\right)\frac{\sinh\left[\frac{2kV_x}{\hbar}(p-\langle p\rangle)\right]}{\frac{2kV_x}{\hbar}(p-\langle p\rangle)}. \tag{4.282}$$

(c) Argue that the last factor in the above expression is negligible for a localized wave packet, and then show that in this limit, the above effective potential is approximately equal to what we will call the **Ehrenfest effective potential** $V_{\mathrm{eff}}^{(\mathrm{E})} := \langle V(x)\rangle$, which follows from the Ehrenfest equation

$$\partial_t \langle p\rangle = -\langle \partial_x V(x)\rangle = -\partial_{\langle x\rangle} V_{\mathrm{eff}}^{(\mathrm{E})}(\langle x\rangle), \tag{4.283}$$

where it turns out the last equality holds for the Gaussian state in a cosine potential.

Problem 4.8

The master equation for a damped harmonic oscillator, coupled to a reservoir in the vacuum state in Lindblad form is

$$\partial_t \rho = -\frac{i}{\hbar}[H,\rho] + \kappa D[a]\rho, \tag{4.284}$$

where κ is the rate of energy decay of the oscillator, and H is the harmonic-oscillator Hamiltonian. Consider the master equation for a damped, *anharmonic* oscillator. A reasonable guess might be to take the same master equation as for the harmonic case, and simply take $H = p^2/2m + V(x)$, where $V(x)$ is an anharmonic potential, while keeping the damping terms the same (and assuming that a is still defined as an appropriate linear combination of x and p). Explain why the Born–Markov master-equation formalism does *not* lead to this master equation, and discuss the most general form for the master equation of a damped, anharmonic oscillator. (*Hint:* what is the spectrum of a harmonic oscillator? An anharmonic oscillator? What does a for a harmonic oscillator look like in the Heisenberg picture? What would it look like if generalized to an anharmonic oscillator? Make sure to keep track of the assumptions in the Born–Markov derivation.)

Chapter 5

Two-Level Atom Interacting with a Classical Field

The interaction of a two-level atom is one of *the* canonical problems in quantum optics. We will now be accounting for quantum coherences between the two levels, which we ignored in our rate-equation treatment in Chapter 3. The approaches we will use here are widely applicable, since a two-level atom is essentially equivalent to a qubit or a spin-1/2 particle. Less obvious is that equivalent phenomena also occur in seemingly different problems such as tunneling in the double-well potential, Bragg diffraction, and neutrino oscillations.[1]

5.1 Atom–Field Interaction

We begin our treatment with a general description of the atom-field interaction. We will assume the field is monochromatic with angular frequency ω to model the field due to a laser:

$$\mathbf{E}(t) = \hat{\varepsilon} E_0 \cos(\omega t). \tag{5.1}$$

Here, $\hat{\varepsilon}$ is the unit polarization vector of the field. Note that we are ignoring the spatial dependence of the field, only writing down the field at the location of the atom. This is appropriate in the **dipole approximation** or **long-wavelength approximation**, where we assume that the wavelength of the field is much longer than the size of the atom, so that we can neglect any variations of the field over the extent of the atom. This is generally appropriate for optical transitions, since atomic dimensions have Å scales, while optical wavelengths are hundreds of nm. Formally, as the name suggests, the dipole approximation corresponds to the lowest-order contribution in a multipole expansion of the atom–field interaction.

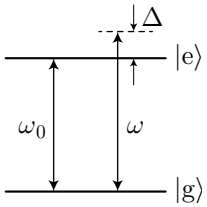
As before, it is convenient to decompose the field into its positive- and negative-rotating components $\mathbf{E}^{(+)}$ and $\mathbf{E}^{(-)}$:

$$\begin{aligned} \mathbf{E}(t) &= \hat{\varepsilon}\frac{E_0}{2}\left(e^{-i\omega t} + e^{i\omega t}\right) \\ &=: \mathbf{E}_0^{(+)} e^{-i\omega t} + \mathbf{E}_0^{(-)} e^{i\omega t} \\ &=: \mathbf{E}^{(+)}(t) + \mathbf{E}^{(-)}(t)\ . \end{aligned} \tag{5.2}$$

That is, $\mathbf{E}^{(\pm)} \sim e^{-i(\pm\omega)t}$.

We will treat the atom as a two-level atom. This is clearly an approximation to a true atom, which has an infinite set of bound states. The justification is that we will consider near-resonant interactions, so that the transitions to other levels are negligible. We will label the ground and excited levels as $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, respectively, and we will denote the resonant frequency by ω_0 (that is, the energy splitting of the pair of states is $\hbar\omega_0$).

[1] Alexander Friedland, "Evolution of the neutrino state inside the Sun," *Physical Review D* **64**, 013008 (2001) (doi: 10.1103/PhysRevD.64.013008).



We will define $\Delta := \omega - \omega_0$ to be the detuning of the laser field from the atomic resonance.

We can write the total Hamiltonian for the atom and field as a sum of the free atomic Hamiltonian $H_{\rm A}$ and the atom–field interaction Hamiltonian $H_{\rm AF}$:

$$H = H_{\rm A} + H_{\rm AF}. \tag{5.3}$$

The atomic free-evolution Hamiltonian is given by

$$H_{\rm A} = \hbar\omega_0 |{\rm e}\rangle\langle{\rm e}|, \tag{5.4}$$
(free evolution)

if we take the ground-state energy to be zero. The atom-field interaction Hamiltonian in the dipole approximation is

$$H_{\rm AF} = -\mathbf{d}\cdot\mathbf{E}, \tag{5.5}$$
(dipole interaction)

where $\mathbf{d}$ is the atomic dipole operator, given in terms of the atomic electron position $\mathbf{r}_{\rm e}$ as

$$\mathbf{d} = -e\mathbf{r}_{\rm e}, \tag{5.6}$$

if we assume a single electron for the atom (i.e., the field interacts predominantly with one electron). We denote the fundamental charge by e, so that the electron charge is $q = -e$. We will not justify this form of the interaction Hamiltonian for now, except to note that it seems reasonable as a dipole-field interaction, and also it is consistent with the classical case (see Problem 5.1).

5.1.1 Parity and the Dipole Operator

We can then use a simple parity argument to gain more information about the form of the dipole operator. The parity operator Π flips the sign of the position operator, and is thus defined by the operator transformation $\Pi\mathbf{r}_{\rm e}\Pi^\dagger = -\mathbf{r}_{\rm e}$. (Note that Π is *unitary*: Π and Π^{-1} correspond to the *same* operation or reversing the position coordinate, so that $\Pi = \Pi^{-1}$—Π is an *involution*—or $\Pi^2 = 1$.) Operating with Π on the right gives $\Pi\mathbf{r}_{\rm e} = -\mathbf{r}_{\rm e}\Pi$, and thus the anticommutator of the parity operator with $\mathbf{r}_{\rm e}$ vanishes:

$$[\Pi, \mathbf{r}_{\rm e}]_+ = 0. \tag{5.7}$$

The matrix elements of the anticommutator clearly vanish,

$$\langle a|[\Pi, \mathbf{r}_{\rm e}]_+|b\rangle = 0, \tag{5.8}$$

but we can also write the matrix elements in the energy basis as

$$\langle a|[\Pi, \mathbf{r}_{\rm e}]_+|b\rangle = \langle a|[\Pi\mathbf{r}_{\rm e} + \mathbf{r}_{\rm e}\Pi]|b\rangle = (\pi_a + \pi_b)\langle a|\mathbf{r}_{\rm e}|b\rangle, \tag{5.9}$$

where π_a and π_b are eigenvalues of Π. We can define these because Π commutes with the atomic Hamiltonian, which has the form $p_{\rm e}^2/2m_{\rm e} - \alpha/|\mathbf{r}_{\rm e}|$, and thus Π and H have simultaneous eigenstates. But $\Pi^2 = 1$, so the possible eigenvalues of Π are ± 1, corresponding to even (+) and odd (−) parity. For both (5.8) and (5.9) to hold, either $\pi_a + \pi_b = 0$ or $\langle a|\mathbf{r}_{\rm e}|b\rangle = 0$. This argument obviously applies just as well to the dipole operator instead of $\mathbf{r}_{\rm e}$. We can then see that the diagonal matrix elements of $\mathbf{d}$ vanish, since $\pi_{\rm g}$ and $\pi_{\rm e}$ are both nonzero:

$$\langle {\rm g}|\mathbf{d}|{\rm g}\rangle = \langle {\rm e}|\mathbf{d}|{\rm e}\rangle = 0. \tag{5.10}$$

The off-diagonal matrix elements $\langle g|\mathbf{d}|e\rangle = \langle e|\mathbf{d}|g\rangle^*$ are nonvanishing, however, *provided that the states have opposite parity*, $\pi_e = -\pi_g$. The point is that the dipole operator couples the ground and excited states, but does not produce any *first-order* shift of either state.

By applying the identity $|e\rangle\langle e| + |g\rangle\langle g|$ on both sides of $\mathbf{d}$, we see that dipole operator admits the decomposition

$$\mathbf{d} = \langle g|\mathbf{d}|e\rangle|g\rangle\langle e| + \langle e|\mathbf{d}|g\rangle|e\rangle\langle g|. \tag{5.11}$$

We can choose the phase of the dipole matrix element $\langle g|\mathbf{d}|e\rangle$ such that it is real, in which case we can write the dipole operator as

$$\mathbf{d} = \langle g|\mathbf{d}|e\rangle(\sigma + \sigma^\dagger), \tag{5.12}$$

(dipole operator in terms of σ)

where $\sigma := |g\rangle\langle e|$ is the atomic lowering operator. For clarity when we get to more complicated atoms, we will always write the dipole matrix element $\langle g|\mathbf{d}|e\rangle$ with the excited state to the right. We can thus write the total atom–field Hamiltonian as

$$H = H_A + H_{AF} = \hbar\omega_0\sigma^\dagger\sigma - \langle g|\mathbf{d}|e\rangle \cdot \mathbf{E}\left(\sigma + \sigma^\dagger\right), \tag{5.13}$$

where $\sigma^\dagger\sigma = |e\rangle\langle e|$ is the excited-state projection operator.

5.1.2 Rotating-Wave Approximation

Just as we decomposed the field into positive- and negative-rotating parts, we can do the same for the dipole operator in the form (5.12):

$$\begin{aligned}\mathbf{d} &= \langle g|\mathbf{d}|e\rangle(\sigma + \sigma^\dagger)\\ &= \mathbf{d}^{(+)} + \mathbf{d}^{(-)},\end{aligned} \tag{5.14}$$

where $\mathbf{d}^{(+)} \sim \sigma$ and $\mathbf{d}^{(-)} \sim \sigma^\dagger$. We do this because the *expectation value* of $\sigma = |g\rangle\langle e|$ has the unperturbed time dependence $e^{-i\omega_0 t}$ (since this is the evolution of $|e\rangle$ under the free atomic Hamiltonian), and thus corresponds to a positive frequency.

Including the same decomposition of the field, the atom–field Hamiltonian becomes

$$\begin{aligned}H_{AF} &= -(\mathbf{d}^{(+)} + \mathbf{d}^{(-)}) \cdot (\mathbf{E}^{(+)} + \mathbf{E}^{(-)})\\ &= -\mathbf{d}^{(+)} \cdot \mathbf{E}^{(+)} - \mathbf{d}^{(-)} \cdot \mathbf{E}^{(-)} - \mathbf{d}^{(+)} \cdot \mathbf{E}^{(-)} - \mathbf{d}^{(-)} \cdot \mathbf{E}^{(+)}.\end{aligned} \tag{5.15}$$

Recalling the time dependences

$$\mathbf{d}^{(\pm)} \sim e^{\mp i\omega_0 t}; \qquad \mathbf{E}^{(\pm)} \sim e^{\mp i\omega t}, \tag{5.16}$$

we see that the first two terms oscillate rapidly as $e^{\pm i(\omega+\omega_0)t}$, while the last two (cross) terms oscillate slowly as $e^{\pm i\Delta t}$. Assuming that $|\omega - \omega_0| \ll \omega + \omega_0$, we can make the **rotating-wave approximation** (RWA).[2] This approximation focuses on slow dynamics, replacing terms rotating at optical frequencies are replaced by their zero average value, which amounts to a coarse-graining on fs time scales. This is reasonable since optical detectors don't respond on fs time scales anyway.

Something that is sometimes not appreciated is that the two-level approximation and the RWA are at the same level of accuracy. That is, it makes no sense to throw one out and keep the other. Both amount to discarding interactions that are far off resonance (the RWA amounts to a very far detuned interaction of a positive-frequency resonance with a negative-frequency field). If the detuning is large enough that the counter-rotating term is not negligible, then neither are the couplings to the other levels.

[2]The name "rotating-wave approximation" comes from nuclear magnetic resonance. A spin-1/2 particle in a magnetic field precesses (rotates) naturally. When you hit it with a linearly polarized microwave field, the field is a superposition of two waves with opposite circular ("rotating") polarization. Only one of the polarization rotates in the same direction as the spin precession, and so the RWA amounts to ignoring the counter-rotating field component. Said another way, it amounts to replacing the linearly polarized wave with a rotating wave. Note that in NMR, the RWA is *exact* for circular polarization, a property that does not carry over to the optical case.

5.1.3 Rabi Frequency

Thus, the atom–field interaction Hamiltonian in the RWA becomes

$$H_{\text{AF}} = -\mathbf{d}^{(+)} \cdot \mathbf{E}^{(-)} - \mathbf{d}^{(-)} \cdot \mathbf{E}^{(+)}. \tag{5.17}$$
(RWA dipole interaction)

Using Eq. (5.2) for the explicit time-dependence of the field along with Eq. (5.14) for the dipole operator, we can write

$$\begin{aligned} H_{\text{AF}} &= -\langle \text{g}|\hat{\varepsilon} \cdot \mathbf{d}|\text{e}\rangle \left(E_0^{(-)} \sigma e^{i\omega t} + E_0^{(+)} \sigma^\dagger e^{-i\omega t} \right) \\ &= \frac{\hbar\Omega}{2} \left(\sigma e^{i\omega t} + \sigma^\dagger e^{-i\omega t} \right), \end{aligned} \tag{5.18}$$

where we have assumed $E_0^{(+)}$ to be real, and we have defined the **Rabi frequency**[3] as

$$\Omega := -\frac{2\langle \text{g}|\hat{\varepsilon} \cdot \mathbf{d}|\text{e}\rangle E_0^{(+)}}{\hbar} = -\frac{\langle \text{g}|\hat{\varepsilon} \cdot \mathbf{d}|\text{e}\rangle E_0}{\hbar}. \tag{5.19}$$
(Rabi frequency)

Note that we generally choose the phase of the dipole matrix element so that $\Omega > 0$. The Rabi frequency characterizes the strength of the atom–field coupling. In the case of a linearly polarized field, the Rabi frequency simplifies to

$$\Omega = -\frac{\langle \text{g}|d_z|\text{e}\rangle E_0}{\hbar} \tag{5.20}$$

if the field is polarized in the z-direction ($\hat{\varepsilon} = \hat{z}$).

5.1.4 Schrödinger Equation

Let's write the atomic state as

$$|\psi\rangle = c_\text{g}|\text{g}\rangle + c_\text{e}|\text{e}\rangle, \tag{5.21}$$

where c_g and c_e carry all the time dependence of the state. With the atomic Hamiltonian H_A (5.4) and interaction H_{AF} (5.18), the Schrödinger equation $i\hbar\partial_t|\psi\rangle = H|\psi\rangle$ then gives

$$\partial_t c_\text{g}|\text{g}\rangle + \partial_t c_\text{e}|\text{e}\rangle = -i\omega_0 c_\text{e}|\text{e}\rangle - i\frac{\Omega}{2}e^{i\omega t}c_\text{e}|\text{g}\rangle - i\frac{\Omega}{2}e^{-i\omega t}c_\text{g}|\text{e}\rangle. \tag{5.22}$$

Projecting with $\langle \text{g}|$ and $\langle \text{e}|$ gives the pair of coupled differential equations,

$$\begin{aligned} \partial_t c_\text{g} &= -i\frac{\Omega}{2}c_\text{e}e^{i\omega t} \\ \partial_t c_\text{e} &= -i\omega_0 c_\text{e} - i\frac{\Omega}{2}c_\text{g}e^{-i\omega t}, \end{aligned} \tag{5.23}$$

which we can now in principle solve for the atomic evolution.

5.1.5 Rotating Frame

We now have a set of coupled equations that involve oscillatory terms at optical frequencies. However, at resonance the precessions are phase-locked and should disappear in the proper coordinates. It it therefore convenient to transform into a corotating frame to eliminate the fast rotation. Thus, we make a transformation into the rotating frame of the laser field by defining the slowly varying excited-state amplitude

$$\tilde{c}_\text{e} := c_\text{e}e^{i\omega t}. \tag{5.24}$$
(rotating-frame transformation)

[3] after Isador Isaac Rabi, who pioneered the field of nuclear magnetic resonance. He was awarded the 1944 Nobel prize for this work.

We can then rewrite the equations of motion as

$$\begin{aligned}\partial_t c_{\mathrm{g}} &= -i\frac{\Omega}{2}\tilde{c}_{\mathrm{e}} \\ \partial_t \tilde{c}_{\mathrm{e}} &= i\Delta\tilde{c}_{\mathrm{e}} - i\frac{\Omega}{2}c_{\mathrm{g}}.\end{aligned} \tag{5.25}$$

In fact, these are the same equations of motion generated by the effective, rotating-frame atomic Hamiltonian

$$\tilde{H}_{\mathrm{A}} = -\hbar\Delta|\mathrm{e}\rangle\langle\mathrm{e}|, \tag{5.26}$$
(rotating-frame free evolution)

where recall that $\Delta := \omega - \omega_0$ is the detuning of the laser from the atomic resonance, and the effective, rotating-frame interaction Hamiltonian

$$\begin{aligned}\tilde{H}_{\mathrm{AF}} &= -\mathbf{d}^{(+)}\cdot\tilde{\mathbf{E}}^{(-)} - \mathbf{d}^{(-)}\cdot\tilde{\mathbf{E}}^{(+)} \\ &= \frac{\hbar\Omega}{2}(\sigma + \sigma^\dagger),\end{aligned} \tag{5.27}$$
(rotating-frame dipole interaction)

where we have defined the stationary field amplitudes

$$\tilde{E}^{(\pm)} := e^{\pm i\omega t}E^{(\pm)}. \tag{5.28}$$

In making the rotating-wave approximation, we have discarded the two terms that would have an explicit time dependence of $e^{\pm i2\omega t}$ in Eq. (5.27), and in transforming to the rotating frame, we have removed *all* of the explicit time dependence from this problem. Essentially, we are hiding the time dependence of the field in the slowly varying amplitude $\tilde{c}_{\mathrm{e}}$, where it cancels most of the natural state evolution. Notice also that $|\mathrm{e}\rangle$ is still an eigenstate of $\tilde{H}_{\mathrm{A}}$, with eigenvalue $\hbar\omega_0 - \hbar\omega = -\hbar\Delta$, so that the rotating-frame transformation has the effect of shifting the excited state down in energy by an amount $\hbar\omega$.

This representation of the problem in the laser frame is interesting when we look at it this way: it shows that this *ac* interaction is equivalent to the problem of two states separated in energy by $\hbar\Delta$ interacting with a *dc* electric field (in the RWA). Because we have eliminated any explicit time dependence, this problem will be easy to solve.

5.1.5.1 Unitary Transformations

Suppose that we have a unitary transformation U, which induces the transformation $|\tilde{\psi}\rangle = U|\psi\rangle$. Then how does the Hamiltonian transform? Both the original and transformed states must satisfy the Schrödinger equation,

$$i\hbar\partial_t|\psi\rangle = H|\psi\rangle, \qquad i\hbar\partial_t|\tilde{\psi}\rangle = \tilde{H}|\tilde{\psi}\rangle, \tag{5.29}$$

where $\tilde{H}$ is the transformed Hamiltonian. We can write the first equation here as

$$i\hbar\partial_t(U^\dagger|\tilde{\psi}\rangle) = HU^\dagger|\tilde{\psi}\rangle. \tag{5.30}$$

Then we can expand the time derivative and operate on the left by U:

$$i\hbar\partial_t|\tilde{\psi}\rangle + i\hbar U\partial_t U^\dagger|\tilde{\psi}\rangle = UHU^\dagger|\tilde{\psi}\rangle. \tag{5.31}$$

Noting that $\partial_t(UU^\dagger) = (\partial_t U)U^\dagger + U\partial_t U^\dagger = 0$, we can rewrite this as

$$i\hbar\partial_t|\tilde{\psi}\rangle = \left[UHU^\dagger + i\hbar(\partial_t U)U^\dagger\right]|\tilde{\psi}\rangle. \tag{5.32}$$

Comparing this to the Schrödinger equation in the transformed variables, we can identify the transformation law

$$\tilde{H} = UHU^\dagger + i\hbar(\partial_t U)U^\dagger \tag{5.33}$$
(time-dependent transformation)

for the Hamiltonian under a general time-dependent, unitary transformation.

Thus we can see that the transformation to the rotating frame is represented by the unitary transformation

$$U = \exp\left(i\omega t|\mathrm{e}\rangle\langle\mathrm{e}|\right). \tag{5.34}$$

This gives the proper form for the rotating-frame state

$$\begin{aligned} |\tilde{\psi}\rangle &= U|\psi\rangle \\ &= U\Big(c_{\mathrm{g}}|\mathrm{g}\rangle + c_{\mathrm{e}}|\mathrm{e}\rangle\Big) \\ &= c_{\mathrm{g}}|\mathrm{g}\rangle + c_{\mathrm{e}}e^{i\omega t}|\mathrm{e}\rangle \\ &= c_{\mathrm{g}}|\mathrm{g}\rangle + \tilde{c}_{\mathrm{e}}|\mathrm{e}\rangle, \end{aligned} \tag{5.35}$$

and it also gives the proper forms for $\tilde{H}_{\mathrm{A}}$ and $\tilde{H}_{\mathrm{AF}}$ from Eqs. (5.26) and (5.27), respectively, when we use the transformation law (5.33) is applied to $H_{\mathrm{A}} + H_{\mathrm{AF}}$.

5.1.5.2 Digression: Field Operators

This rotating-frame transformation, where the atomic ground and excited states become nearly degenerate, has an analogous interpretation in the fully quantum treatment of the problem. The key point is that in a fully quantum treatment, the electric field is represented by an operator. We will defer the detailed derivations until later, but the basic idea is that a single-mode field is represented quantum-mechanically by a harmonic oscillator of frequency ω, where the nth energy level corresponds to the presence of n photons. The other important thing to realize is that the positive-rotating field amplitude is proportional to the field annihilation operator, $E^{(+)} \sim a$, where a is given by in terms of the photon number states as

$$a = \sum_{n=1}^{\infty} |n-1\rangle\langle n|\sqrt{n}, \tag{5.36}$$

while the negative-rotating amplitude corresponds to the creation operator, $E^{(-)} \sim a^{\dagger}$. It is not hard to see, for example, that generic expectation values of the field operators have the right time dependence to match the classical case.

In terms of the quantized field, the combined Hamiltonian becomes

$$\begin{aligned} H_{\mathrm{quantum}} &= H_{\mathrm{A}} + H_{\mathrm{AF}} \\ &= \hbar\omega_0|\mathrm{e}\rangle\langle\mathrm{e}| + \hbar g\left(\sigma a^{\dagger}e^{i\omega t} + \sigma^{\dagger}ae^{-i\omega t}\right), \end{aligned} \tag{5.37}$$

where $2g$ is often called the **one-photon Rabi frequency**, as we will see in our discussion of the Jaynes–Cummings model in Chapter 10. This is very similar in form to the semiclassical Hamiltonian, where the atom–field interaction is given by Eq. (5.18):

$$\begin{aligned} H_{\mathrm{semiclassical}} &= H_{\mathrm{A}} + H_{\mathrm{AF}} \\ &= \hbar\omega_0|\mathrm{e}\rangle\langle\mathrm{e}| + \frac{\hbar\Omega}{2}\left(\sigma e^{i\omega t} + \sigma^{\dagger}e^{-i\omega t}\right), \end{aligned} \tag{5.38}$$

except for the presence of the field operators and the different coupling constant.

This quantum Hamiltonian is in the interaction picture with respect to the field evolution. We can see this because the field operators carry the explicit time dependence of the field (written out explicitly here), and because the field should evolve according to a harmonic-oscillator Hamiltonian,

$$H_{\mathrm{F}} = \hbar\omega\left(a^{\dagger}a + \frac{1}{2}\right), \tag{5.39}$$

which is obviously missing here. We can transform out of the interaction picture by using the transformation

$$U_{\rm I} = \exp\left(iH_{\rm F}t/\hbar\right). \tag{5.40}$$

The resulting Hamiltonian is

$$\begin{aligned}\tilde{H}_{\rm quantum} &= \hbar\omega_0|{\rm e}\rangle\langle{\rm e}| + \hbar g\left(\sigma a^\dagger + \sigma^\dagger a\right) + \hbar\omega\left(a^\dagger a + \frac{1}{2}\right)\\ &= \tilde{H}_{\rm A} + \tilde{H}_{\rm AF} + \tilde{H}_{\rm F},\end{aligned} \tag{5.41}$$

where the tildes indicate operators after transformation (i.e., in the Schrödinger picture), and we have hidden constant phase factors in the atom–field Hamiltonian. This Hamiltonian is then in the Schrödinger picture with respect to the field, where the field time dependence is generated by the presence of $H_{\rm F}$. Note that we didn't shift the energy of the excited state down, but in a sense we shifted the energy of the ground state up, which amounts to the same thing. This happened because the Hamiltonian here couples states of the form

$$|{\rm g}, n+1\rangle \longrightarrow |{\rm e}, n\rangle, \tag{5.42}$$

where the integer refers to photon-number states. The splitting between these states is $-\hbar\Delta$, as we saw in the semiclassical rotating frame.

In the classical limit, the average photon number N of the laser field is very large, and in a coherent (Gaussian) state of the field, the fractional uncertainty in N becomes vanishingly small. Hence, the field operator a can be replaced by $\sqrt{N}$. With the argument above, the free atom and field Hamiltonians conspire to give an atomic Hamiltonian with the correct splitting in the rotating frame. Upon making the identification $\Omega/2 = g\sqrt{N}$, we also recover the correct form for the rotating interaction Hamiltonian (5.27). Hence we have shown that the transformation to the rotating frame also arises as a transformation from the interaction picture to the Schrödinger picture with respect to the quantized field. We have also established explicitly how the fully quantized atom–field treatment reduces to the present semiclassical model in the classical-field limit.

With the expressions (5.37) and (5.38) for the atom–field interaction in hand, we can make one final remark about the RWA. Since $\mathbf{E}^{(+)}$ annihilates a photon from the laser field, the terms left in the interaction Hamiltonian from the RWA correspond to raising the atomic state while lowering the field state ($d_z^{(-)}E^{(+)} \longrightarrow \sigma^\dagger a$) and lowering the atomic state while raising the field state ($d_z^{(+)}E^{(-)} \longrightarrow \sigma a^\dagger$). Near resonance, these interactions are energy-conserving. The rotating terms that we neglected are of the form ($d_z^{(+)}E^{(+)} \longrightarrow \sigma a$) and ($d_z^{(-)}E^{(-)} \longrightarrow \sigma^\dagger a^\dagger$), corresponding to lowering the atom but annihilating a photon, and raising the atom but creating a photon, respectively. These processes violate energy conservation by an energy of about two photons' worth, and should thus be much less important than the energy-conserving ones. Invoking the RWA, then, amounts to keeping only the energy-conserving (resonant) terms in the interaction Hamiltonian.

5.2 Rabi Flopping

Now we can solve the coupled amplitude equations (5.25) in the rotating frame to look at the driven atomic dynamics. We will do this first in the case of exact resonance, and then in the more general, nearly resonant case.

5.2.1 Resonant Interaction

In the case of exact resonance ($\Delta = 0$), the coupled equations reduce to

$$\begin{aligned}\partial_t c_{\rm g} &= -i\frac{\Omega}{2}\tilde{c}_{\rm e}\\ \partial_t \tilde{c}_{\rm e} &= -i\frac{\Omega}{2}c_{\rm g}.\end{aligned} \tag{5.43}$$

We can easily decouple these by differentiating them and substituting in the original equations. For example,

$$\partial_t^2 c_{\rm g} = -i\frac{\Omega}{2}\partial_t \tilde{c}_{\rm e} = -\left(\frac{\Omega}{2}\right)^2 c_{\rm g}. \tag{5.44}$$

This has the form of an undamped harmonic oscillator of frequency $\Omega/2$. The equations of motion are invariant under the exchange g $\longleftrightarrow$ e, and so $c_{\rm e}$ satisfies the same uncoupled equation as $c_{\rm g}$. Thus, we may write the uncoupled equations as

$$\begin{aligned} \partial_t^2 c_{\rm g} &= -\left(\frac{\Omega}{2}\right)^2 c_{\rm g} \\ \partial_t^2 \tilde{c}_{\rm e} &= -\left(\frac{\Omega}{2}\right)^2 \tilde{c}_{\rm e}. \end{aligned} \tag{5.45}$$

The general solution for $c_{\rm g}$ is

$$c_{\rm g}(t) = A\sin\left(\frac{1}{2}\Omega t\right) + B\cos\left(\frac{1}{2}\Omega t\right). \tag{5.46}$$

To put this solution in terms of the initial conditions $c_{\rm g}(0)$ and $\tilde{c}_{\rm e}(0)$, we first set $t = 0$ in the $c_{\rm g}$ solution to find that $c_{\rm g}(0) = B$. Differentiating $c_{\rm g}(t)$ and using the original equation of motion for $c_{\rm g}$, we find

$$\partial_t c_{\rm g}(t) = \frac{1}{2}\Omega\left[A\cos\left(\frac{1}{2}\Omega t\right) - B\sin\left(\frac{1}{2}\Omega t\right)\right] = -\frac{i}{2}\Omega\tilde{c}_{\rm e}(t). \tag{5.47}$$

Setting $t = 0$ then gives $A = -i\tilde{c}_{\rm e}(0)$. Thus, the general solution for $c_{\rm g}$ comes from Eq. (5.46), and the general solution for $\tilde{c}_{\rm e}$ comes from Eq. (5.47), which gives

$$\begin{aligned} c_{\rm g}(t) &= c_{\rm g}(0)\cos\left(\frac{1}{2}\Omega t\right) - i\tilde{c}_{\rm e}(0)\sin\left(\frac{1}{2}\Omega t\right) \\ \tilde{c}_{\rm e}(t) &= \tilde{c}_{\rm e}(0)\cos\left(\frac{1}{2}\Omega t\right) - ic_{\rm g}(0)\sin\left(\frac{1}{2}\Omega t\right) \end{aligned}$$

(two-level atom solution, $\Delta = 0$) (5.48)

as the general solution for the two amplitudes.

5.2.1.1 Example: Initially Unexcited Atom

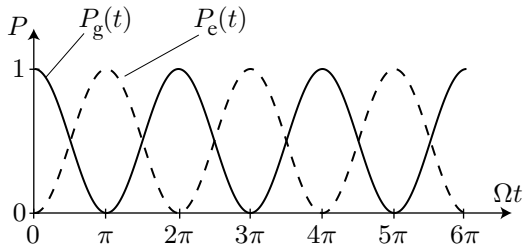
For an atom initially in the ground state, $c_{\rm g}(0) = 1$ and $\tilde{c}_{\rm e}(0) = 0$. The general solution then becomes

$$\begin{aligned} c_{\rm g}(t) &= \cos\left(\frac{1}{2}\Omega t\right) \\ \tilde{c}_{\rm e}(t) &= -i\sin\left(\frac{1}{2}\Omega t\right). \end{aligned} \tag{5.49}$$

The ground- and excited-state populations are thus

$$\begin{aligned} P_{\rm g}(t) &= |c_{\rm g}(t)|^2 = \cos^2\left(\frac{1}{2}\Omega t\right) = \frac{1}{2}(1+\cos\Omega t) \\ P_{\rm e}(t) &= |\tilde{c}_{\rm e}(t)|^2 = \sin^2\left(\frac{1}{2}\Omega t\right) = \frac{1}{2}(1-\cos\Omega t). \end{aligned} \tag{5.50}$$

Thus, we see explicitly the significance of the Rabi frequency: the population oscillates between the ground and excited levels at the angular frequency Ω. This oscillation phenomenon is referred to as **Rabi flopping**.



Roughly, the ''upswings'' in population from $|g\rangle$ to $|e\rangle$ correspond to light absorption, while the downswings correspond to stimulated emission back into the original field (thus far we are ignoring *spontaneous* emission). The period of the oscillation is $T = 2\pi/\Omega$, and thus some particular times are important. If the field is turned on for a duration $T/2$ [i.e., $\Omega(T/2) = \pi$], an atom initially in the ground state is promoted to the excited state with unit probability. A pulse of this form is called a **π-pulse**. On the other hand, if the field is turned on for a duration $T/4$ [i.e., $\Omega(T/4) = \pi/2$], an atom initially in the ground state ends up in a superposition of the ground and excited states. A pulse of this form is called a **$\pi/2$-pulse**.

Of course, these ''pulses'' refer to square-profile pulses, since we have assumed a constant field amplitude E_0, but more generally we can consider the above analysis to be valid even when the field amplitude is time-dependent, so long as it varies slowly on the optical time scales. In this case, the Rabi frequency itself is time-dependent, and we can generally define a π-pulse to be any pulse with an ''area'' of π,

$$\int dt\, \Omega(t) = \pi, \tag{5.51}$$

and a $\pi/2$ pulse is any pulse with area $\pi/2$.

5.2.2 Nearly Resonant Interaction

In the case of nonzero detuning Δ, we need to solve Eqs. (5.25). We can again decouple these by differentiating and eliminating appropriate variables, with the result

$$\begin{aligned} \left(\partial_t^2 - i\Delta\partial_t + \frac{\Omega^2}{4}\right) c_g &= 0 \\ \left(\partial_t^2 - i\Delta\partial_t + \frac{\Omega^2}{4}\right) \tilde{c}_e &= 0. \end{aligned} \tag{5.52}$$

Rewriting the decoupled equations as

$$\begin{aligned} \left(\partial_t - i\frac{\Delta}{2} + i\frac{\tilde{\Omega}}{2}\right)\left(\partial_t - i\frac{\Delta}{2} - i\frac{\tilde{\Omega}}{2}\right) c_g &= 0 \\ \left(\partial_t - i\frac{\Delta}{2} + i\frac{\tilde{\Omega}}{2}\right)\left(\partial_t - i\frac{\Delta}{2} - i\frac{\tilde{\Omega}}{2}\right) \tilde{c}_e &= 0, \end{aligned} \tag{5.53}$$

where $\tilde{\Omega}$ is the **generalized Rabi frequency**,

$$\tilde{\Omega} := \sqrt{\Omega^2 + \Delta^2}, \tag{5.54}$$

it is easy to see that any function that causes either factor to vanish will solve the equation, and thus the solutions are linear combinations of functions of the form $\exp(i\Delta t/2 \pm i\tilde{\Omega}t/2)$. We will thus write the solutions as

$$\begin{aligned} c_g(t) &= e^{i\Delta t/2}\left[A_g \cos\left(\frac{1}{2}\tilde{\Omega}t\right) + B_g \sin\left(\frac{1}{2}\tilde{\Omega}t\right)\right] \\ \tilde{c}_e(t) &= e^{i\Delta t/2}\left[A_e \cos\left(\frac{1}{2}\tilde{\Omega}t\right) + B_e \sin\left(\frac{1}{2}\tilde{\Omega}t\right)\right]. \end{aligned} \tag{5.55}$$

Setting $t = 0$ in these equations gives

$$A_\mathrm{g} = c_\mathrm{g}(0), \qquad A_\mathrm{e} = \tilde{c}_\mathrm{e}(0), \tag{5.56}$$

while differentiating the solutions and comparing to the equations of motion gives

$$\begin{aligned} \partial_t c_\mathrm{g}(0) &= i\frac{\Delta}{2}c_\mathrm{g}(0) + \frac{\tilde{\Omega}}{2}B_\mathrm{g} = -i\frac{\Omega}{2}\tilde{c}_\mathrm{e}(0) \qquad &\implies B_\mathrm{g} &= -\frac{i}{\tilde{\Omega}}\left[\Delta c_\mathrm{g}(0) + \Omega\tilde{c}_\mathrm{e}(0)\right] \\ \partial_t \tilde{c}_\mathrm{e}(0) &= i\frac{\Delta}{2}\tilde{c}_\mathrm{e}(0) + \frac{\tilde{\Omega}}{2}B_\mathrm{e} = i\Delta\tilde{c}_\mathrm{e}(0) - i\frac{\Omega}{2}\tilde{c}_\mathrm{e}(0) &\implies B_\mathrm{e} &= \frac{i}{\tilde{\Omega}}\left[\Delta\tilde{c}_\mathrm{e}(0) - \Omega c_\mathrm{g}(0)\right]. \end{aligned} \tag{5.57}$$

Thus, the general solution in terms of initial conditions reads

$$\begin{aligned} c_\mathrm{g}(t) &= e^{i\Delta t/2}\left[c_\mathrm{g}(0)\cos\left(\frac{1}{2}\tilde{\Omega}t\right) - \frac{i}{\tilde{\Omega}}\left[\Delta c_\mathrm{g}(0) + \Omega c_\mathrm{e}(0)\right]\sin\left(\frac{1}{2}\tilde{\Omega}t\right)\right] \\ \tilde{c}_\mathrm{e}(t) &= e^{i\Delta t/2}\left[\tilde{c}_\mathrm{e}(0)\cos\left(\frac{1}{2}\tilde{\Omega}t\right) + \frac{i}{\tilde{\Omega}}\left[\Delta\tilde{c}_\mathrm{e}(0) - \Omega c_\mathrm{g}(0)\right]\sin\left(\frac{1}{2}\tilde{\Omega}t\right)\right]. \end{aligned}$$

(two-level atom solution, arbitrary Δ) (5.58)

5.2.2.1 Example: Initially Unexcited Atom

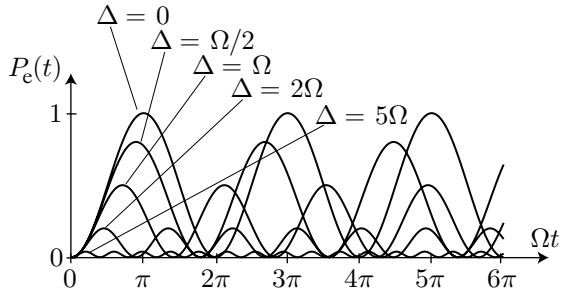
Again, for an atom initially in the ground state, $c_\mathrm{g}(0) = 1$ and $\tilde{c}_\mathrm{e}(0) = 0$, and the general solution becomes

$$\begin{aligned} c_\mathrm{g}(t) &= e^{i\Delta t/2}\left[\cos\left(\frac{1}{2}\tilde{\Omega}t\right) - i\frac{\Delta}{\tilde{\Omega}}\sin\left(\frac{1}{2}\tilde{\Omega}t\right)\right] \\ \tilde{c}_\mathrm{e}(t) &= -ie^{i\Delta t/2}\frac{\Omega}{\tilde{\Omega}}\sin\left(\frac{1}{2}\tilde{\Omega}t\right). \end{aligned} \tag{5.59}$$

The excited-state population is thus

$$P_\mathrm{e}(t) = \frac{\Omega^2}{\tilde{\Omega}^2}\sin^2\left(\frac{1}{2}\tilde{\Omega}t\right) = \frac{\Omega^2}{\tilde{\Omega}^2}\left(\frac{1}{2} - \frac{1}{2}\cos\tilde{\Omega}t\right). \tag{5.60}$$

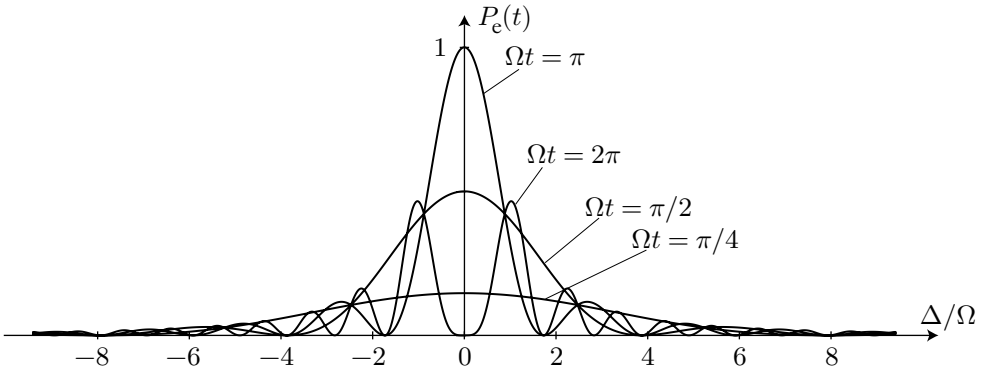
We can thus notice two things. First, the Rabi oscillations now occur at the generalized Rabi frequency $\tilde{\Omega} \geq \Omega$, so the oscillation rate increases as the magnitude of the detuning increases. For weak fields ($|\Delta| \gg \Omega$), $\tilde{\Omega} \approx |\Delta|$, while for strong fields ($|\Delta| \ll \Omega$), $\tilde{\Omega} \approx \Omega$.



The other thing to notice is that the amplitude of the oscillations is reduced. The modulation depth (maximum excitation probability) is

$$\frac{\Omega^2}{\tilde{\Omega}^2} = \frac{\Omega^2}{\Omega^2 + \Delta^2}, \tag{5.61}$$

which reduces to Ω^2/Δ^2 for weak excitation and to 1 for strong excitation.



5.3 Dressed States

One thing that we can conclude from the Rabi-flopping behavior is that the old eigenstates $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ are no longer eigenstates of the coupled system. However, we can still find the new eigenstates. We start by writing the equations of motion (5.23) for the coefficients in the matrix form

$$\partial_t \begin{bmatrix} \tilde{c}_\mathrm{e} \\ c_\mathrm{g} \end{bmatrix} = -i \begin{bmatrix} -\Delta & \Omega/2 \\ \Omega/2 & 0 \end{bmatrix} \begin{bmatrix} \tilde{c}_\mathrm{e} \\ c_\mathrm{g} \end{bmatrix} = -\frac{i}{\hbar}\tilde{H} \begin{bmatrix} \tilde{c}_\mathrm{e} \\ c_\mathrm{g} \end{bmatrix}, \tag{5.62}$$

where we identify the rotating-frame Hamiltonian

$$\tilde{H} = \left(\tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF}\right) = \hbar \begin{bmatrix} -\Delta & \Omega/2 \\ \Omega/2 & 0 \end{bmatrix} \tag{5.63}$$

in the uncoupled energy basis. We leave the diagonalization of this Hamiltonian as an exercise, but it can be shown that the eigenvalues are

$$E_\pm = -\frac{\hbar\Delta}{2} \pm \frac{\hbar\tilde{\Omega}}{2}, \tag{5.64}$$

(dressed-state energies)

and the corresponding eigenvectors are given by an effective rotation of the uncoupled states,

$$\begin{aligned} |+\rangle &= \sin\theta|\mathrm{g}\rangle + \cos\theta|\mathrm{e}\rangle \\ |-\rangle &= \cos\theta|\mathrm{g}\rangle - \sin\theta|\mathrm{e}\rangle, \end{aligned} \tag{5.65}$$

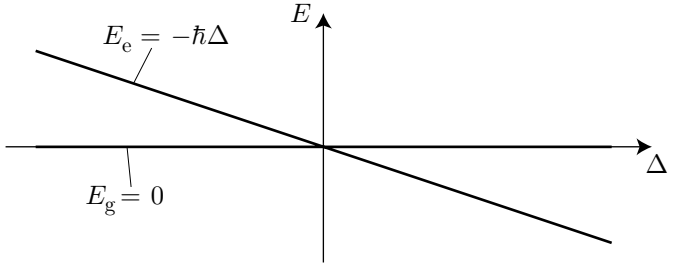
(dressed states)

where by convention the state $|+\rangle$ has the higher energy, and the **Stückelberg angle** θ is defined via

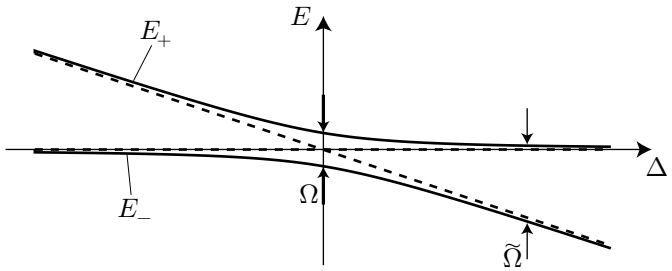
$$\tan 2\theta = -\frac{\Omega}{\Delta} \quad \left(0 \le \theta < \frac{\pi}{2}\right). \tag{5.66}$$

(Stückelberg angle)

These are the **dressed states** of the atom, and we see from Eq. (5.64) that the coupling to the field causes an **avoided crossing** in the energy level structure of the atom. That is, in the uncoupled case, the energies of the ground and excited states—the **bare states**—are 0 and $-\hbar\Delta$, respectively. The energy curves cross when $\Delta = 0$.



The coupling between the two levels lifts the degeneracy, however, converting the level crossing into a hyperbolic avoided crossing. Avoided crossings happen commonly in quantum mechanics, and here we have a simple model for this phenomenon.



In the limit of large detuning, when the coupling is small, we can approximately identify the coupled eigenstates with the uncoupled eigenstates. Near resonance, however, the states are mixed, and the energies are shifted. The energy shifts are the **ac Stark shifts**, or **lamp shifts**, which we will treat in more detail later.

5.3.1 Rabi Oscillations in the Dressed-State Picture

We saw that due to the interaction of the field, the atom undergoes Rabi oscillations, where the magnitudes of the probability amplitudes oscillate in time. How does this work in the dressed-state picture, where the only possible evolution of the amplitudes is a change in *phase*? As an illustration, we'll work this out explicitly for the resonant case $\Delta = 0$. In this case, the dressed states are even and odd superpositions of the bare states,

$$|\pm\rangle = |\mathrm{g}\rangle \pm |\mathrm{e}\rangle \tag{5.67}$$

(we are dropping the normalization factor $1/\sqrt{2}$ here), while the bare states are

$$\begin{aligned} |\mathrm{g}\rangle &= |+\rangle + |-\rangle \\ |\mathrm{e}\rangle &= |+\rangle - |-\rangle \end{aligned} \tag{5.68}$$

in terms of the dressed states. Then if the atom starts in the ground state $|\mathrm{g}\rangle = |+\rangle + |-\rangle$, the dressed-state phases will evolve according to the appropriate energies:

$$\begin{aligned} |\psi(t)\rangle &= |+\rangle\, e^{-iE_+ t/\hbar} + |-\rangle\, e^{-iE_- t/\hbar} \\ &= |+\rangle\, e^{-i\Omega t/2} + |-\rangle\, e^{i\Omega t/2}. \end{aligned} \tag{5.69}$$

Dropping an irrelevant overall phase,

$$|\psi(t)\rangle = |+\rangle + |-\rangle\, e^{i\Omega t}. \tag{5.70}$$

Thus, it is the *relative phase* of the two dressed states that changes in time, and causes the Rabi flopping. The frequency splitting between the dressed states is just Ω, so the evolution must be periodic with frequency Ω. At time $\Omega t = \pi$,

$$|\psi(t = \pi/\Omega)\rangle = |+\rangle + |-\rangle\, e^{i\pi} = |+\rangle - |-\rangle = |\mathrm{e}\rangle, \tag{5.71}$$

and the atom is in the excited state. At time $\Omega t = 2\pi$,

$$|\psi(t = 2\pi/\Omega)\rangle = |+\rangle + |-\rangle\, e^{i2\pi} = |+\rangle + |-\rangle = |\mathrm{g}\rangle, \tag{5.72}$$

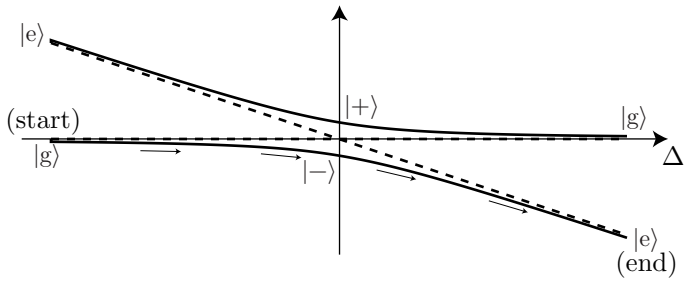
and the atom is back in the ground state. This is precisely what we found in our analysis above.

Of course, the same general picture holds off resonance. In this case, the dressed-state splitting becomes $\tilde{\Omega}$, so the oscillations occur at the generalized Rabi frequency. Also, from Eq. (5.65) the bare states are not equal superpositions of the dressed states, so the off-resonant Rabi flopping will not result in a complete transfer of population, say from $|\mathrm{g}\rangle$ to $|\mathrm{e}\rangle$. This intuitive picture gives a simple way of looking at the analytic, off-resonant solutions we found above.

5.3.2 Adiabatic Passage and Landau–Zener Crossings

Suppose you have an atom in the ground state, and you want to tranfer it exactly into the excited state. Sure, you can just use a resonant π-pulse. But lasers aren't perfect: they have jitter in both frequency and amplitude, the amplitude jitter often being more important directly on resonance. Instead of working really hard to stabilize the laser, there is an easy trick that you can use, provided you can *chirp* (sweep) the laser frequency. The idea is to chirp the laser frequency across the atomic resonance. As long as the chirp is slow enough, it turns out that the atom will flip to the excited state with nearly unit probability. This effect is called **adiabatic passage**.

We can see how this works as follows. Suppose we start the chirp well below resonance, $\Delta \ll -\Omega$ (assuming $\Omega > 0$ by convention). From Eqs. (5.65) and (5.66), we can see that the phase angle $\theta \approx 0$, and thus $|-\rangle \approx |\mathrm{g}\rangle$. As the chirp slowly proceeds, the adiabatic theorem guarantees that the atom will remain in the $|-\rangle$ dressed state. When the chirp ends well above resonance, $\Delta \gg \Omega$, we can see that the phase angle increases to $\theta \approx \pi/2$ (passing through $\theta = \pi/4$ on resonance), and thus $|-\rangle \approx |\mathrm{e}\rangle$.



Thus, by mixing the two states, the atom adiabatically passes from $|\mathrm{g}\rangle$ to $|\mathrm{e}\rangle$: the avoided crossing *exchanges the identities of the uncoupled eigenstates*. Of course, we could have chirped the other way, in which case the atom would have followed the $|+\rangle$ dressed state from $|\mathrm{g}\rangle$ to $|\mathrm{e}\rangle$.

Clearly this happens if the chirp is "slow enough." For a very fast chirp, the atom–field coupling has effectively no time to become manifest, and the atom starting in $|\mathrm{g}\rangle$ will jump ("tunnel") across the avoided crossing as if it weren't there. For intermediate chirps, we might expect the atom to end up in a superposition of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, since some of the probablity will tunnel through the gap. but how slow is slow enough? This problem can be solved completely, and the solution was given independently by Landau, Zener, and Stückelberg; this is commonly called the **Landau–Zener crossing** problem.[4]

To solve the problem, recall the atom–field interaction Hamiltonian from Eq. (5.18) in the nonrotating frame:

$$H_{\mathrm{AF}} = \frac{\hbar\Omega}{2}\left(\sigma e^{i\omega t} + \sigma^\dagger e^{-i\omega t}\right). \tag{5.73}$$

This came from a monochromatic field $E^{(+)}$ with phase $e^{-i\omega t}$. To generalize this to a chirped field, we note that the frequency is the *rate at which phase accumulates*, so that in general we can write the phase as $e^{-i\phi(t)}$, with $\omega \equiv d\phi(t)/dt$. Thus, the generalized interaction Hamiltonian has the same form as above, but with the replacement

$$e^{-i\omega t} \longrightarrow \exp\left(-i\int_0^t \omega(t')\,dt'\right), \tag{5.74}$$

where $\omega(t)$ is the instantaneous frequency. We carry through the derivation of the Schrödinger equation as before, defining the slowly varying excited-state amplitudes by

$$\tilde{c}_{\mathrm{e}} = c_{\mathrm{e}} \exp\left(i\int_0^t \omega(t')\,dt'\right) \tag{5.75}$$

[4] L. D. Landau, "Zur Theorie der Energieübertragung. II.," *Physikalische Zeitschrift der Sowjetunion* **2**, 46 (1932); Clarence Zener, "Non-Adiabatic Crossing of Energy Levels," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, **137**, 696 (1932); E. C. G. Stueckelberg, *Helvetica Physica Acta* **5**, 369 (1932). See also Jan R. Rubbmark, Michael M. Kash, Michael G. Littman, and Daniel Kleppner, "Dynamical effects at avoided level crossings: A study of the Landau–Zener effect using Rydberg atoms," *Physical Review A* **23**, 3107 (1981) (doi: 10.1103/PhysRevA.23.3107).

so that the rotating-frame atomic and atom–field Hamiltonians become

$$\begin{aligned} \tilde{H}_{\mathrm{A}} &= -\hbar\Delta(t)\sigma^\dagger\sigma \\ \tilde{H}_{\mathrm{AF}} &= \frac{\hbar\Omega}{2}\left(\sigma + \sigma^\dagger\right), \end{aligned} \tag{5.76}$$

where $\Delta(t) := \omega(t) - \omega_0$. The equations of motion for the bare-state amplitudes are thus

$$\begin{aligned} \partial_t c_{\mathrm{g}} &= -i\frac{\Omega}{2}\tilde{c}_{\mathrm{e}} \\ \partial_t \tilde{c}_{\mathrm{e}} &= i\Delta(t)\tilde{c}_{\mathrm{e}} - i\frac{\Omega}{2}c_{\mathrm{g}}. \end{aligned} \tag{5.77}$$

Decoupling by differentiation and back substitution as usual gives

$$\begin{aligned} \left(\partial_t^2 - i\Delta\partial_t - i(\partial_t\Delta) + \frac{\Omega^2}{4}\right)\tilde{c}_{\mathrm{e}} &= 0 \\ \left(\partial_t^2 - i\Delta\partial_t + \frac{\Omega^2}{4}\right)c_{\mathrm{g}} &= 0. \end{aligned} \tag{5.78}$$

Now we transform into yet another rotating frame defined by the new variables

$$\begin{aligned} c'_{\mathrm{e}} &:= \tilde{c}_{\mathrm{e}} \exp\left(-\frac{i}{2}\int_0^t \Delta(t')\,dt'\right) \\ c'_{\mathrm{g}} &:= c_{\mathrm{g}} \exp\left(-\frac{i}{2}\int_0^t \Delta(t')\,dt'\right), \end{aligned} \tag{5.79}$$

in terms of which the equations (5.78) become

$$\begin{aligned} \left(\partial_t^2 - \frac{i}{2}(\partial_t\Delta) + \frac{\Delta^2}{4} + \frac{\Omega^2}{4}\right)c'_{\mathrm{e}} &= 0 \\ \left(\partial_t^2 + \frac{i}{2}(\partial_t\Delta) + \frac{\Delta^2}{4} + \frac{\Omega^2}{4}\right)c'_{\mathrm{g}} &= 0. \end{aligned} \tag{5.80}$$

Assuming a purely linear chirp, $\Delta = \alpha t$ ($\alpha > 0$), we see that

$$\begin{aligned} \left(\partial_t^2 + \frac{\Omega^2}{4} - \frac{i\alpha}{2} + \frac{\alpha^2}{4}t^2\right)c'_{\mathrm{e}} &= 0 \\ \left(\partial_t^2 + \frac{\Omega^2}{4} + \frac{i\alpha}{2} + \frac{\alpha^2}{4}t^2\right)c'_{\mathrm{g}} &= 0, \end{aligned} \tag{5.81}$$

and then defining the new variables

$$\nu := i\frac{\Omega^2}{4\alpha} = i|\nu|, \qquad z := e^{-i\pi/4}\sqrt{\alpha}\,t, \tag{5.82}$$

the uncoupled equations of motion become

$$\begin{aligned} \left(\partial_z^2 + \nu + \frac{1}{2} - \frac{z^2}{4}\right)c'_{\mathrm{e}} &= 0 \\ \left(\partial_z^2 + (\nu - 1) + \frac{1}{2} - \frac{z^2}{4}\right)c'_{\mathrm{g}} &= 0. \end{aligned} \tag{5.83}$$

These are both in the form of **Weber's equation**,

$$\left(\partial_z^2 + \nu + \frac{1}{2} - \frac{z^2}{4}\right)y = 0, \tag{5.84}$$

whose solutions are the **parabolic-cylinder functions** $D_\nu(z)$, $D_\nu(-z)$, $D_{-(\nu+1)}(iz)$, and $D_{-(\nu+1)}(-iz)$.[5]

We will now be concerned with the solution for the excited-state amplitude $c'_{\rm e}$. The leading-order asymptotic expansion of $D_n(z)$ (for large $|z|$) is

$$D_\nu(z) \sim e^{-z^2/4} z^\nu \tag{5.85}$$

for $|\arg(z)| < 3\pi/4$. Letting $\nu \longrightarrow -(\nu+1)$ and $z \longrightarrow i|z|e^{-i\pi/4}$,

$$D_{-(\nu+1)}\left(i|z|e^{-i\pi/4}\right) \sim e^{-i\pi(\nu+1)/4}\, e^{-i|z|^2/4}\, |z|^{-(\nu+1)}. \tag{5.86}$$

We can conclude from this that $D_{-(\nu+1)}(-iz)$ is a function that vanishes for large $|z|$, if z follows the direction $e^{i3\pi/4} = -e^{-i\pi/4}$ (vanishing as $|z|^{-1}$ since ν is imaginary), and thus vanishes in the limit $t \longrightarrow -\infty$. Thus, we can write our solution for the excited-state amplitude

$$c'_{\rm e} = A D_{-(\nu+1)}(-iz), \tag{5.87}$$

where A is an undetermined constant, since this represents a particular solution to Weber's equation with the correct initial condition

$$c'_{\rm e}(t \longrightarrow -\infty) = 0. \tag{5.88}$$

Note that we are idealizing the chirp, assuming that it extends to all frequencies, and thus we treating this as a scattering-type problem, where the boundary conditions are applied at $t = \pm\infty$.

Now to determine the coefficient. We use the bare-state equations of motion (5.77), which in terms of new variables becomes

$$c'_{\rm g} = \frac{i}{\sqrt{\nu}}\left(\partial_z + \frac{z}{2}\right) c'_{\rm e}. \tag{5.89}$$

We can then use the asymptotic expansion (5.86) to find $|c'_{\rm g}|$ in the limit $t \longrightarrow -\infty$:

$$\left|c'_{\rm g}\right| = \frac{|A|}{\sqrt{|\nu|}} e^{\pi|\nu|/4}. \tag{5.90}$$

Our other boundary condition is that the ground state is initially fully populated,

$$\left|c'_{\rm g}(t \longrightarrow -\infty)\right| = 1, \tag{5.91}$$

and thus we have

$$|A| = \sqrt{|\nu|} e^{-\pi|\nu|/4}, \tag{5.92}$$

which fixes the form of the solution, up to an overall phase.

To look at the $t \longrightarrow \infty$ limit, we use an alternate form of the asymptotic expression,

$$D_\nu(z) \sim e^{-z^2/4} z^\nu - \frac{\sqrt{2\pi}}{\Gamma(-\nu)} e^{-i\pi\nu} e^{z^2/4} z^{-(\nu+1)}, \tag{5.93}$$

valid for $\arg z \in (-5\pi/4, -\pi/4)$. Letting $\nu \longrightarrow -(\nu+1)$ and $z \longrightarrow -i|z|e^{-i\pi/4}$, and keeping only the leading-order term, we find

$$D_{-(\nu+1)}\left(-i|z|e^{i\pi/4}\right) \sim \frac{\sqrt{2\pi}}{\Gamma(\nu+1)} e^{i\pi\nu/4} e^{i|z|^2/4} |z|^\nu, \tag{5.94}$$

in which case the excited-state population becomes

$$\begin{aligned} P_{\rm e}(t \longrightarrow \infty) &= |c'_{\rm e}|^2 = |A|^2 \left| D_{-(\nu+1)}\left(-i|z|e^{-i\pi/4}\right)\right|^2 \\ &= \frac{2\pi|\nu|e^{-\pi|\nu|}}{\Gamma(\nu+1)\Gamma(-\nu+1)} \\ &= 2e^{-\pi|\nu|}\sinh(\pi|\nu|) \\ &= 1 - e^{-2\pi|\nu|} \end{aligned} \tag{5.95}$$

[5]The theory of Weber's equation and the parabolic-cylinder functions that we will use here is covered by E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th reprinted ed. (Cambridge, 1945), section 16.5.

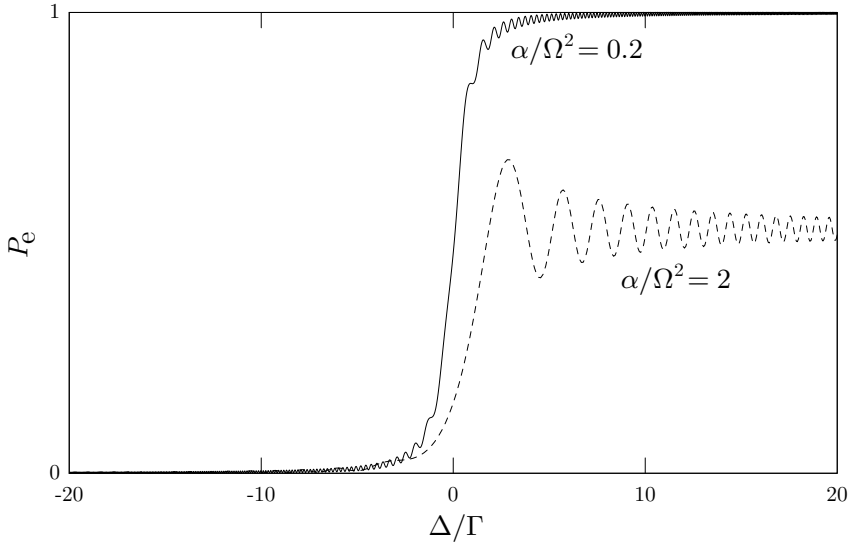
in the limit $t \longrightarrow \infty$. Thus, the fraction that *didn't* adiabatically follow the dressed state is

$$P_{\rm lost} = e^{-2\pi|\nu|} = \exp\left(-\frac{\pi\Omega^2}{2|\alpha|}\right) = \exp\left(-\frac{\pi\Omega^2}{2|\partial_t\Delta|}\right).$$

(Landau–Zener tunnel probability) (5.96)

This is the Landau–Zener result. In the limit of a strong field and slow sweep ($\Omega^2 \gg |\partial_t\Delta|$), nothing is lost—all the population adiabatically follows the dressed state and makes the transition. In the opposite regime of a weak field and fast sweep ($\Omega^2 \ll |\partial_t\Delta|$), the population jumps (tunnels) across the gap, and everything stays in the original state.

The simulated evolution for an atom undergoing a chirp from $\Delta/\Omega = -20$ to 20 for two different chirp rates, slow ($\alpha/\Omega^2 = 0.2$) and fast ($\alpha/\Omega^2 = 2$), are shown here.



The Landau–Zener predictions for the final excited-state population are 0.9996 and 0.544, respectively.

What we have discussed here is just adiabatic passage. We also need to consider the effects of the fact that the excited state will decay due to spontaneous emission. Thus, we also need adiabatic *rapid* passage,[6] since the chirp must be fast compared to decay rate of excited state for the analysis to be valid. The field must also be strong enough so that the dressed states are well resolved, in spite of homogeneous broadening of the levels. Thus, if the Rabi frequency is much larger than the decay rate, and the field is chirped quickly enough, the atom can still make a complete transition.

5.4 Bloch Sphere

Consider the Pauli operators

$$\begin{aligned}
\sigma_x &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \sigma + \sigma^\dagger \\
\sigma_y &= \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} = i\left(\sigma - \sigma^\dagger\right) \\
\sigma_z &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = |{\rm e}\rangle\langle{\rm e}| - |{\rm g}\rangle\langle{\rm g}| = \sigma^\dagger\sigma - \sigma\sigma^\dagger = [\sigma^\dagger, \sigma],
\end{aligned} \tag{5.97}$$

[6] Adiabatic rapid passage was first observed in NH_3 by sweeping the resonant frequency via the dc Stark effect (shifting the resonant frequency by applying an external dc electric field). See Michael M. T. Loy, "Observation of Population Inversion by Optical Adiabatic Rapid Passage," *Physical Review Letters* **32**, 814 (1974) (doi: 10.1103/PhysRevLett.32.814).

which satisfy the commutation and anticommutation relations

$$\begin{aligned}[\sigma_\alpha, \sigma_\beta] &= 2i\varepsilon_{\alpha\beta\gamma}\sigma_\gamma \\ [\sigma_\alpha, \sigma_\beta]_+ &= 2\delta_{\alpha\beta}.\end{aligned} \tag{5.98}$$

These operators work on (rotating-frame) states with the ordering

$$\begin{bmatrix} \tilde{c}_\mathrm{e} \\ c_\mathrm{g} \end{bmatrix}. \tag{5.99}$$

The idea behind the **Bloch sphere**[7] is to use the expectation values $\langle\sigma_\alpha\rangle$ as dynamical coordinates for the atomic evolution.

Let's first connect these variables to the density matrix for the atom. It is sufficient to use the relations (in the rotating frame)

$$\begin{aligned}\langle\sigma\rangle &= \mathrm{Tr}[|\mathrm{g}\rangle\langle\mathrm{e}|\rho] = \mathrm{Tr}[\langle\mathrm{e}|\tilde{\rho}|\mathrm{g}\rangle] = \tilde{\rho}_\mathrm{eg} \\ \left\langle\sigma^\dagger\right\rangle &= \mathrm{Tr}[|\mathrm{e}\rangle\langle\mathrm{g}|\rho] = \mathrm{Tr}[\langle\mathrm{g}|\tilde{\rho}|\mathrm{e}\rangle] = \tilde{\rho}_\mathrm{ge} \\ \left\langle\sigma^\dagger\sigma\right\rangle &= \mathrm{Tr}[|\mathrm{e}\rangle\langle\mathrm{e}|\tilde{\rho}] = \rho_\mathrm{ee} \\ \left\langle\sigma\sigma^\dagger\right\rangle &= \mathrm{Tr}[|\mathrm{g}\rangle\langle\mathrm{g}|\tilde{\rho}] = \rho_\mathrm{gg}.\end{aligned} \tag{5.100}$$

Here the twiddles indicate coherences in the rotating frame. To make this more explicit, $\tilde{\rho}$ is the slowly varying state, which in a pure state has the form $|\tilde{\psi}\rangle\langle\tilde{\psi}|$. The corresponding density matrix for a pure state then has the form $\tilde{\rho}_{\alpha\beta} = \tilde{c}_\alpha\tilde{c}^*_\beta$ (with $\tilde{c}_\mathrm{g} \equiv c_\mathrm{g}$). Then the rotating-frame populations are independent of the choice of frame,

$$\begin{aligned}\tilde{\rho}_\mathrm{gg} &= c_\mathrm{g}c^*_\mathrm{g} = \rho_\mathrm{gg} \\ \tilde{\rho}_\mathrm{ee} &= \tilde{c}_\mathrm{e}\tilde{c}^*_\mathrm{e} = c_\mathrm{e}c^*_\mathrm{e} = \rho_\mathrm{ee},\end{aligned} \tag{5.101}$$

while the coherences in the rotating frame differ from the usual coherences by a rotating phase factor,

$$\begin{aligned}\tilde{\rho}_\mathrm{ge} &= c_\mathrm{g}\tilde{c}^*_\mathrm{e} = c_\mathrm{g}c_\mathrm{e}e^{-i\omega t} = \rho_\mathrm{ge}e^{-i\omega t} \\ \tilde{\rho}_\mathrm{eg} &= \rho_\mathrm{eg}e^{i\omega t},\end{aligned} \tag{5.102}$$

and thus the rotating-frame coherences are called the **slowly varying coherences**.

Now using the evolution equations for the coefficients $\tilde{c}_\mathrm{e}$ and c_g [Eqs. (5.23)], we can compute the equations of motion for the excited-state population,

$$\begin{aligned}\partial_t\rho_\mathrm{ee} &= \tilde{c}^*_\mathrm{e}(\partial_t\tilde{c}_\mathrm{e}) + \mathrm{c.c.} \\ &= i\Delta\tilde{c}_\mathrm{e}\tilde{c}^*_\mathrm{e} - i\frac{\Omega}{2}c_\mathrm{g}\tilde{c}^*_\mathrm{e} + \mathrm{c.c.} \\ &= i\frac{\Omega}{2}(\tilde{\rho}_\mathrm{eg} - \tilde{\rho}_\mathrm{ge}),\end{aligned} \tag{5.103}$$

the ground-state population,

$$\begin{aligned}\partial_t\rho_\mathrm{gg} &= -\partial_t\rho_\mathrm{ee} \\ &= -i\frac{\Omega}{2}(\tilde{\rho}_\mathrm{eg} - \tilde{\rho}_\mathrm{ge}),\end{aligned} \tag{5.104}$$

[7] after Felix Bloch, who derived the equation of motion for a spin in a magnetic field, which has the spherical representation. See F. Bloch, "Nuclear Induction," *Physical Review* **70**, 460 (1946) (doi: 10.1103/PhysRev.70.460).

and the coherences,

$$\begin{aligned}\partial_t\tilde{\rho}_{\rm ge} &= c_{\rm g}\partial_t\tilde{c}_{\rm e}^* + \tilde{c}_{\rm e}^*\partial_t c_{\rm g}\\ &= -i\Delta c_{\rm g}\tilde{c}_{\rm e}^* + i\frac{\Omega}{2}c_{\rm g}c_{\rm g}^* - i\frac{\Omega}{2}\tilde{c}_{\rm e}\tilde{c}_{\rm e}^*\\ &= -i\Delta\tilde{\rho}_{\rm ge} - i\frac{\Omega}{2}(\rho_{\rm ee}-\rho_{\rm gg}),\\ \partial_t\tilde{\rho}_{\rm eg} &= \partial_t\tilde{\rho}_{\rm ge}^*\\ &= i\Delta\tilde{\rho}_{\rm eg} + i\frac{\Omega}{2}(\rho_{\rm ee}-\rho_{\rm gg}).\end{aligned} \tag{5.105}$$

Of course, these four equations for the density matrix elements are equivalent to the Schrödinger–von Neumann equation

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A}+\tilde{H}_{\rm AF},\tilde{\rho}\right] \tag{5.106}$$

in the rotating frame. Again, without the transformation to the rotating frame, the equations of motion would have explicit time dependences, representing the relative precession of the atomic and field phases.

Given the above relations for the density matrix elements, how many degrees of freedom are there? There are four matix elements, and if each is complex, then there are eight possible independent, real numbers. The populations must be real, so this removes two free variables. The populations much further sum up to unity, removing another free variable. Finally, the constraint $\rho_{\rm ge}=\rho_{\rm eg}^*$ removes two more free variables, leaving only three independent, real numbers to represent the quantum state. This motivates the idea of using a three-vector (in $\mathbb{R}^3$) to represent the atomic state.

To proceed with this idea, we start with the relations

$$\begin{aligned}\langle\sigma_x\rangle &= \langle\sigma\rangle + \langle\sigma^\dagger\rangle = \tilde{\rho}_{\rm eg}+\tilde{\rho}_{\rm ge}\\ \langle\sigma_y\rangle &= i\langle\sigma\rangle - i\langle\sigma^\dagger\rangle = i(\tilde{\rho}_{\rm eg}-\tilde{\rho}_{\rm ge})\\ \langle\sigma_z\rangle &= \rho_{\rm ee}-\rho_{\rm gg},\end{aligned} \tag{5.107}$$

for the Bloch variables, and then we use the equations of motion for the density matrix elements (5.103)-(5.105) to write

$$\begin{aligned}\partial_t\langle\sigma_x\rangle &= \Delta\langle\sigma_y\rangle\\ \partial_t\langle\sigma_y\rangle &= -\Delta\langle\sigma_x\rangle - \Omega\langle\sigma_z\rangle\\ \partial_t\langle\sigma_z\rangle &= \Omega\langle\sigma_y\rangle\,.\end{aligned} \tag{5.108}$$

(Bloch-vector equations of motion)

Note that these equations may be rewritten in terms of the **Bloch vector** $\langle\boldsymbol{\sigma}\rangle := \langle\sigma_x\rangle\,\hat{x}+\langle\sigma_y\rangle\,\hat{y}+\langle\sigma_z\rangle\,\hat{z}$ as

$$\partial_t\langle\boldsymbol{\sigma}\rangle = -\Delta\hat{z}\times\langle\boldsymbol{\sigma}\rangle + \Omega\hat{x}\times\langle\boldsymbol{\sigma}\rangle\,, \tag{5.109}$$

(Bloch-vector equation of motion)

which we can also write this as a torque equation in terms of a single "precession vector" $\boldsymbol{\wp}$ as[8]

$$\partial_t\langle\boldsymbol{\sigma}\rangle = \boldsymbol{\wp}\times\langle\boldsymbol{\sigma}\rangle\,, \tag{5.110}$$

(Bloch-vector equation of motion)

where

$$\boldsymbol{\wp} := \Omega\hat{x} - \Delta\hat{z}, \tag{5.111}$$

(precession vector)

in analogy with $\boldsymbol{\tau}=\partial_t\mathbf{L}=\boldsymbol{\Omega}\times\mathbf{L}$, with $\mathbf{L}$ the angular momentum, $\boldsymbol{\tau}$ the torque, and $\boldsymbol{\Omega}$ the angular frequency vector for the precession (or the magnetic field vector in the case of **Larmor precession** of

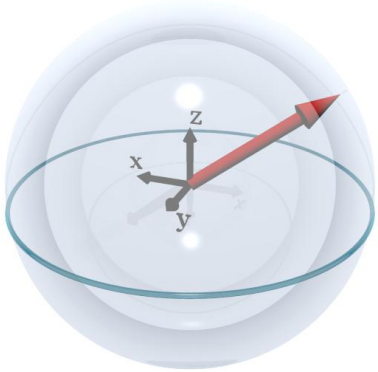
[8]The symbol $\wp$ is pronounced "squiggle."

a magnetic moment).[9] This picture of the two-level atom in terms of a precessing spin is known as the **Feynman–Vernon–Hellwarth representation**.[10]

One property that follows immediately from this representation is that the length of the Bloch vector is a constant of the motion. That's because the change in $\langle \boldsymbol{\sigma} \rangle$ is always normal to it. Furthermore, if we assume a pure quantum state, we can see that

$$\begin{aligned} \left| \langle \boldsymbol{\sigma} \rangle \right|^2 &= \langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2 \\ &= (\tilde{\rho}_{\mathrm{eg}} + \tilde{\rho}_{\mathrm{ge}})^2 - (\tilde{\rho}_{\mathrm{eg}} - \tilde{\rho}_{\mathrm{ge}})^2 + (\rho_{\mathrm{ee}} - \rho_{\mathrm{gg}})^2 \\ &= 4\tilde{\rho}_{\mathrm{eg}}\tilde{\rho}_{\mathrm{ge}} + \rho_{\mathrm{ee}}^2 + \rho_{\mathrm{gg}}^2 - 2\rho_{\mathrm{ee}}\rho_{\mathrm{gg}} \\ &= (\rho_{\mathrm{ee}} + \rho_{\mathrm{gg}})^2 = 1, \end{aligned} \tag{5.112}$$

since $\tilde{\rho}_{\mathrm{eg}}\tilde{\rho}_{\mathrm{ge}} = \rho_{\mathrm{ee}}\rho_{\mathrm{gg}}$ for a pure state. Thus, the Bloch vector for a pure state lies on a sphere of unit radius, which we call the **Bloch sphere**. Thus, of the three independent real numbers in the density matrix, one of these is fixed by the purity of the state. The remaining two numbers, as we will see, correspond to the degree of atomic excitation and a phase angle.

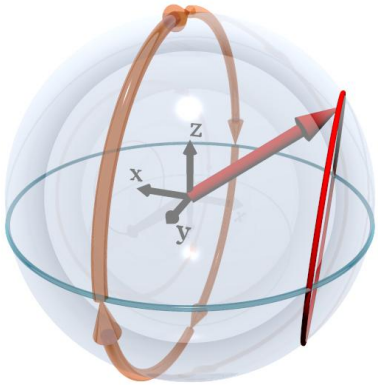


We can see that the FVH representation is a compact, handy way to visualize the evolution. The trajectories lie on the unit sphere, and their evolution is generated simply by a constant-speed rotation of the whole sphere. The rotation axis and angular speed are determined by the precession vector $\boldsymbol{\wp}$, whose magnitude is simply $\tilde{\Omega}$.

From Eqs. (5.107), we can interpret the meanings of the components of the Bloch vector. The vertical (z) component of the Bloch vector represents the degree of atomic excitation (population inversion): the Bloch vector points straight down for an atom in $|\mathrm{g}\rangle$ and straight up for an atom in $|\mathrm{e}\rangle$. When the driving field is on resonance ($\Delta = 0$), the precession vector is $\boldsymbol{\wp} = \Omega\hat{x}$.

[9] See Richard Feynman, Robert B. Leighton, and Matthew L. Sands, *The Feynman Lectures in Physics* (Addison–Wesley, 1963), Chapter 20.

[10] Richard P. Feynman, Frank L. Vernon, and Robert W. Hellwarth, "Geometrical Representation of the Schrödinger Equation for Solving Maser Problems," *Journal of Applied Physics* **28**, 49 (1957) (doi: 10.1063/1.1722572).



If the atom is initially in the ground state, the trajectory follows the great circle given by the intersection of the Bloch sphere with the $x = 0$ plane. Thus, it passes through the excited state.

Now we turn to the transverse (x and y) components of the Bloch vector. Recalling that the dipole operator has the form

$$\mathbf{d} = \langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle(\sigma + \sigma^\dagger), \tag{5.113}$$

we can easily see that

$$\langle \mathbf{d} \rangle = \langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle \langle \sigma_x \rangle \,. \tag{5.114}$$

Thus, $\langle \sigma_x \rangle$ represents the atomic dipole. The other transverse component $\langle \sigma_y \rangle$, represents the alternate "quadrature" of the dipole moment. That is, $\langle \sigma_x \rangle$ represents the real part of $\left\langle \mathbf{d}^{(+)} \right\rangle$,

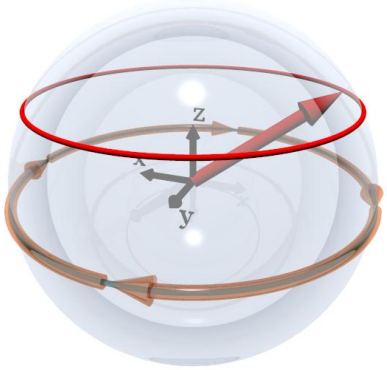
$$\langle \sigma_x \rangle \sim \mathrm{Re}\left[\langle \sigma \rangle\right] \sim \mathrm{Re}\left[\left\langle \mathbf{d}^{(+)} \right\rangle\right], \tag{5.115}$$

while $\langle \sigma_y \rangle$ represents the imaginary part of $\left\langle \mathbf{d}^{(+)} \right\rangle$,

$$\langle \sigma_y \rangle \sim -\mathrm{Im}\left[\langle \sigma \rangle\right] \sim -\mathrm{Im}\left[\left\langle \mathbf{d}^{(+)} \right\rangle\right]. \tag{5.116}$$

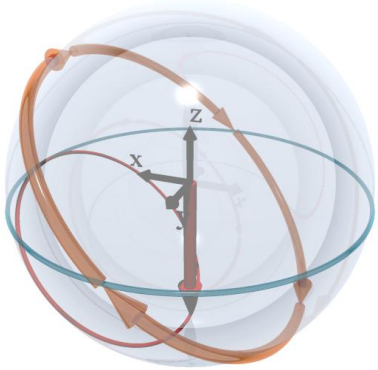
Note that there can only be a dipole moment when the atom is in a *superposition* of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, since the diagonal matrix elements of $\mathbf{d}$ vanish.

The other picture of the transverse Bloch-vector components is as follows. When the atom is in a superposition of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, the azimuthal angle (represented by the x and y components of the Bloch vector) represents the *relative phase* of the ground and excited states. In the absence of an external field, $\Omega = 0$. Then the precession vector is $\boldsymbol{\wp} = -\Delta\hat{z}$.



Thus, the Bloch sphere simply spins about the z-axis, causing azimuthal rotation of the trajectories. This evolution represents the relative phase evolution of the ground and excited states. But remember that we are in the *rotating* frame, where the rotation rate is the field frequency ω. To go back to the original variables, you just have to add ω to the precession frequency to get $\omega_0 \hat{z}$ in the stationary coordinates. We can therefore see that the free evolution is just the precession of the excited state phase relative to that of the ground state.

With a nearly resonant driving field with nonzero detuning, the rotation axis is tilted, being a combination of the previous two rotations. If the atom starts in the ground state, the trajectory never quite makes it exactly to the excited state.



This a nice way to visualize how the off-resonant excitation ends up being incomplete. Furthermore, the rate at which the Rabi oscillations occur is given by the magnitude

$$|\boldsymbol{\wp}| = \sqrt{\Omega^2 + \Delta^2} = \tilde{\Omega}, \tag{5.117}$$

as we saw from the direct solution to the Schrödinger equation.

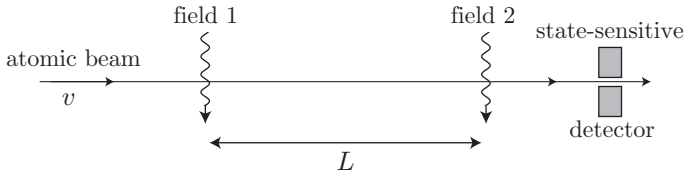
5.4.1 Atomic Timekeeping and Ramsey Fringes

One nice application of the Bloch sphere is to understanding **Ramsey fringes**,[11] which form the basis for atomic time and frequency standards as well as the benchmark for demonstrating quantum coherence.

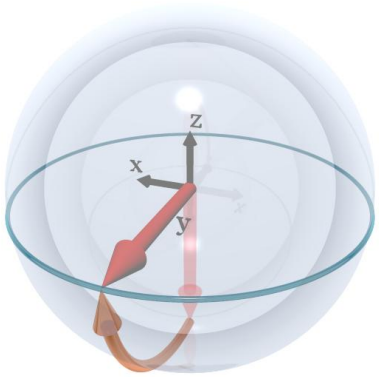
Suppose that we want to let a field interact with an atom in order to compare their frequencies. The main limitation in doing this is the interaction time, since long interaction times are required to resolve small frequency splittings (as dictated by the "time-frequency uncertainty relation"). In principle, one can let the atom and field interact for a long time, but this poses a number of difficult problems. For example, for configurations such as the atomic beam, it is difficult to maintain a uniform interaction over the entire length of the beam, since the interaction region must be large (say, meters) for a sensitive measurment. Furthermore, constraints on the apparatus itself (such as the vacuum system) may not permit the driving field to enter in certain regions. And even if it *is* possible to have a uniform field, the nearly resonant field will cause an energy (Stark) shift of the transition (as we saw from the dressed-state solutions), so that a precise comparison isn't possible anyway.

So what do we do? Well, there is a clever trick called **Ramsey's method of separated, oscillatory fields**. Suppose we have an beam of atoms moving at velocity v. We allow two identical fields (laser or microwave fields, depending on the transition, but both fields are derived from the same source) of width ℓ to cross the beam a distance L apart. Mathematically, we will idealize the fields as spatially uniform, but it is straightforward to generalize this to arbitrary beam profiles. The beams are then followed by a state-sensitive detector (Stern–Gerlach apparatus) to measure the excitation probability.

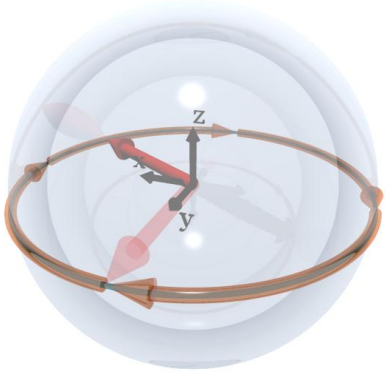
[11] Ramsey, a former student of Rabi, shared the 1989 Nobel prize for the method of separated, oscillatory fields. See Norman F. Ramsey, "A New Molecular Beam Resonance Method," *Physical Review* **76**, 996 (1949) (doi: 10.1103/PhysRev.76.996); Norman F. Ramsey, "A Molecular Beam Resonance Method with Separated Oscillating Fields," *Physical Review* **78**, 695 (1950) (doi: 10.1103/PhysRev.78.695); Norman F. Ramsey, *Molecular Beams* (Oxford, 1956).



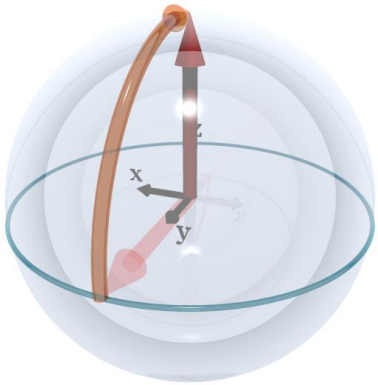
To see how this works, we will assume that the field is very close to the atomic resonance, so that $|\Delta| \ll \Omega$, where Ω is the Rabi frequency for each field. In doing so, we can ignore the fact that the Rabi oscillations do not quite occur about the x-axis. Now letting $\tau = \ell/v$ be the interaction time of each field with the passing atoms, the first field causes a Rabi oscillation with an accumulated phase of $\tilde{\Omega}\tau \approx \Omega\tau$. We assume the atoms start in the ground state, and we will choose the field amplitude Ω such that the field drives a $\pi/2$-pulse (i.e., $\Omega\tau = \pi/2$). Then the interaction with the first field puts the atom in an equal superposition of the ground and excited states.



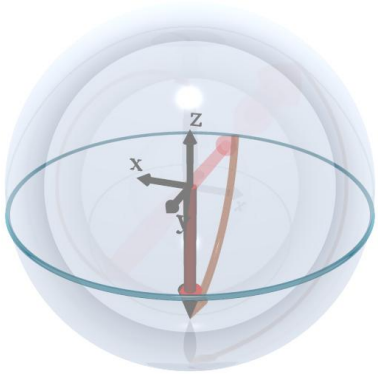
Then in between the fields, the atom undergoes free evolution—precession about the $-z$-axis at rate Δ—for a time $T = L/v$. The accumulated phase is thus $-\Delta T$.



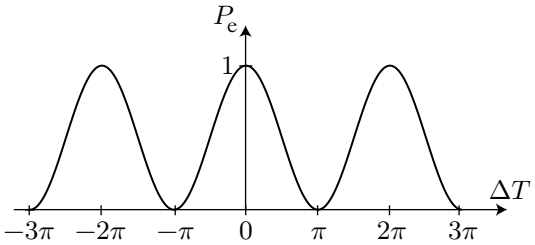
The final field causes another $\pi/2$-pulse, but its effect depends on the state of the atom after the precession stage. If the atom ends up with its initial phase after the precession, which happens if ΔT is an integer multiple of 2π, then the effect of the second $\pi/2$-pulse continues the evolution from before and promotes the atom to the excited state.



On the other hand, if the atom ends up with the *opposite* phase after precession, as happens when ΔT is an odd-integer multiple of π, then the second $\pi/2$-pulse has the opposite effect: the atom returns to the ground state.



For other final phases after the precession stage, the final excited-state population interpolates sinusoidally between these extreme values. We thus see that the output signal (the excited-state population) is sinusoidal in T with period $2\pi/\Delta$. Similarly the output signal is sinusoidal in Δ with period $2\pi/T$.



These oscillations are what are referred to as Ramsey fringes. Essentially, we have built something like an optical Mach–Zehnder interferometer, but where the two arms of the interferometer correspond to the internal states of the atom, and the beamsplitters correspond to the $\pi/2$-pulses. Alternately, we can think of this experiment as a sort of Young double slit, but where the slits are separated in *time* (and thus the fringes appear as a function of frequency). Since the output signal varies between 0 and 1, we can write

$$P_e = \cos^2\left(\frac{\Delta T}{2}\right) = \frac{1}{2}\left(1 + \cos \Delta T\right), \tag{5.118}$$

and we can see that the width (FWHM) of the central fringe is π/T in angular frequency. Thus, the accuracy of the comparison of the atom and field frequencies increases as T increases, as we expect.

In the more general case, where we are not restricted to very small detunings, the excitation probability can be written (Problem 5.6)

$$P_{\rm e} = 4\frac{\Omega^2}{\tilde{\Omega}^2}\sin^2\left(\frac{\tilde{\Omega}\tau}{2}\right)\left[\cos\left(\frac{\Delta T}{2}\right)\cos\left(\frac{\tilde{\Omega}\tau}{2}\right) - \frac{\Delta}{\tilde{\Omega}}\sin\left(\frac{\Delta T}{2}\right)\sin\left(\frac{\tilde{\Omega}\tau}{2}\right)\right]^2,$$

(Ramsey excitation proability) (5.119)

where τ is the interaction time of each of the two laser pulses of Rabi frequency Ω, and T is the time interval (time spent ''in the dark'') between pulses. For larger detunings, then, the tipped Rabi oscillations lead to a reduction in the fringe contrast. A more serious reduction in the fringe contrast for larger detunings occurs in atomic beams due to the longitudinal velocity spread, which causes slightly different interaction times and thus slightly different fringe spacings.

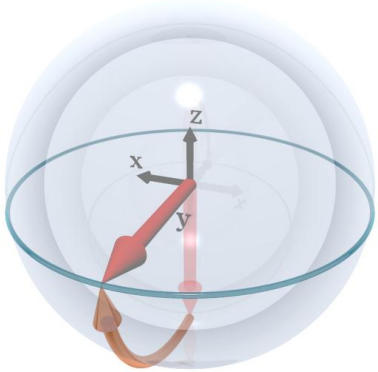
This method is precisely the one used for atomic clocks. The current frequency standard corresponds to the energy difference between two spin states of an isolated ^{133}Cs atom, which is defined to have a transition frequency of 9.192 631 770 GHz. In typical cesium frequency standards, a ''flywheel'' oscillator with very good short-term stability, such as a crystal oscillator or a hydrogen maser, has its frequency periodically compared to the cesium transition frequency by a similar atomic beam measurement. Of course, there are many tricks in getting this method to work well and to compensate for systematic effects that we won't get into here. One of the best atomic-beam clocks was NIST-7,[12] a cesium-beam atomic clock with a drift region of $L = 1.53$ m, an interaction region of $\ell = 2.3$ cm, and a mean beam velocity of 230 m/s. It operated from 1993-1999, and had an uncertainty of 5×10^{-15}. The current U.S. standard clock operated by the National Institute of Standards and Technology (NIST) is NIST-F1, a ''fountain'' clock, where a sample of ultracold atoms is tossed upwards and returns to the interaction region under free fall to reach long interaction times, and as of 2005, the uncertainty is about 5×10^{-16}.

The measure of clock stability is the ratio $\omega_0/\delta\omega$, where $\delta\omega$ is the frequency uncertainty. Traditional approaches to cesium clocks has focused on making $\delta\omega$ as small as possible. However, the ratio can also be made large by choosing a much larger transition frequency, such as in the optical regime. Candidates for future standards include optical transitions in single trapped ions, where the transition shift due the trapping fields averages to zero, or atoms trapped in ''magic wavelength'' optical lattices, which we will discuss soon.

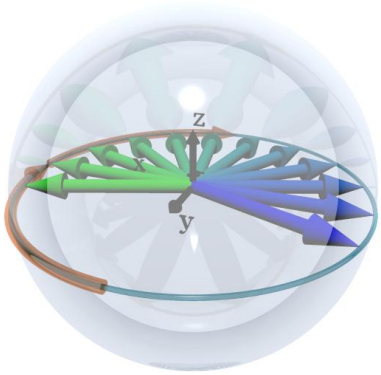
5.4.2 Spin Echoes and Photon Echoes

Ramsey-type experiments work a lot less well when **dephasing** between members of an ensemble occurs. This type of issue crops up with **inhomogeneous broadening**, where each atom effectively has a slightly different resonance frequency, as happens with Doppler broadening in atomic vapors or local field effects on atoms embedded in crystalline media (and has something of a similar effect of a velocity spread in the Ramsey experiment, which causes effectively different drift times for different atoms). To see the problem, let's walk through the Ramsey experiment in the presence of inhomogeneous broadening. The first step is the $\pi/2$-pulse to put the atoms in a superposition of the ground and excited states.

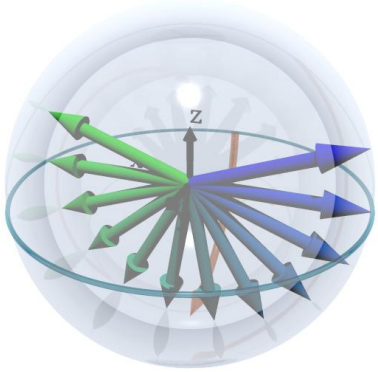
[12] J. H. Shirley, W. D. Lee, and R. E. Drullinger, ''Accuracy evaluation of the primary frequency standard NIST-7,'' *Metrologia* **38**, 427 (2001) (doi: 10.1088/0026-1394/38/5/7).



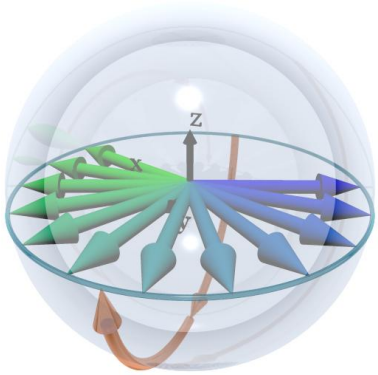
During the free-drift time, each atom precesses at a slightly different frequency, leading to a spread in phase angles that increases with time. As shown here, the blue vectors precess more quickly than the green vectors.



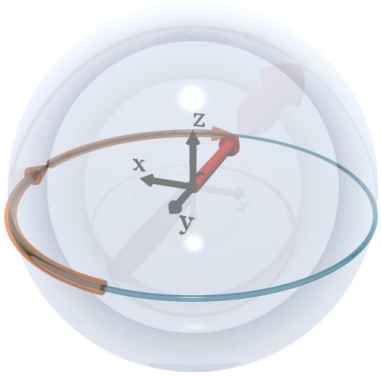
Now when the second $\pi/2$-pulse (the ''interrogation pulse'') comes, the Bloch sphere rotates appropriately. But in a situation that would put all the atoms in the ground state, only a small fraction of the atoms actually makes it to the right state. In the limit of large dephasing, the atoms are spread uniformly around the equator, and thus after the interrogation pulse, the average excited-state population is just $1/2$, independent of the drift time. The Ramsey fringes damp away in a drift time of order $1/\delta\omega_0$, where $\delta\omega_0$ measures the inhomogeneously broadened width of the atomic transition. This damping of the ensemble-averaged dipole moment due to dephasing is sometimes called **free-induction decay**.



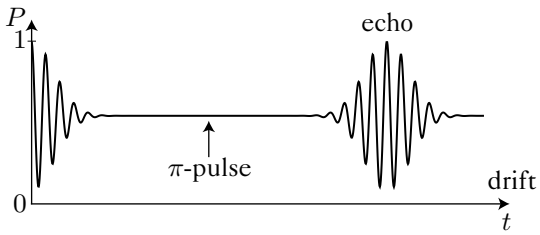
The fact that the Ramsey fringes damp away makes it *look* like irreversible decoherence of the atomic polarization. But just how reversible is it? Let's try something else—after the atoms have dephased, hit the atoms with a π-pulse. This effectively reflects the orientation of the dipoles.



Now as the evolution continues, the dipoles begin to come back together to the same phase. In the diagram, the faster (blue) dipoles are now *behind* the slower (green) ones, and thus the slower ones can now "catch up." The other way to look at this is that the reflection due to the π-pulse is effectively equivalent to flipping the precession axis, and thus reversing the direction of time. The dipoles thus come back to their common original location. Actually, that's not quite right: they come back to the *mirror image* of the original orientation, if we account fully for the effect of the reflection of the π-pulse.



When the dipoles rephase, the Ramsey fringes become visible again. If the π-pulse is applied a time T after the original preparation pulse, the spin echo occurs at time $2T$. This phenomenon in the case of nuclear spins is called the **spin echo**.[13]



In the optical case, there is a more direct, dramatic manifestation. Suppose that an atomic sample (a ruby crystal in the original experiments) is irradiated by two coherent light pulses separated by some time

[13] E. L. Hahn, "Spin Echoes," *Physical Review* **80**, 580 (1950) (doi: 10.1103/PhysRev.80.580).

T. The atoms then spontaneously emit another pulse of light a time T *after* the second pulse. We can see how this works based on the above analysis. The first laser pulse comes in and polarizes the atoms. After the dipoles dephase, the atoms can only emit incoherently, so the radiation is nondirectional and even suppressed, as we have seen is the case for classical dipoles radiating out of phase. The other way to say this is that the polarization wave decays away. However, upon applying the π-pulse, the dipoles rephase, and the polarization wave recurs. The recurrent polarization wave emits a pulse of light, an ''echo'' of the original excitation. This is called the **photon echo**.[14]

5.4.3 Adiabatic Following

We will close this discussion of the Bloch sphere by revisiting the problem of rapid adiabatic passage. Suppose a bunch of atoms begin in the ground state, so that the Bloch vector points along $-z$. Now apply an intense but very far detuned field ($|\Delta| \gg \Omega$ and thus $\tilde{\Omega} \approx |\Delta|$), and say that it is above resonance so that the precession vector $\boldsymbol{\wp} \approx |\Delta|\hat{z}$ is aligned with the Bloch vector. The precession of the Bloch vector is very simple, since it just stays in place. Now start sweeping the detuning through resonance, so that the precession vector moves through the x-axis and up towards the $+z$-axis. As long as we change the direction of $\boldsymbol{\wp}$ slowly on the time scale of the Rabi frequency Ω, the Bloch vector will follow the precession vector. When the detuning is swept until it is far *below* resonance, the precession vector has flipped by this time, and the Bloch vector has flipped as well. This is the ''classical'' view of the adiabatic passage problem that we treated quantum mechanically in Section 5.3.2.

5.5 Optical Bloch Equations

Now let's return to the evolution of the density operator. Recall that we have the Schrödinger–von Neumann equation in both the ''laboratory'' frame,

$$\partial_t \rho = -\frac{i}{\hbar}\left[H_{\rm A} + H_{\rm AF}, \rho\right], \tag{5.120}$$

and in the rotating frame,

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}, \tilde{\rho}\right]. \tag{5.121}$$

In the latter case, we have already worked out the equations of motion for the density-matrix elements in Eqs. (5.103)-(5.105):

$$\begin{aligned}
\partial_t \rho_{\rm ee} &= i\frac{\Omega}{2}(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}) \\
\partial_t \rho_{\rm gg} &= -i\frac{\Omega}{2}(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}) \\
\partial_t \tilde{\rho}_{\rm ge} &= -i\Delta\tilde{\rho}_{\rm ge} - i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}) \\
\partial_t \tilde{\rho}_{\rm eg} &= i\Delta\tilde{\rho}_{\rm eg} + i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}).
\end{aligned} \tag{5.122}$$

To model spontanous emission, we need to add extra terms. We will do so now by simply putting them in, but we will justify them later. With $\Delta = \Omega = 0$, the extra terms have the form

$$\begin{aligned}
\partial_t \rho_{\rm ee} &= -\Gamma\rho_{\rm ee} \\
\partial_t \rho_{\rm gg} &= +\Gamma\rho_{\rm ee} \\
\partial_t \tilde{\rho}_{\rm ge} &= -\gamma_\perp\tilde{\rho}_{\rm ge} \\
\partial_t \tilde{\rho}_{\rm eg} &= -\gamma_\perp\tilde{\rho}_{\rm eg}.
\end{aligned} \tag{5.123}$$

[14] Photon echoes with pulse separations of around 100 ns (with 10 ns pulses) were observed in ruby ($T_2^* \sim 0.1$ ns) by I. D. Abella, N. A. Kurnit, and S. R. Hartmann, ''Photon Echoes,'' *Physical Review* **141**, 391 (1966) (doi: 10.1103/PhysRev.141.391).

Let's look at these and understand them. The excited-state population now decays at a rate of Γ, and to compensate for this, a similar term puts the decayed population into the ground state. These terms have exactly the same form as the rate-equation terms for spontaneous emission, if we identify $\rho_{\rm ee}$ and $\rho_{\rm gg}$ as the relative number densities $N_{\rm e}/N$ and $N_{\rm g}/N$ in the excited and ground states, respectively. We thus identify $\Gamma = A_{21}$ as the excited-state decay rate. Since Γ is the rate of relaxation of the z-component of the Bloch vector to the ground state, it is also called the **longitudinal decay rate**.

The coherences also damp at the rate $\gamma_\perp$, which are introduced phenomenologically now, but which we will justify later via the quantum theory of damping. For now we note that in general $\gamma_\perp \geq \Gamma/2$, and in fact we can write

$$\gamma_\perp = \frac{\Gamma}{2} + \gamma_{\rm c}, \tag{5.124}$$

where $\gamma_{\rm c}$ models additional coherence decay *beyond* the minimum rate of $\Gamma/2$ needed for consistency with spontaneous emission. Thus $\gamma_{\rm c}$ models dephasing effects such as atom–atom collisions that do not affect the populations. Since $\gamma_\perp$ is the rate at which the coherences damp it is also the rate at which the *transverse* components (transverse to z) of the Bloch vector damp, and hence $\gamma_\perp$ is called the **transverse decay rate**. The original and still common notation[15] for these decay rates is in terms of the **longitudinal relaxation time** $T_1 = 1/\Gamma$ and **transverse relaxation time** $T_2 = 1/\gamma_\perp$. Note that the notation T_2^* is used when there is *inhomogeneous* broadening, and would include inhomogeneous dephasing as well as other sources of damping (e.g., collisions), so that $T_2^* \leq T_2$.

We can thus combine the damping terms with the Hamiltonian-evolution terms in (5.122) to obtain the **optical Bloch equations**:

$$\begin{aligned}
\partial_t \rho_{\rm ee} &= i\frac{\Omega}{2}(\tilde\rho_{\rm eg} - \tilde\rho_{\rm ge}) - \Gamma\rho_{\rm ee} \\
\partial_t \rho_{\rm gg} &= -i\frac{\Omega}{2}(\tilde\rho_{\rm eg} - \tilde\rho_{\rm ge}) + \Gamma\rho_{\rm ee} \\
\partial_t \tilde\rho_{\rm ge} &= -(\gamma_\perp + i\Delta)\tilde\rho_{\rm ge} - i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}) \\
\partial_t \tilde\rho_{\rm eg} &= -(\gamma_\perp - i\Delta)\tilde\rho_{\rm eg} + i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}).
\end{aligned} \tag{5.125}$$

(optical Bloch equations)

That is, these are the extension of Bloch's original equations for nuclear magnetic resonance to the optical regime.

Note that we may also write the damped optical Bloch equations in terms of the Bloch vector as

$$\partial_t \langle\sigma_\alpha\rangle = \epsilon_{\alpha\mu\nu}\wp_\mu\langle\sigma_\nu\rangle - \gamma_\alpha\Big(\langle\sigma_\alpha\rangle + \delta_{\alpha z}\Big),$$

(optical Bloch equations, Bloch-vector form) (5.126)

where again $\wp_\alpha = \Omega\delta_{\alpha x} - \Delta\delta_{\alpha z}$ is the precession vector, and $\gamma_\alpha = \gamma_\perp(\delta_{\alpha x} + \delta_{\alpha y}) + \Gamma\delta_{\alpha z}$ gives the damping rates for the three Bloch-vector components as we discussed above (note that there is no implied summation in the $\gamma_\alpha\langle\sigma_\alpha\rangle$ term. Writing the components out separately gives

$$\begin{aligned}
\partial_t\langle\sigma_x\rangle &= \Delta\langle\sigma_y\rangle - \gamma_\perp\langle\sigma_x\rangle \\
\partial_t\langle\sigma_y\rangle &= -\Delta\langle\sigma_x\rangle - \Omega\langle\sigma_z\rangle - \gamma_\perp\langle\sigma_y\rangle \\
\partial_t\langle\sigma_z\rangle &= \Omega\langle\sigma_y\rangle - \Gamma\Big(\langle\sigma_z\rangle + 1\Big),
\end{aligned}$$

(optical Bloch equations, Bloch-vector form) (5.127)

where we can see explicitly that the damping terms push the transverse components towards zero, while they push the longitudinal component towards the ground-state value $\langle\sigma_z\rangle = -1$.

[15] F. Bloch, *op. cit.*

5.5.1 Steady State

Since we now have damping in the equations of motion, there exist steady-state solutions ($\partial_t \tilde{\rho} = 0$) to the optical Bloch equations (5.125). To find these, we first set $\partial_t \tilde{\rho}_{\mathrm{eg}} = 0$, which gives

$$\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty) = \frac{i\Omega}{\gamma_\perp - i\Delta}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right) = -\frac{\Omega(\Delta - i\gamma_\perp)}{\gamma_\perp^2 + \Delta^2}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right). \tag{5.128}$$

The complex conjugate of this equation is

$$\tilde{\rho}_{\mathrm{ge}}(t \longrightarrow \infty) = -\frac{i\Omega}{\gamma_\perp + i\Delta}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right) = -\frac{\Omega(\Delta + i\gamma_\perp)}{\gamma_\perp^2 + \Delta^2}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right), \tag{5.129}$$

which we can subtract from the previous equation to obtain

$$(\tilde{\rho}_{\mathrm{eg}} - \tilde{\rho}_{\mathrm{ge}})(t \longrightarrow \infty) = \frac{2i\Omega\gamma_\perp}{\gamma_\perp^2 + \Delta^2}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right). \tag{5.130}$$

Now we can set $\partial_t \rho_{\mathrm{ee}} = 0$ to obtain

$$\rho_{\mathrm{ee}}(t \longrightarrow \infty) = i\frac{\Omega}{2\Gamma}\left(\tilde{\rho}_{\mathrm{eg}} - \tilde{\rho}_{\mathrm{ge}}\right) = -\frac{\Omega^2(\gamma_\perp/\Gamma)}{\gamma_\perp^2 + \Delta^2}\left(\rho_{\mathrm{ee}} - \frac{1}{2}\right). \tag{5.131}$$

Solving for ρ_{ee}, we find the steady-state excitation

$$\rho_{\mathrm{ee}}(t \longrightarrow \infty) = \frac{\Omega^2}{2\gamma_\perp\Gamma}\frac{1}{1 + \dfrac{\Delta^2}{\gamma_\perp^2} + \dfrac{\Omega^2}{\gamma_\perp\Gamma}}. \tag{5.132}$$
(steady-state excitation)

We can put this result into Eq. (5.128) to obtain the steady-state coherence

$$\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty) = -\frac{i\Omega}{2\gamma_\perp}\frac{1 + \dfrac{i\Delta}{\gamma_\perp}}{1 + \dfrac{\Delta^2}{\gamma_\perp^2} + \dfrac{\Omega^2}{\gamma_\perp\Gamma}}. \tag{5.133}$$
(steady-state coherence)

The other elements of the density matrix are of course given by $\rho_{\mathrm{ee}} + \rho_{\mathrm{gg}} = 1$ and $\tilde{\rho}_{\mathrm{ge}} = \tilde{\rho}_{\mathrm{eg}}^*$.

We can simplify the notation here somewhat by defining the **saturation parameter**

$$s := \frac{\Omega^2/\gamma_\perp\Gamma}{1 + \Delta^2/\gamma_\perp^2}. \tag{5.134}$$
(saturation parameter)

The saturation parameter is proportional to the intensity, and it has a Lorentzian frequency profile with full width at half maximum of $2\gamma_\perp$. We can then write the steady-state solutions as

$$\begin{aligned}\rho_{\mathrm{ee}}(t \longrightarrow \infty) &= \frac{s/2}{1+s} \\ |\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty)|^2 &= \frac{\Gamma}{4\gamma_\perp}\frac{s}{(1+s)^2}.\end{aligned} \tag{5.135}$$
(steady-state solutions to optical Bloch equations)

In this form it is easier to see that we get generally the same result that we got for the rate equations in Eq. (3.6): for small intensities, the excitation increases linearly with s (as $s/2$), and in the limit of large intensity ($s \longrightarrow \infty$), the largest possible excitation is *half* the population ($\rho_{\mathrm{ee}} \longrightarrow 1/2$). Furthermore, although the excitation ρ_{ee} increases monotonically with s, we can see that the expectation value of the

dipole moment, which is proportional to the real part of $\tilde{\rho}_{\mathrm{ge}}$, increases as $\sqrt{s}$ for small excitations but decreases back to zero for very large excitations. You might get the false impression from this that a highly excited atom does not radiate! This is not quite true, and we will return to this point shortly.

Most often we will be concerned with the "pure" case of homogeneous (natural) broadening, with $\gamma_\perp = \Gamma/2$. That is, there is no additional damping of the coherences due to collisions. In this case, the saturation parameter becomes

$$s := \frac{2\Omega^2/\Gamma^2}{1 + (2\Delta/\Gamma)^2},$$

(saturation parameter, homogeneous broadening) (5.136)

the steady-state population becomes

$$\rho_{\mathrm{ee}}(t \longrightarrow \infty) = \frac{s/2}{1+s} = \frac{\Omega^2/\Gamma^2}{1 + \left(\frac{2\Delta}{\Gamma}\right)^2 + 2\frac{\Omega^2}{\Gamma^2}},$$

(steady-state excitation, homogeneous broadening) (5.137)

and the steady-state coherence is

$$\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty) = -\frac{i\Omega}{\Gamma} \frac{1 + \frac{i2\Delta}{\Gamma}}{1 + \left(\frac{2\Delta}{\Gamma}\right)^2 + 2\frac{\Omega^2}{\Gamma^2}}.$$

(steady-state coherence, homogeneous broadening) (5.138)

These solutions will be important in our discussion of resonance fluorescence.

5.5.2 Damped Rabi Oscillations

5.5.2.1 Laplace Transform

Now let's consider solutions to the optical Bloch equations. Recall (from Section 4.1) that we can write the Liouville–von Neumann equation for the density operator as

$$\partial_t \tilde{\rho} = \mathcal{L}\tilde{\rho}, \tag{5.139}$$

where $\mathcal{L}$ is the Liouvillian superoperator, and effectively has larger tensor rank than the density operator. We can write this in component form as

$$\partial_t \tilde{\rho}_\alpha = \mathcal{L}_{\alpha\beta}\tilde{\rho}_\beta, \tag{5.140}$$

where α is a composite index, so that the density matrix is a column vector (i.e., for a two-level atom, α takes on the values ee, eg, ge, and gg). The Liouvillian then acts as a matrix in this notation. In the general case of a linear, time-independent equation of this form, we can obtain a solution via the Laplace transform. To review this method, start with the identity

$$y(t) = \int_0^t dt'\, \dot{y}(t') + y_0, \tag{5.141}$$

where $y_0 = y(t=0)$. Then we define the Laplace transform as

$$\begin{aligned}\mathscr{L}[y](s) &:= \int_0^\infty dt\, e^{-st} y(t) \\ &= \int_0^\infty dt\, e^{-st}\left[y_0 + \int_0^t dt'\, \dot{y}(t')\right] \\ &= \frac{y_0}{s} + \int_0^\infty dt'\, \dot{y}(t') \int_{t'}^\infty dt\, e^{-st} \\ &= \frac{y_0}{s} + \frac{1}{s}\int_0^\infty dt'\, e^{-st'}\, \dot{y}(t') \\ &= \frac{y_0}{s} + \frac{1}{s}\mathscr{L}[\dot{y}](s).\end{aligned} \tag{5.142}$$

Here, we used the identity (5.141), and the fact that the two-dimensional integral is over the $t' < t$ region, so that we can interchange the order of integration via

$$\int_0^\infty dt \int_0^t dt' = \int_0^\infty dt' \int_{t'}^\infty dt. \tag{5.143}$$

Thus, we can solve our result (5.142) for $\mathscr{L}[\dot{y}]$ to find the transform of the time derivative

$$\mathscr{L}[\dot{y}] = s\mathscr{L}[y] - y_0. \tag{5.144}$$
(Laplace transform of time derivative)

We can now use this result to take the Laplace transform of the Liouville–von Neumann equation to find

$$s\mathscr{L}[\tilde{\rho}] - \tilde{\rho}(0) = \mathcal{L}\mathscr{L}[\tilde{\rho}], \tag{5.145}$$

assuming $\mathcal{L}$ is time-independent. Thus, the Laplace transform conveniently changes a system of coupled differential equations into an algebraic problem. Now we can solve for $\mathscr{L}[\tilde{\rho}]$, with the result

$$\mathscr{L}[\tilde{\rho}] = \frac{1}{s - \mathcal{L}}\,\tilde{\rho}(0). \tag{5.146}$$

Note that the addition of a scalar and an operator here should be interpreted in the sense

$$s - \mathcal{L} \longrightarrow s\delta_{\alpha\beta} - \mathcal{L}_{\alpha\beta}. \tag{5.147}$$

The operator $(s-\mathcal{L})^{-1}$ is called the **resolvent of the Liouvillian**, and gives the decoupled form for the Laplace transform of the solutions in terms of the initial condition. The solutions are then given by inverse Laplace transforms:

$$\tilde{\rho}(t) = \mathscr{L}^{-1}\left[\frac{1}{s-\mathcal{L}}\,\tilde{\rho}(0)\right]. \tag{5.148}$$
(general solution to master equation)

Again, we have assumed that $\mathcal{L}$ is time-independent: this is where it helps to use the density operator $\tilde{\rho}$ in the rotating frame.

5.5.2.2 Torrey's Solutions

To apply this method to the two-level atom, it is useful to do so in a slightly different form.[16] Starting with the optical Bloch equations in the form (5.127), we can write these in matrix form as

$$\begin{aligned}\partial_t\langle\boldsymbol{\sigma}\rangle &= \begin{bmatrix} -\gamma_\perp & \Delta & 0 \\ -\Delta & -\gamma_\perp & -\Omega \\ 0 & \Omega & -\Gamma \end{bmatrix}\langle\boldsymbol{\sigma}\rangle - \begin{bmatrix} 0 \\ 0 \\ \Gamma \end{bmatrix} \\ &=: \mathbf{Q}\langle\boldsymbol{\sigma}\rangle - \boldsymbol{\Gamma}.\end{aligned} \tag{5.149}$$

[16] Here we follow H. C. Torrey, "Transient Nutations in Nuclear Magnetic Resonance," *Physical Review* **76**, 1059 (1949) (doi: 10.1103/PhysRev.76.1059). See also L. Allen and J. H. Eberly, *Optical Resonance and Two-Level Atoms* (Wiley, 1975), section 3.5, p. 62.

We thus have a smaller matrix to deal with than the Laplacian, but at the expense of an extra constant in the equation. Now taking the Laplace transform of this equation, and using the fact that $\mathscr{L}[1] = 1/s$, as we used in Eq. (5.142), we find

$$s\mathscr{L}[\langle\boldsymbol{\sigma}\rangle] - \langle\boldsymbol{\sigma}\rangle_0 = \mathbf{Q}\mathscr{L}[\langle\boldsymbol{\sigma}\rangle] - \frac{1}{s}. \tag{5.150}$$

Rearranging this, we get a slightly modified form of the resolvent solution:

$$\mathscr{L}[\langle\boldsymbol{\sigma}\rangle] = \frac{1}{s(s-\mathbf{Q})}\left(s\langle\boldsymbol{\sigma}\rangle_0 - \boldsymbol{\Gamma}\right). \tag{5.151}$$

The analog of the resolvent operator here is

$$\frac{1}{s-\mathbf{Q}} = \frac{1}{f(s)}\begin{bmatrix} (s+\gamma_\perp)(s+\Gamma)+\Omega^2 & (s+\Gamma)\Delta & -\Delta\Omega \\ -(s+\Gamma)\Delta & (s+\gamma_\perp)(s+\Gamma) & -(s+\gamma_\perp)\Omega \\ -\Delta\Omega & (s+\gamma_\perp)\Omega & (s+\gamma_\perp)^2+\Delta^2 \end{bmatrix}, \tag{5.152}$$

where

$$f(s) = \det(s-\mathbf{Q}) = (s+\gamma_\perp)\Omega^2 + (s+\Gamma)\left((s+\gamma_\perp)^2+\Delta^2\right). \tag{5.153}$$

Looking at this solution, we can see that each component of $\mathscr{L}[\langle\boldsymbol{\sigma}\rangle]$ can be written in the form

$$\frac{g(s)}{sf(s)}, \tag{5.154}$$

where $f(s)$ is a cubic polynomial in s, and $g(s)$ is also at most cubic in s [and thus of smaller degree than $f(s)$]. We can thus write $f(s)$ in terms of three roots $-a_{1,2,3}$,

$$f(s) = (s+a_1)(s+a_2)(s+a_3), \tag{5.155}$$

and comparison to the form (5.153), where all the polynomial coefficients are positive, tells us that the product $a_1a_2a_3 > 0$. For this to hold, as well as to maintain the positivity of the other coefficients, one of these numbers must be positive, and the other two must either also be real and positive, or they must form a complex-conjugate pair. Thus, we may write

$$f(s) = (s+c)(s+a-ib)(s+a+ib) = (s+c)\left[(s+a)^2+b^2\right], \tag{5.156}$$

where $a, c > 0$, and b is real for conjugate roots, but could otherwise be imaginary. This form for $f(s)$ implies the partial-fraction decomposition

$$\frac{g(s)}{sf(s)} = \frac{A}{s+c} + \frac{B(s+a)+C}{(s+a)^2+b^2} + \frac{D}{s}, \tag{5.157}$$

where A, B, C, and D are constants that depend both on the initial conditions and on which component of $\langle\boldsymbol{\sigma}\rangle$ we are considering. Computing the inverse Laplace transform of this expression, we can see that the solution for any component of $\langle\boldsymbol{\sigma}\rangle$ can be written in the form

$$\langle\sigma_\alpha(t)\rangle = Ae^{-ct} + Be^{-at}\cos bt + \frac{C}{b}e^{-at}\sin bt + D. \tag{5.158}$$

(general form of solution to optical Bloch equations)

We can thus see that the general solution is reasonably simple, and even without finding the explicit forms of the coefficients, we can interpret the different terms. The first term represents exponential decay of the populations and coherences, as we would expect from the damping terms. The second and third terms represent exponentially damped Rabi oscillations with different phases. The final term is just the steady-state solution that we derived in Section 5.5.1.

5.5.2.3 Exact Resonance

Unfortunately, since the general solution depends on the roots of $f(s)$, which do not have a simple form, we don't gain much intuition by trying to write them down. Torrey[17] identified three regimes where the solutions are reasonably simple: (1) exact resonance ($\Delta = 0$), (2) damping such that $\gamma_\perp = \Gamma$, and (3) strong excitation, such that $\Omega \gg \Gamma, \gamma_\perp$. A fourth case is the weak-excitation regime ($\Omega \ll \Gamma, \gamma_\perp$), which we will see later reproduces the results of the classical Lorentz model. We will only consider the first case ($\Delta = 0$), for an atom initially in the ground state ($\langle\sigma_x(0)\rangle = \langle\sigma_y(0)\rangle = 0$, $\langle\sigma_z(0)\rangle = -1$), to get a feel for the damped solutions. For $\Delta = 0$, and for homogeneous broadening ($\gamma_\perp = \Gamma/2$),

$$f(s) = (s + \Gamma/2)\Omega^2 + (s + \Gamma)(s + \Gamma/2)^2. \tag{5.159}$$

One root (the real one) is $s = -\Gamma/2$, and thus

$$c = \frac{\Gamma}{2}, \tag{5.160}$$

giving the decay rate of the damped but nonoscillating term. The other two roots are

$$s = -\frac{3\Gamma}{4} \pm i\Omega_\Gamma, \tag{5.161}$$

where

$$\Omega_\Gamma := \sqrt{\Omega^2 - \left(\frac{\Gamma}{4}\right)^2} \tag{5.162}$$

is the Rabi flopping frequency in the presence of damping. Note that in the limit $\Omega \longrightarrow 0$, the roots become $s = -\Gamma$ and $s = -\Gamma/2$, which is what we expect for the longitudinal and transverse decay rates in the absence of driving. Thus, we have fixed the other two roots,

$$a = \frac{3\Gamma}{4}, \qquad b = \Omega_\Gamma, \tag{5.163}$$

in the notation above, giving the decay rate and oscillation frequency, respectively, of the oscillating terms.

x-component: Now we need to determine the coefficients of the different terms for the three components. First for the $\langle\sigma_x\rangle$ case. Starting with the steady-state coefficient D, we note from Eq. (5.138) that on resonance the steady-state value of $\tilde{\rho}_{\rm eg}$ is purely imaginary. But $\langle\sigma_x\rangle$ is the real part of $\tilde{\rho}_{\rm eg}$, so $D = 0$. From (5.157), we see that we can recover A from

$$A = \lim_{s\to -c}\left(\frac{g(s)}{sf(s)}(s + c)\right) = 0, \tag{5.164}$$

since $g(s) = 0$ on resonance and $\langle\sigma_x(0)\rangle = 0$. Also, from $\langle\sigma_x(0)\rangle = 0$ and Eq. (5.158), we see that $A+B+D = 0$, and thus $B = 0$. The remaining coefficient to determine is C. From Eq. (5.149), the initial value of the derivative is

$$\partial_t\langle\sigma_x(0)\rangle = \Delta\langle\sigma_y(0)\rangle - \frac{\Gamma}{2}\langle\sigma_x(0)\rangle = 0. \tag{5.165}$$

Similarly differentiating Eq. (5.158),

$$\partial_t\langle\sigma_x(0)\rangle = -cA - aB + C, \tag{5.166}$$

and comparing these two expressions gives $C = 0$. Thus,

$$\langle\sigma_x(t)\rangle = 0,$$

(solution to optical Bloch equations, $\Delta = 0$, $\gamma_\perp = \Gamma/2$, $\rho(0) = |{\rm g}\rangle\langle{\rm g}|$) (5.167)

as we expect from the resonant case.

[17] H. C. Torrey, *op. cit.*

y-component: We can obtain the steady state here from $\langle\sigma_y\rangle = i\tilde{\rho}_{\rm eg} - i\tilde{\rho}_{\rm ge}$. Using Eq. (5.138), we find

$$D = \frac{2\Omega/\Gamma}{1+2\dfrac{\Omega^2}{\Gamma^2}}. \tag{5.168}$$

The A coefficient is again given by

$$A = \lim_{s\to -c}\left(\frac{g(s)}{sf(s)}(s+c)\right) = \lim_{s\to -\Gamma/2}\left(\frac{(s+\Gamma/2)\Omega(s+\Gamma)}{(s+\Gamma/2)\Omega^2+(s+\Gamma)(s+\Gamma/2)^2}\frac{s+\Gamma/2}{s}\right) = 0. \tag{5.169}$$

Setting $\langle\sigma_y(0)\rangle = 0$ in Eq. (5.158), we see that $A+B+D=0$, and thus $B=-D$. Finally, matching the initial time derivatives of $\langle\sigma_y(0)\rangle$ from Eqs. (5.149) and (5.158) gives

$$C = cA + aB - \Delta\langle\sigma_x(0)\rangle - \frac{\Gamma}{2}\langle\sigma_y(0)\rangle - \Omega\langle\sigma_z(0)\rangle = -\frac{3\Omega/2}{1+2\dfrac{\Omega^2}{\Gamma^2}} + \Omega = \Omega\left(\frac{\Omega^2-\Gamma^2/4}{\Omega^2+\Gamma^2/2}\right). \tag{5.170}$$

The complete solution is thus

$$\langle\sigma_y(t)\rangle = \frac{\Omega\Gamma}{\Omega^2+\Gamma^2/2}\left[1 - e^{-(3\Gamma/4)t}\left(\cos\Omega_\Gamma t - \frac{\Omega^2-\Gamma^2/4}{\Gamma\Omega_\Gamma}\sin\Omega_\Gamma t\right)\right].$$

(solution to optical Bloch equations, $\Delta=0$, $\gamma_\perp=\Gamma/2$, $\rho(0)=|{\rm g}\rangle\langle{\rm g}|$) (5.171)

We can clearly see the damped oscillations in the dipole moment at the frequency Ω_Γ.

z-component: The steady state comes from $\langle\sigma_z\rangle = \rho_{\rm ee} - \rho_{\rm gg} = 2\rho_{\rm ee} - 1$. Using Eq. (5.137) for the steady-state solution, we find

$$D = \frac{2\Omega^2/\Gamma^2}{1+2\dfrac{\Omega^2}{\Gamma^2}} - 1 = -\frac{1}{1+2\dfrac{\Omega^2}{\Gamma^2}} = \frac{\Omega^2}{\Omega^2+\Gamma^2/2} - 1. \tag{5.172}$$

Note that $D\longrightarrow 0$ as $\Omega\longrightarrow\infty$, as we expect it to. The A coefficient is once again given by

$$A = \lim_{s\to -c}\left(\frac{g(s)}{sf(s)}(s+c)\right) = \lim_{s\to -\Gamma/2}\left(\frac{-(s+\Gamma/2)^2(s+\Gamma)}{(s+\Gamma/2)\Omega^2+(s+\Gamma)(s+\Gamma/2)^2}\frac{s+\Gamma/2}{s}\right) = 0. \tag{5.173}$$

Setting $\langle\sigma_z(0)\rangle = -1$ in Eq. (5.158) gives $A+B+D=-1$, and thus

$$B = -D - 1 = -\frac{\Omega^2}{\Omega^2+\Gamma^2/2}. \tag{5.174}$$

Finally, again matching the initial time derivatives of $\langle\sigma_z(0)\rangle$ from Eqs. (5.149) and (5.158),

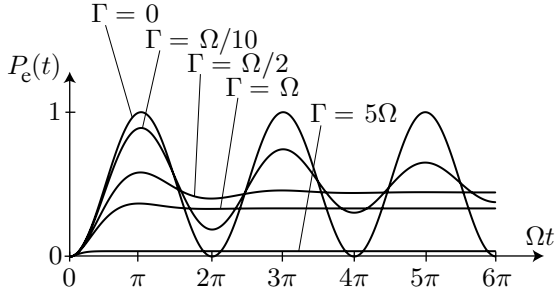
$$C = cA + aB + \Omega\langle\sigma_y(0)\rangle - \Gamma\langle\sigma_z(0)\rangle - \Gamma = \frac{(-3\Gamma/4)\Omega^2}{\Omega^2+\Gamma^2/2}. \tag{5.175}$$

The complete solution is thus

$$\langle\sigma_z(t)\rangle = -1 + \frac{\Omega^2}{\Omega^2+\Gamma^2/2}\left[1 - e^{-(3\Gamma/4)t}\left(\cos\Omega_\Gamma t + \frac{3\Gamma}{4\Omega_\Gamma}\sin\Omega_\Gamma t\right)\right].$$

(solution to optical Bloch equations, $\Delta=0$, $\gamma_\perp=\Gamma/2$, $\rho(0)=|{\rm g}\rangle\langle{\rm g}|$) (5.176)

We can again clearly see the damped Rabi oscillations at the damped Rabi frequency Ω_Γ.



The damped Rabi oscillations in the excited state population $P_\mathrm{e} \equiv \rho_\mathrm{ee} = (\langle\sigma_z\rangle + 1)/2$ are shown above.

5.5.3 Operator Form

The optical Bloch equations (5.125) can be written in the equivalent and compact operator form of a **master equation** for the density operator as

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF}, \tilde{\rho}\right] + \Gamma\mathcal{D}[\sigma]\tilde{\rho} + \frac{\gamma_\mathrm{c}}{2}\mathcal{D}[\sigma_z]\tilde{\rho},$$

(optical Bloch equations, operator form) (5.177)

where we have defined the **Lindblad superoperator**

$$\mathcal{D}[c]\rho := c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right). \tag{5.178}$$

(Lindblad superoperator)

The last two terms in the master equation correspond to decay of the excited state and extra dephasing due to collisions, respectively (the latter turns out to have the form of a measurement of σ_z, and thus causes increased uncertainty in the transverse spin components). Again, this is a "superoperator" on ρ because it operates on it from both sides. We will see that this form is universal, in the sense that all *Markovian* master equations (those where $\partial_t\rho(t)$ depends on ρ at time t and not any other time) always have damping terms in this **Lindblad form**—that is, any form where the master equation is written purely in terms of Hamiltonian commutators and Lindblad superoperators.

Note also that in many cases, these damping terms *must* be written as separate terms from the Hamiltonian part. For example, take the naturally damped atom,

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF}, \tilde{\rho}\right] + \Gamma\mathcal{D}[\sigma]\tilde{\rho}. \tag{5.179}$$

It is common in the literature to define an effective, non-Hermitian Hamiltonian by

$$\tilde{H}_\mathrm{eff} := \tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF} - i\frac{\hbar\Gamma}{2}\sigma^\dagger\sigma. \tag{5.180}$$

In this way, we can absorb *some* of the Lindblad terms into the Hamiltonian:

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_\mathrm{eff}\tilde{\rho} - \tilde{\rho}H_\mathrm{eff}^\dagger\right] + \Gamma\sigma\tilde{\rho}\sigma^\dagger. \tag{5.181}$$

However, one term remains. What is the meaning of this? Consider that

$$\sigma\tilde{\rho}\sigma^\dagger = \rho_\mathrm{ee}|\mathrm{g}\rangle\langle\mathrm{g}|. \tag{5.182}$$

Thus, this operator only enters the $\partial_t\rho_\mathrm{gg}$ equation, and in fact it is the term that *returns the decayed population to the ground state*. Thus, the non-Hermitian Hamiltonian can correctly handle both the decay of the excited state and the coherence decay. However, on its own, it does not preserve the trace of the density operator, as we can see from the form $\exp(-iH_\mathrm{eff}t/\hbar)$ of the time-evolution operator. Renormalization at each time step, then, introduces an extra term that cannot be written in terms of a Hamiltonian.

5.5.4 Orders of Magnitude

Now that we have introduced a couple of time scales, it is useful to estimate their magnitude. First, the Rabi frequency. Suppose we have a Gaussian beam with beam-waist parameter ($1/e^2$ intensity radius) w_0.[18] In terms of the total beam power P, the intensity at the beam center is

$$I = \frac{2P}{\pi w_0^2}. \tag{5.183}$$

[18] Daniel A. Steck, *Classical and Modern Optics* (2006), chapter 6, available online at `http://steck.us/teaching`.

In terms of the field amplitude, the intensity is

$$I = \frac{E_0^2}{2\eta_0}, \tag{5.184}$$

where η_0 is the impedence of the vacuum, so that

$$E_0 = \sqrt{2\eta_0 I}. \tag{5.185}$$

We may thus write the Rabi frequency as

$$\Omega = \frac{|\langle \mathrm{g}|d_z|\mathrm{e}\rangle|\sqrt{2\eta_0 I}}{\hbar} = \left[\frac{4\eta_0 |\langle \mathrm{g}|d_z|\mathrm{e}\rangle|^2 P}{\pi\hbar^2 w_0^2}\right]^{1/2}. \tag{5.186}$$

The dipole matrix element $\langle \mathrm{g}|d_z|\mathrm{e}\rangle$ is of the order of ea_0, where e is the fundamental charge and $a_0 \approx 0.529$ Å is the Bohr radius. On the D_2 transition (780 nm) in ^{87}Rb, depending on the details of the interaction, the matrix element can be as high as about 3.0 $ea_0 \approx 2.5 \times 10^{-29}$ C $\cdot$ m. A beam power of 10 mW is achievable by a very modest diode laser, and a beam waist of $w_0 = 1$ mm is fairly typical. Putting in these numbers, we find $\Omega/2\pi \approx 84$ MHz. Much larger local intensities are possible by using larger lasers (on the order of a couple of W for the largest cw lasers at 780 nm) and smaller beam waists (on the order of 10 μm is easy to achieve), giving Rabi frequencies in the tens of GHz range. Even larger values can be achieved with pulsed lasers, until the fields are strong enough that the whole two-level treatment breaks down.

We have also introduced the spontaneous decay rate Γ. Back when we studied the rate equations, we related this quantity to the Einstein B coefficient. Shortly, we will connect this to the dipole matrix element. But typically, for atomic transitions in the optical, decay rates $\Gamma/2\pi$ are on the order of several MHz (6.1 MHz for ^{87}Rb, corresponding to a lifetime of 26 ns). The decay rates turn out to scale as ω_0^3 (Chapter 11), so these can become larger as the transition becomes more energetic. For ''dipole forbidden'' transitions, the decay rates can be substantially smaller.

5.6 Consistency with Other Models

5.6.1 Classical Limit

We will now show the connection between the optical Bloch equations and the classical Lorentz atom. Formally, these two problems are *equivalent* in the case of a weak drive. The general idea is to construct the *damped* version of the quantum harmonic oscillator, and then show that it reduces to both the Lorentz model and the optical Bloch equations in the weak-excitation limit.

5.6.1.1 Review: Harmonic Oscillator in Quantum Mechanics

We will take the following results for the quantum harmonic oscillator to be given. The Hamiltonian is

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 x^2 = \hbar\omega_0\left(a^\dagger a + \frac{1}{2}\right), \tag{5.187}$$

where the creation ($a^\dagger$) and annihilation (a) operators are defined via the relation

$$a = \frac{1}{\sqrt{2}}\left(\frac{x}{x_0} + i\frac{x_0 p}{\hbar}\right) \tag{5.188}$$

and its Hermitian adjoint. The length scale x_0 is given by

$$x_0 = \sqrt{\frac{\hbar}{m\omega_0}}. \tag{5.189}$$

The phase-space operators can be written in terms of the ladder operators as

$$x = \frac{x_0}{\sqrt{2}}\left(a + a^\dagger\right) \qquad p = \frac{\hbar}{i\sqrt{2}\,x_0}\left(a - a^\dagger\right). \tag{5.190}$$

We have the usual commutation relations

$$[x, p] = i\hbar \qquad [a, a^\dagger] = 1, \tag{5.191}$$

which are equivalent to each other via the above definitions.

We will denote the eigenstates of the Hamiltonian ("Fock states") by $|n\rangle$ for nonnegative n, with corresponding eigenvalues

$$E_n = \hbar\omega_0\left(n + \frac{1}{2}\right). \tag{5.192}$$

These states all have $\langle x\rangle = \langle p\rangle = 0$ and moments

$$\left\langle n\left|x^2\right|n\right\rangle = \frac{\hbar}{m\omega_0}\left(n + \frac{1}{2}\right), \qquad \left\langle n\left|p^2\right|n\right\rangle = m\hbar\omega_0\left(n + \frac{1}{2}\right). \tag{5.193}$$

In this basis, the ladder operators have the effect

$$a|n\rangle = \sqrt{n}\,|n-1\rangle \qquad a^\dagger|n\rangle = \sqrt{n+1}\,|n+1\rangle. \tag{5.194}$$

The eigenstate of the annihilation operator is the **coherent state**

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} e^{-|\alpha|^2/2}|n\rangle, \tag{5.195}$$

with eigenvalue α:

$$a|\alpha\rangle = \alpha|\alpha\rangle. \tag{5.196}$$

Note that the occupation probabilities for the number states form a Poisson distribution of mean α. The effect of the creation operator on the coherent state is more complicated:

$$a^\dagger|\alpha\rangle = (\partial_\alpha + \alpha^*)\,|\alpha\rangle. \tag{5.197}$$

The ground state $|0\rangle$ is a special case of a coherent state. The general coherent state $|\alpha\rangle$ has the same Gaussian probability-density profile as the ground state, but the centroid oscillates with frequency ω_0 and amplitude $\sqrt{2}\,x_0|\alpha|$ in position and $\sqrt{2}\,\hbar|\alpha|/x_0$ in momentum.

5.6.1.2 Evolution of the Means: Damped Quantum Harmonic Oscillator

We can add damping to the harmonic oscillator by adding an extra component in Lindblad form to the master equation:

$$\partial_t\rho = -\frac{i}{\hbar}[H, \rho] + \gamma D[a]\rho.$$

(damped quantum harmonic oscillator) (5.198)

Again, the Lindblad superoperator is given by

$$D[a]\rho := a\rho a^\dagger - \frac{1}{2}\left(a^\dagger a\rho + \rho a^\dagger a\right). \tag{5.199}$$

We already know from the discussion of the Moyal bracket that the *Hamiltonian* part of this evolution is classical. So now let's check this again and show the correspondence of this oscillator *with damping* with the

classical damped oscillator. The master equation implies the equation of motion for the expectation value of an arbitrary operator A:

$$\begin{aligned}\partial_t\langle A\rangle &= \mathrm{Tr}[A\partial_t\rho] \\ &= -\frac{i}{\hbar}\langle[A,H]\rangle + \gamma\left\langle a^\dagger A a - \frac{1}{2}\left(a^\dagger a A + A a^\dagger a\right)\right\rangle \\ &= -\frac{i}{2m\hbar}\langle[A,p^2]\rangle - \frac{im\omega_0^2}{\hbar}\langle[A,x^2]\rangle + \frac{\gamma}{2}\left\langle a^\dagger[A,a] + [a^\dagger,A]a\right\rangle.\end{aligned} \tag{5.200}$$

Recall that we are in the Schrödinger picture, so the time dependence is contained only in ρ, and not in A. For the expected position, we can use the commutation relation $[x,f(p)] = i\hbar\partial_p f(p)$ to evaluate $[x,H]$; we can also use $[x,a] = -x_0/\sqrt{2} = -[x,a^\dagger]$ to evaluate the dissipation term:

$$\left\langle a^\dagger[x,a] + [a^\dagger,x]a\right\rangle = -\frac{x_0}{\sqrt{2}}\left\langle a + a^\dagger\right\rangle = -\langle x\rangle\,. \tag{5.201}$$

The resulting position equation is

$$\partial_t\langle x\rangle = \frac{\langle p\rangle}{m} - \frac{\gamma}{2}\langle x\rangle. \tag{5.202}$$

Similarly, we can find the equation for the expected momentum by using $[p,f(x)] = -i\hbar\partial_x f(x)$ to evaluate the Hamiltonian term and $[p,a] = [p,a^\dagger] = \hbar/i\sqrt{2}\,x_0$ to evaluate the dissipation term. The resulting equation is

$$\partial_t\langle p\rangle = -m\omega_0^2\langle x\rangle - \frac{\gamma}{2}\langle p\rangle. \tag{5.203}$$

It may look funny to have damping terms on both Eqs. (5.202) and (5.203), but differentiating Eq. (5.202) and eliminating $\partial_t\langle p\rangle$ gives

$$\partial_t^2\langle x\rangle + \gamma\partial_t\langle x\rangle + \left(\omega_0^2 + \frac{\gamma^2}{4}\right)\langle x\rangle = 0$$

(centroid evolution, quantum damped harmonic oscillator) (5.204)

for the wave-packet centroid. This has the same form as for a classical damped oscillator:

$$\ddot{x} + \gamma\dot{x} + \omega_0^2 x = 0. \tag{5.205}$$

Note that we identify the frequency ω_0 in Eq. (5.204) as the renormalized oscillation frequency ω_γ of the damped oscillator, given by $\omega_\gamma^2 = \omega_0^2 - \gamma^2/4$, and not the true resonance frequency ω_0 that appears in the classical formula (5.205).

Adding a dipole interaction Hamiltonian for coupling to an external applied field,

$$H_{\mathrm{AF}} = -\mathbf{d}\cdot\mathbf{E} = exE_0^{(+)}e^{-i\omega t} + \text{c.c.}, \tag{5.206}$$

where e is the fundamental charge ($e > 0$), modifies the momentum equation to read (using $[p,H_{\mathrm{AF}}] = -i\hbar eE_0^{(+)}e^{-i\omega t} + \text{c.c.}$)

$$\partial_t\langle p\rangle = -m\omega_0^2\langle x\rangle - \frac{\gamma}{2}\langle p\rangle - \left(eE_0^{(+)}e^{-i\omega t} + \text{c.c.}\right). \tag{5.207}$$

Rewriting the equations as a second-order equation, we find

$$\partial_t^2\langle x\rangle + \gamma\partial_t\langle x\rangle + \left(\omega_0^2 + \frac{\gamma^2}{4}\right)\langle x\rangle = -\frac{eE_0^{(+)}}{m}e^{-i\omega t} + \text{c.c.},$$

(centroid evolution, with external drive) (5.208)

which is the equation for the driven Lorentz atom in the dipole approximation, if we again associate the wave-packet centroid with the classical electron position and interpret the frequency ω_0 properly. Of course, for typical atomic dipole transitions in the optical, $\gamma/\omega \sim 10^{-8}$, so that the difference between the resonance and damped oscillation frequencies is negligible.

Note also that the damped harmonic oscillator in the weak-driving limit occupies mostly the ground state, with small population in $|1\rangle$. The populations in higher-energy states are negligible, so we can identify the harmonic oscillator with the weakly driven, two-level atom by taking $|1\rangle \longrightarrow |\mathrm{e}\rangle$, $|0\rangle \longrightarrow |\mathrm{g}\rangle$, and $a \longrightarrow \sigma = |\mathrm{g}\rangle\langle\mathrm{e}|$. Then the master equation (5.198), including the dipole interaction becomes

$$\partial_t \rho = -i\omega_0[\sigma^\dagger\sigma, \rho] - \frac{i}{\hbar}[H_{\mathrm{AF}}, \rho] + \gamma D[\sigma]\rho, \tag{5.209}$$

which generates the usual optical Bloch equations for the two-level atom if we identify $\gamma = \Gamma$. The same interaction Hamiltonian above gives the dipole coupling of the two-level atom to a monochromatic field. Thus, we see in the weak-excitation limit ($s \ll 1$) that the Lorentz atom accurately describes the dynamics of a quantum-mechanical atom interacting with a classical monochromatic field.

Note that there is a subtlety involved in introducing the atom–field interaction H_{AF}. Really, in Eq. (5.208), we should have ended up with the same result, but with a factor of the oscillator strength f_0 multiplying the right-hand side. This is the same replacement $e/m \longrightarrow ef_0/m$ that we discussed in the classical treatment, which turns out to be necessary to get the quantitatively correct answer. How does this work out here? We have to be more careful about the replacement $\sigma \longrightarrow a$ in going from the two-level atom to the harmonic oscillator. Quantum mechanically, the dipole operator is

$$d_z = \langle\mathrm{g}|d_z|\mathrm{e}\rangle\left(\sigma + \sigma^\dagger\right), \tag{5.210}$$

while the classical dipole moment is

$$ex = \frac{ex_0}{\sqrt{2}}\left(a + a^\dagger\right) = \sqrt{\frac{e^2\hbar}{2m\omega_0}}\left(a + a^\dagger\right). \tag{5.211}$$

To put in the oscillator strength, we let $e/m \longrightarrow ef_0/m$:

$$ex = \sqrt{\frac{e^2\hbar f_0}{2m\omega_0}}\left(a + a^\dagger\right), \tag{5.212}$$

to make this expression quantitatively correct. Thus, to make the identification $\sigma \longrightarrow a$, we must also identify the coefficients

$$|\langle\mathrm{g}|d_z|\mathrm{e}\rangle|^2 = \frac{e^2\hbar f_0}{2m\omega_0}, \tag{5.213}$$

which is the correct relation between the oscillator strength and dipole matrix element. (See also Problem 5.10, where the same result comes out of comparing the classical and quantum expressions for the polarizability.)

5.6.1.3 Evolution of the Variances

The solution to a driven, damped harmonic oscillator turns out to be a coherent state, so let's see this explicitly. Let's evaluate the equation of motion for $\langle x^2\rangle$. Using $[x^2, p^2] = 2i\hbar[x, p]_+$ for the Hamiltonian term (with $[a, b]_+ := ab + ba$) and $[x^2, a] = -\sqrt{2}\,x_0 x = -[x^2, a^\dagger]$, we find

$$\begin{aligned}\partial_t\langle x^2\rangle &= \frac{1}{m}\langle[x, p]_+\rangle - \gamma\left[\frac{x_0}{2\hbar}\left(a^\dagger x - xa\right)\right] \\ &= \frac{1}{m}\langle[x, p]_+\rangle - \gamma\left(\langle x^2\rangle - \frac{\hbar}{2m\omega_0}\right).\end{aligned} \tag{5.214}$$

Using the variance definition

$$V_x := \langle x^2\rangle - \langle x\rangle^2, \tag{5.215}$$

the equation of motion becomes

$$\partial_t V_x = \partial_t \langle x^2 \rangle - 2\langle x \rangle \partial_t \langle x \rangle = \frac{2}{m} C_{xp} - \gamma \left(V_x - \frac{\hbar}{2m\omega_0} \right), \tag{5.216}$$

where the symmetrized covariance is

$$C_{xp} := \frac{1}{2} \langle [x, p]_+ \rangle - \langle x \rangle \langle p \rangle. \tag{5.217}$$

Similarly, for the momentum variance

$$V_p := \langle p^2 \rangle - \langle p \rangle^2, \tag{5.218}$$

we can use the commutators $[p^2, a] = [p^2, a^\dagger] = -2i\hbar p/\sqrt{2}x_0$ to obtain

$$\partial_t V_p = -2m\omega_0^2 C_{xp} - \gamma \left(V_p - \frac{m\omega_0 \hbar}{2} \right). \tag{5.219}$$

For the covariance, the Hamiltonian part requires the commutators

$$\begin{aligned} \left[[x, p]_+, x^2 \right] &= -4i\hbar x^2 \\ \left[[x, p]_+, p^2 \right] &= 4i\hbar p^2 \\ [[x, p]_+, a] &= i\hbar a^\dagger \\ \left[[x, p]_+, a^\dagger \right] &= i\hbar a. \end{aligned} \tag{5.220}$$

To derive the last two relations, it is useful to make the identification $[x, p]_+ = (\hbar/2i)(a^2 - {a^\dagger}^2)$, along with $[{a^\dagger}^2, a] = -2a^\dagger$ and $[a^2, a^\dagger] = 2a$. The equation of motion for the anticommutator is

$$\partial_t \langle [x, p]_+ \rangle = \frac{2}{m} \langle p^2 \rangle - 2m\omega_0^2 \langle x^2 \rangle - \gamma \langle [x, p]_+ \rangle, \tag{5.221}$$

so that the covariance equation becomes

$$\partial_t C_{xp} = \frac{1}{2} \partial_t \langle [x, p]_+ \rangle - \langle p \rangle \partial_t \langle x \rangle - \langle x \rangle \partial_t \langle p \rangle = \frac{1}{m} V_p - m\omega_0^2 V_x - \gamma C_{xp}. \tag{5.222}$$

Collecting all the equations for the means and variances together,

$$\begin{aligned} \partial_t \langle x \rangle &= \frac{1}{m} \langle p \rangle - \frac{\gamma}{2} \langle x \rangle \\ \partial_t \langle p \rangle &= -m\omega_0^2 \langle x \rangle - \frac{\gamma}{2} \langle p \rangle \\ \partial_t V_x &= \frac{2}{m} C_{xp} - \gamma \left(V_x - \frac{\hbar}{2m\omega_0} \right) \\ \partial_t V_p &= -2m\omega_0^2 C_{xp} - \gamma \left(V_p - \frac{m\omega_0 \hbar}{2} \right) \\ \partial_t C_{xp} &= \frac{1}{m} V_p - m\omega_0^2 V_x - \gamma C_{xp}. \end{aligned}$$

(damped quantum harmonic oscillator, evolution of means and variances) (5.223)

This is sufficient to completely characterize a Gaussian state for the damped harmonic oscillator. Of course to consider forcing, we still need to add a dipole interaction Hamiltonian. But as a check, the steady state here should be the ground state of the harmonic oscillator. However, when we compute the steady state of the above equations, we find

$$\langle x \rangle = \langle p \rangle = 0, \tag{5.224}$$

(steady-state means)

while

$$V_x = \frac{\hbar}{2m\omega_0}, \qquad V_p = \frac{m\omega_0\hbar}{2}, \qquad C_{xp} = 0. \tag{5.225}$$
(steady-state variances)

This is a state of minimum uncertainty (where the generalized uncertainty relation is $V_x V_p - C_{xp}^2 \geq \hbar^2/4$), and thus must be Gaussian.[19] The variances are indeed the same as for $|0\rangle$ from Eq. (5.193), so the oscillator damps to the ground state.

In the driven case, note that the means and variances of Eqs. (5.223) are uncoupled. For a sinusoidal drive of the form (5.206), the means evolve according to the classical equation of motion (5.208), so that in steady state the wave packet centroid oscillates with an amplitude given by the steady state of Eq. (5.208). Hence, the steady state is just a coherent state $|\alpha\rangle$ with amplitude

$$|\alpha| = \frac{e|E_0^{(+)}|/m}{(\omega^2 - \omega_0^2)^2 + \gamma^2\omega^2}, \tag{5.226}$$

again showing the classical nature of the solution.

5.6.2 Rate-Equation Limit

We can also show that under certain conditions, the optical Bloch equations (5.125) reduce to the rate equations of Chapter 3. Recall the form of the optical Bloch equations:

$$\begin{aligned} \partial_t \rho_{ee} &= i\frac{\Omega}{2}(\tilde{\rho}_{eg} - \tilde{\rho}_{ge}) - \Gamma\rho_{ee} \\ \partial_t \tilde{\rho}_{eg} &= -(\gamma_\perp - i\Delta)\tilde{\rho}_{eg} + i\frac{\Omega}{2}(\rho_{ee} - \rho_{gg}). \end{aligned} \tag{5.227}$$

In the case of strong collisional damping $\gamma_\perp \gg \Omega, \Gamma$, we can note that the coherences will be damped very quickly, whereas the populations will continue to evolve on much longer time scales. We can exploit this separation of time scales and make the **adiabatic approximation**, where we focus only on the slow population dynamics by assuming the coherences are always approximately in equilibrium. Thus, approximate steady state of the coherence equations give

$$\begin{aligned} (\gamma_\perp - i\Delta)\tilde{\rho}_{eg} &= \quad i\frac{\Omega}{2}(\rho_{ee} - \rho_{gg}) \quad (\text{for } \partial_t\tilde{\rho}_{eg} \approx 0) \\ (\gamma_\perp + i\Delta)\tilde{\rho}_{ge} &= -i\frac{\Omega}{2}(\rho_{ee} - \rho_{gg}) \quad (\text{for } \partial_t\tilde{\rho}_{ge} \approx 0). \end{aligned} \tag{5.228}$$

Adding these two equations together gives

$$(\tilde{\rho}_{eg} + \tilde{\rho}_{ge}) = i\frac{\Delta}{\gamma_\perp}(\tilde{\rho}_{eg} - \tilde{\rho}_{ge}), \tag{5.229}$$

while subtracting them gives

$$\gamma_\perp(\tilde{\rho}_{eg} - \tilde{\rho}_{ge}) - i\Delta(\tilde{\rho}_{eg} + \tilde{\rho}_{ge}) = i\Omega(\rho_{ee} - \rho_{gg}). \tag{5.230}$$

Combining these two relations to obtain the adiabatic difference of the coherences, we find

$$\gamma_\perp\left(1 + \frac{\Delta^2}{\gamma_\perp^2}\right)(\tilde{\rho}_{eg} - \tilde{\rho}_{ge}) = i\Omega(\rho_{ee} - \rho_{gg}). \tag{5.231}$$

We can now put this into the first population equation to obtain the adiabatic evolution equation

$$\partial_t\rho_{ee} = -\Gamma\rho_{ee} - \frac{\Omega^2}{2\gamma_\perp(1 + \Delta^2/\gamma_\perp^2)}(\rho_{ee} - \rho_{gg}). \tag{5.232}$$

[19] Eugen Merzbacher, *Quantum Mechanics*, 3rd ed. (Wiley, 1998), pp. 219-20.

This result is now formally equivalent to the rate equation for nearly monochromatic light [Eq. (3.17),

$$\partial_t N_2 = -A_{21} N_2 - \sigma(\omega) \frac{I}{\hbar\omega} (N_2 - N_1) ,$$

(rate-equation limit of optical Bloch equations) (5.233)

where we have assumed no degeneracy for the two-level atom ($g_1 = g_2 = 1$).

To compare the Einstein equation to the adiabatic result, we can clearly identify

$$\rho_{ee} \longrightarrow N_2/N, \qquad \rho_{gg} \longrightarrow N_1/N, \qquad \Gamma \longrightarrow A_{21}. \tag{5.234}$$

Comparing the coefficients of the stimulated emission and absorption terms is less straightforward but very useful. Recall from Eq. (3.18) that the resonant cross section is

$$\sigma(\omega) = A_{21} \frac{\lambda^2}{4} s(\omega), \tag{5.235}$$

while for a Lorentzian line shape $s(\omega)$, we have from Eq. (3.19)

$$s(\omega) = \frac{\Delta\omega}{2\pi \left[(\Delta\omega/2)^2 + \Delta^2 \right]}. \tag{5.236}$$

Comparing the denominator of $s(\omega)$ to the similar denominator of Eq. (5.232), we conclude that we must identify the transverse decay rate with the line width:

$$\gamma_\perp \longrightarrow \frac{\Delta\omega}{2}. \tag{5.237}$$

We may thus write the line-shape function as

$$s(\omega) = \frac{\gamma_\perp}{\pi \left(\gamma_\perp^2 + \Delta^2 \right)}. \tag{5.238}$$

Now identifying the coefficients of the spontaneous emission and absorption terms,

$$\frac{\gamma_\perp \Omega^2}{2 \left(\gamma_\perp^2 + \Delta^2 \right)} \longrightarrow \frac{\sigma(\omega) I}{\hbar\omega} = \frac{\Gamma \lambda^2 \gamma_\perp I}{4\pi \left(\gamma_\perp^2 + \Delta^2 \right) \hbar\omega}. \tag{5.239}$$

Thus, we have

$$\Omega^2 = \frac{\Gamma \lambda^2 I}{2\pi\hbar\omega} = \frac{2\pi\Gamma c^2 I}{\hbar\omega^3}. \tag{5.240}$$

Using $\Omega = -\langle g|d_z|e\rangle E_0/\hbar$ for z-polarized light, and $I = E_0^2/2\eta = (\epsilon_0 c/2) E_0^2$, we find

$$\Gamma = \frac{\omega^3}{\pi\epsilon_0 \hbar c^3} |\langle g|d_z|e\rangle|^2. \tag{5.241}$$

This analysis is only valid near resonance, so $\omega \approx \omega_0$, and we take the atom to be spherically symmetric, so that $|\langle g|d_x|e\rangle|^2 = |\langle g|d_y|e\rangle|^2 = |\langle g|d_z|e\rangle|^2$, with the result

$$\Gamma = \frac{\omega_0^3}{3\pi\epsilon_0 \hbar c^3} |\langle g|\mathbf{d}|e\rangle|^2.$$

(relation between decay rate and dipole matrix elements) (5.242)

This is, in fact, the correct relation between the atomic decay rate and the atomic dipole matrix element that comes out of a full quantum electrodynamics calculation, as we will later see in Section 11.4.

5.6.2.1 Saturation Intensity

We can make a connection to the earlier notion of the **saturation intensity**. From our rate equation analysis, specifically Eq. (3.27), the population inversion for an exactly resonant drive is given by

$$\frac{N_2 - N_1}{N} = -\frac{1}{1 + \dfrac{2\sigma_0 I}{\hbar\omega_0 A_{21}}} = -\frac{1}{1 + \dfrac{I}{I_{\text{sat}}}}, \tag{5.243}$$

where we again have ignored any degeneracy. We have also defined the saturation intensity as

$$I_{\text{sat}} := \frac{\hbar\omega_0 A_{21}}{2\sigma_0}, \tag{5.244}$$
(saturation intensity)

as we did before in Eq. (3.31).

We can do the same thing for the optical Bloch equations. Recalling that the steady-state population in the excited state from Eq. (5.132) is

$$\rho_{\text{ee}}(t \longrightarrow \infty) = \frac{\Omega^2}{2\gamma_\perp \Gamma} \frac{1}{1 + \dfrac{\Delta^2}{\gamma_\perp^2} + \dfrac{\Omega^2}{\gamma_\perp \Gamma}}, \tag{5.245}$$

we can write

$$\rho_{\text{ee}}(t \longrightarrow \infty) = \frac{\Omega^2/\Gamma^2}{1 + 2\Omega^2/\Gamma^2} \tag{5.246}$$

for the case of exact resonance ($\Delta = 0$) and homongenous broadening ($\Gamma = 2\gamma_\perp$). Then the population inversion is given by

$$\rho_{\text{ee}} - \rho_{\text{gg}} = 2\rho_{\text{ee}} - 1 = -\frac{1}{1 + 2\Omega^2/\Gamma^2}. \tag{5.247}$$

Since Ω^2 scales as the intensity, we can similarly define the saturation intensity for the two-level atom to match (5.243):

$$\frac{I}{I_{\text{sat}}} \equiv \frac{2\Omega^2}{\Gamma^2}.$$
(saturation intensity related to Rabi frequency) (5.248)

Using $\Omega = -\langle \text{g}|\hat{\varepsilon}\cdot\mathbf{d}|\text{e}\rangle E_0/\hbar$ for arbitrary light polarization and $I = E_0^2/2\eta = (\epsilon_0 c/2)E_0^2$, this relation gives

$$I_{\text{sat}} = \frac{c\epsilon_0 \Gamma^2 \hbar^2}{4|\langle \text{g}|\hat{\varepsilon}\cdot\mathbf{d}|\text{e}\rangle|^2}. \tag{5.249}$$
(saturation intensity)

Thus, for example, we can write

$$\rho_{\text{ee}}(t \longrightarrow \infty) = \left(\frac{1}{2}\right) \frac{I/I_{\text{sat}}}{1 + 4\Delta^2/\Gamma^2 + I/I_{\text{sat}}}, \tag{5.250}$$
(steady-state population)

for the excited-state population in the case of homogenout broadening. Similarly, we may write

$$s = \frac{I/I_{\text{sat}}}{1 + 4\Delta^2/\Gamma^2} \tag{5.251}$$
(saturation parameter)

for the saturation parameter. The saturation effect here represents the *nonlinear response of the two-level atom to the field*, which is not predicted by the classical (linear) Lorentz model.

But now, are the two saturation intensities from Eqs. (5.244) and (5.249) equivalent? Consider the case of linearly polarized light ($\hat{\varepsilon} = \hat{z}$). Using

$$\Gamma = \frac{\omega_0^3}{\pi\epsilon_0\hbar c^3}|\langle \mathrm{g}|d_z|\mathrm{e}\rangle|^2 \tag{5.252}$$

to eliminate the dipole matrix element in Eq. (5.249), we find

$$I_{\mathrm{sat}} = \frac{\hbar\omega_0^3\Gamma}{4\pi c^2}. \tag{5.253}$$

In the case of the rate equations, the resonant cross section for linearly polarized light and random orientation of the atom from Eq. (3.21) is

$$\sigma_0 = \frac{\lambda_0^2}{2\pi}. \tag{5.254}$$

Substitution of this cross section into Eq. (5.244) and taking $A_{21} = \Gamma$ then gives exactly the same expression (5.253) for the saturation intensity.

5.6.2.2 Validity of the Rate-Equation Limit

As we discussed above, the optical Bloch equations can be accurately represented by the rate equations when the collisional damping rate is large, because the coherences are quickly damped away. The rate equations are also valid in the case of incoherent (broadband) excitation, since there the dipole is driven by many frequencies, and thus the dephasing of the different frequency components mimics fast damping.

5.7 Spectrum of Resonance Fluorescence

We will now consider the radiation (**resonance fluorescence**) due to a single, isolated atom driven by a monochromatic field. The optical Wiener–Khinchin theorem that we wrote down before in Eq. (2.21) was

$$\int_0^\infty I(\mathbf{r},\omega)e^{-i\omega\tau}\,d\omega = \frac{2}{\eta}\left\langle E^{(-)}(\mathbf{r},t)E^{(+)}(\mathbf{r},t+\tau)\right\rangle. \tag{5.255}$$

Inverting this relation, keeping in mind that $I(\omega)$ will be centered near a large, positive frequency (of the driving laser), gives the intensity spectrum in terms of the field autocorrelation function:

$$I(\mathbf{r},\omega) = \frac{1}{\pi\eta}\int_{-\infty}^\infty \left\langle E^{(-)}(\mathbf{r},t)E^{(+)}(\mathbf{r},t+\tau)\right\rangle e^{i\omega\tau}\,d\tau. \tag{5.256}$$

Note that the spectrum is real due to the time-inversion symmetry of the correlation function. Note also that we are maintaining a particular ordering of the positive- and negative-frequency components of the electric field, a point we will return to below.

Applying this to the scattered light from a driven, two-level atom, recall that we use ω to refer to the frequency of the driving field and ω_0 to refer to the atomic resonance frequency. Thus, we will use ω_{s} to refer to the frequency of the scattered light. Also, recall from Eq. (1.43) that in the radiation zone, the electric field of an oscillating dipole is

$$\mathbf{E}^{(+)}(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0 c^2}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}\right]\frac{\ddot{d}^{(+)}(t_{\mathrm{r}})}{r}. \tag{5.257}$$

This classical expression is still appropriate for radiation from a quantum dipole, although obviously we will need to remember that $d(t)$ here is a Heisenberg-picture operator. We will work in the near-resonant regime, and thus we assume that any scattered light will be close to ω_0 in frequency. Thus, we make the approximation

$$\ddot{d}^{(+)} \approx -\omega_0^2 d^{(+)}. \tag{5.258}$$

We also note that

$$|(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}|^2 = 1-|\hat{\varepsilon}\cdot\hat{r}|^2\,, \tag{5.259}$$

which allows us to use the angular distribution function

$$f_{\hat{\varepsilon}}(\theta,\phi) = \frac{3}{8\pi}\left(1-|\hat{r}\cdot\hat{\varepsilon}|^2\right). \tag{5.260}$$

from before. Putting all this together, we can write down the scattered spectrum as

$$I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \frac{\omega_0^4}{6\pi^2\epsilon_0 c^3 r^2} f_{\hat{\varepsilon}}(\theta,\phi)\int_{-\infty}^{\infty} d\tau\, e^{i\omega_{\rm s}\tau}\left\langle d^{(-)}(t)d^{(+)}(t+\tau)\right\rangle, \tag{5.261}$$

where we used $\eta = 1/\epsilon_0 c$. The angle brackets here imply both a time average and an expectation value, since we are now dealing with operators instead of classical quantities (Heisenberg-picture operators, that is, since they now carry the explicit time dependence). We can now separate out the dipole matrix elements from the dipole product to obtain

$$I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \frac{\omega_0^4|\langle {\rm g}|\mathbf{d}|{\rm e}\rangle|^2}{6\pi^2\epsilon_0 c^3 r^2} f_{\hat{\varepsilon}}(\theta,\phi)\int_{-\infty}^{\infty} d\tau\, e^{i\omega_{\rm s}\tau}\left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle. \tag{5.262}$$

Now we see the importance of the so called **normal ordering**, where $E^{(+)}$ appears to the right of $E^{(-)}$ in the expectation value in Eq. (5.256): it implies that the atomic lowering operator is to the right in the expectation value, which thus vanishes if the atom is in the ground state. With any other ordering, the expectation value would be nonzero for an atom in the ground state, and would thus need to be compensated by another explicit term. Finally, we can use the result (5.242) from our rate-equation analysis to write this expression in terms of the spontaneous decay rate:

$$I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \frac{\hbar\omega_0\Gamma}{2\pi r^2} f_{\hat{\varepsilon}}(\theta,\phi)\int_{-\infty}^{\infty} d\tau\, e^{i\omega_{\rm s}\tau}\left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle. \tag{5.263}$$

The spectral content is entirely in the integral factor, and thus we may define the (unnormalized) spectral function

$$S(\omega_{\rm s}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\omega_{\rm s}\tau}\left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle, \tag{5.264}$$

(radiation spectrum)

in terms of which the intensity spectral density becomes

$$I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \frac{\hbar\omega_0\Gamma}{r^2} f_{\hat{\varepsilon}}(\theta,\phi) S(\omega_{\rm s}). \tag{5.265}$$

(fluorescence spectral density)

Note that the expectation value here is in the standard frame, whereas the optical Bloch equations are in the rotating frame. To use the rotating-frame solutions, we transform as follows:

$$\langle\psi|\sigma^\dagger(t)\sigma(t+\tau)|\psi\rangle = \langle\tilde{\psi}|e^{i\omega t}\sigma^\dagger(t)\sigma(t+\tau)e^{-i\omega(t+\tau)}|\tilde{\psi}\rangle = e^{-i\omega\tau}\langle\tilde{\psi}|\sigma^\dagger(t)\sigma(t+\tau)|\tilde{\psi}\rangle, \tag{5.266}$$

so that we may use the spectral function

$$S(\omega_{\rm s}) := \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i(\omega_{\rm s}-\omega)\tau}\left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle \tag{5.267}$$

(radiation spectrum, rotating frame)

if we take the expectation value with respect to rotating-frame solutions. We will do this henceforth. Thus, to find the spectrum of the scattered light, we need to compute the two-time atomic correlation function [which is of course proportional to $g^{(1)}(\tau)$] and then compute its Fourier transform. We will do so shortly, but first we will make a few comments about the total scattered radiation.

5.7.1 Scattering Cross Section, Line Shape, and Power Broadening

To compute the total scattered intensity, we can integrate the spectrum over all frequencies, keeping in mind that this is a one-sided spectrum:

$$\int_0^\infty S(\omega_s)\, d\omega_s = \langle \sigma^\dagger \sigma \rangle = \rho_{ee}(t \longrightarrow \infty). \tag{5.268}$$

Here, the steady-state population is appropriate in view of the time average. Thus, the scattered intensity is proportional to the excited-state population, as we would expect from the rate-equation model. We can, of course, define a *normalized* spectrum by

$$s(\omega_s) := \frac{S(\omega_s)}{\rho_{ee}(t \longrightarrow \infty)}. \tag{5.269}$$
(normalized spectrum)

The total scattered intensity is thus

$$I_{sc} = \frac{\hbar\omega_0\Gamma}{r^2} f_{\hat{\varepsilon}}(\theta,\phi)\rho_{ee}(t \longrightarrow \infty), \tag{5.270}$$

and we obtain the total scattered power by integrating the intensity over a spherical shell of radius r:

$$P_{sc} = \hbar\omega_0\Gamma\rho_{ee}(t \longrightarrow \infty). \tag{5.271}$$

The photon scattering rate is given by dividing the scattered power by the photon energy $\hbar\omega_0$ (again, assuming that scattering occurs near resonance),

$$R_{sc} = \Gamma\rho_{ee}(t \longrightarrow \infty), \tag{5.272}$$
(photon scattering rate)

and we see that this is simply the excited-state decay rate multiplied by the excited-state population.

A common way to describe the total scattered power is the **scattering cross section** σ_{sc}, which we define as the power radiated by the atom divided by the incident energy flux. That is, the scattered power is $\sigma_{sc}I$, where I is the intensity of the driving field. This is, of course, the same quantity that we defined in Section 1.2.1 in our treatment of the Lorentz atom. Recalling from Eq. (5.250) that we can write the excited-state population in steady state as

$$\rho_{ee}(t \longrightarrow \infty) = \left(\frac{1}{2}\right) \frac{I/I_{sat}}{1 + 4\Delta^2/\Gamma^2 + I/I_{sat}}, \tag{5.273}$$

we see that we can write the scattering cross section as

$$\sigma_{sc} = \frac{\sigma_0}{1 + 4\Delta^2/\Gamma^2 + I/I_{sat}}, \tag{5.274}$$
(scattering cross section)

where the on-resonance, small-signal cross section is given by

$$\sigma_0 := \frac{\hbar\omega_0\Gamma}{2I_{sat}}. \tag{5.275}$$
(on-resonance, small-signal cross section)

Obviously, the cross section falls to zero as the driving laser is tuned away from resonance, or due to saturation as the incident intensity becomes large.

We will now examine the absorption line shape, which is just the frequency dependence of the cross section. It is somewhat more convenient to examine the excited-state population, which we saw is proportional to the cross section. We can write the population in the form of Eq. (5.137) as

$$\rho_{ee}(t \longrightarrow \infty) = \frac{\Omega^2/\Gamma^2}{1 + \left(\dfrac{2\Delta}{\Gamma}\right)^2 + 2\dfrac{\Omega^2}{\Gamma^2}}. \tag{5.276}$$
(absorption line shape)

For a weak driving field, this reduces to

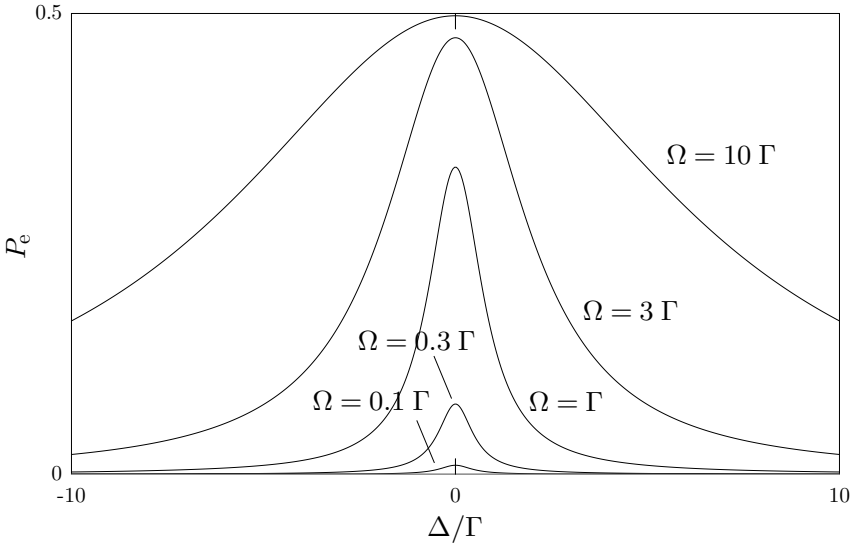
$$\rho_{\rm ee}(t \longrightarrow \infty) = \frac{\Omega^2/4}{\Delta^2 + \Gamma^2/4}. \qquad \text{(absorption line shape, weak drive)} \tag{5.277}$$

This line shape is Lorentzian with a width (FWHM) of Γ, and a maximum value of $(\Omega/\Gamma)^2 \ll 1$. Hence, Γ is also referred to as the transition linewidth (representing the *angular* frequency width of the transition).

In the strong-field limit, the excited-state population becomes

$$\rho_{\rm ee}(t \longrightarrow \infty) = \frac{\Omega^2/4}{\Delta^2 + \Omega^2/2}. \qquad \text{(absorption line shape, strong drive)} \tag{5.278}$$

This line shape is also Lorentzian, but now with a much larger width (FWHM) of $\sqrt{2}\,\Omega$, and a maximum value of $1/2$ as we expect for a saturated transition.



Thus, the line shape of the transition is effectively larger due to the strong coupling to the field. This phenomenon is called **power broadening** of the transition.

5.7.2 Coherent and Incoherent Scattering

Before grinding through the full solution of the $g^{(1)}(\tau)$ coherence function, we can gain some insight by first considering the asymptotic limit. Given that $\sigma(t)$ is a stochastically fluctuating operator, due to the random nature of spontaneous emission, it should become uncorrelated with itself at very different times. Thus,

$$\lim_{\tau \longrightarrow \infty} \left\langle \sigma^\dagger(t)\sigma(t+\tau) \right\rangle = \left\langle \sigma^\dagger(t) \right\rangle \left\langle \sigma(t+\tau) \right\rangle = \left| \tilde{\rho}_{\rm eg}(t \longrightarrow \infty) \right|^2 . \tag{5.279}$$

Thus, the coherence function decays to a possibly nonzero constant. The Fourier transform of this dc component leads to a delta function in the radiated spectrum, which we can refer to as the "coherent" component of the spectrum (even though it turns out not to be coherent to all orders). The decaying part of the correlation function leads to a broadened or "incoherent" component of the spectrum, which we will evaluate below. Formally, we decompose the scattering rate as

$$\begin{aligned} R_{\rm sc} &= R_{\rm sc}^{(\rm coh)} + R_{\rm sc}^{(\rm inc)} \\ &= \Gamma |\left\langle \sigma \right\rangle|^2 + \Gamma \left[\left\langle \sigma^\dagger \sigma \right\rangle - |\left\langle \sigma \right\rangle|^2 \right] \\ &= \Gamma |\tilde{\rho}_{\rm eg}(t \longrightarrow \infty)|^2 + \Gamma \left[\rho_{\rm ee}(t \longrightarrow \infty) - |\tilde{\rho}_{\rm eg}(t \longrightarrow \infty)|^2 \right]. \end{aligned}$$

(coherent and incoherent scattering rates) (5.280)

That is, the coherent part is due to the *square of the mean* of the dipole moment, which corresponds to what we found in the classical analysis (recall that the classical electron "position" was in fact the mean electron position). The incoherent part includes the *mean square* dipole moment, and thus accounts for the fluctuations of the dipole moment within the (possibly fictitious) ensemble.

As for the actual spectral content of the coherent part, a constant value for the coherence function in Eq. (5.267) gives a perfectly defined spectral peak

$$S^{(\mathrm{coh})}(\omega_\mathrm{s}) = \frac{1}{\pi}\int_{-\infty}^{\infty} d\tau\, e^{i(\omega_\mathrm{s}-\omega)\tau}|\tilde{\rho}_\mathrm{eg}(t\longrightarrow\infty)|^2 = 2|\tilde{\rho}_\mathrm{eg}(t\longrightarrow\infty)|^2\delta(\omega_\mathrm{s}-\omega). \tag{5.281}$$
(elastic spectral peak)

That is, the coherent part of the spectrum is exactly at the driving laser frequency, and thus represents *elastically* scattered light. (In the regime of far red detuning, the elastic scattering is often referred to as **Rayleigh scattering**.) We recall that this was also the classical prediction for the scattered light in steady state, since the steady state of the Lorentz atom is sinusoidal oscillation at the driving frequency. The incoherent part will be spread over a range of frequencies and thus will represent *inelastic* scattering.

What fraction of the scattered light is coherent? We can calculate this simply using the steady-state solutions of the optical Bloch equations in the forms (5.135):

$$\frac{R_\mathrm{sc}^{(\mathrm{coh})}}{R_\mathrm{sc}} = \frac{|\tilde{\rho}_\mathrm{eg}(\,t\longrightarrow\infty)|^2}{\rho_\mathrm{ee}(t\longrightarrow\infty)} = \frac{\dfrac{\Gamma}{4\gamma_\perp}\dfrac{s}{(1+s)^2}}{\dfrac{s/2}{1+s}} = \frac{\Gamma}{2\gamma_\perp}\frac{1}{(1+s)}. \tag{5.282}$$

Here, s is again the saturation parameter, given by

$$s = \frac{\Omega^2/\gamma_\perp\Gamma}{1+\Delta^2/\gamma_\perp^2}. \tag{5.283}$$

For homogeneous broadening (no collisions), $\gamma_\perp = \Gamma/2$, and thus

$$\frac{R_\mathrm{sc}^{(\mathrm{coh})}}{R_\mathrm{sc}} = \frac{1}{1+s}, \tag{5.284}$$
(fraction of elastic scattering)

where the saturation parameter becomes

$$s = \frac{2\Omega^2/\Gamma^2}{1+(2\Delta/\Gamma)^2}, \tag{5.285}$$

In this case, for small saturation parameter (small driving intensity), the scattered light is *completely elastic*. As the driving laser intensity increases, the elastic component vanishes, and the light is all inelastically scattered. In the presence of collisional damping ($\gamma_\perp > \Gamma/2$), there is some inelastic scattering even for a vanishingly small driving field.

In the homogeneously broadened case, the steady-state solutions again give explicit expressions for the components. The total scattering rate is

$$R_\mathrm{sc} = \Gamma\rho_\mathrm{ee}(t\longrightarrow\infty) = \frac{\Gamma}{2}\frac{s}{1+s}, \tag{5.286}$$
(total scattering rate)

while the coherent scattering rate is

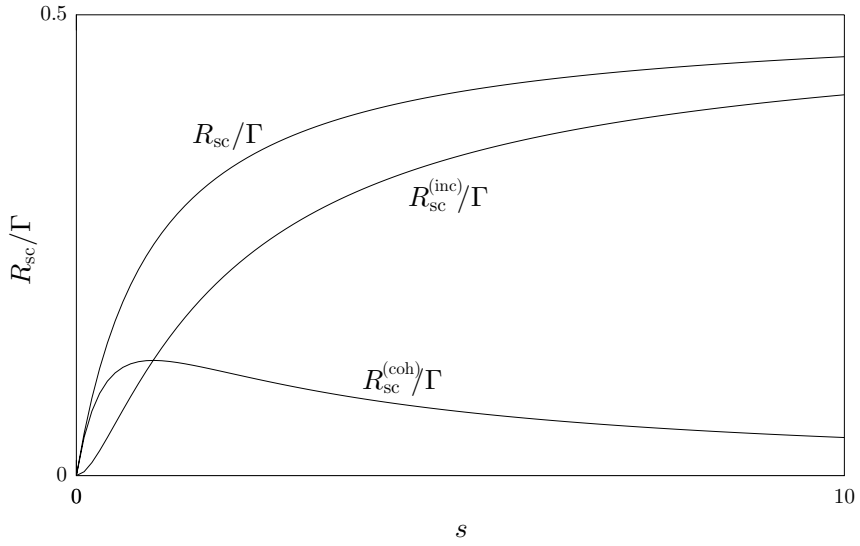
$$R_\mathrm{sc}^{(\mathrm{coh})} = \Gamma\,|\tilde{\rho}_\mathrm{eg}(\,t\longrightarrow\infty)|^2 = \frac{\Gamma}{2}\frac{s}{(1+s)^2}. \tag{5.287}$$
(coherent scattering rate)

The incoherent scattering rate is the difference of these two expressions:

$$R_{\rm sc}^{\rm (inc)} = R_{\rm sc} - R_{\rm sc}^{\rm (coh)} = \frac{\Gamma}{2}\frac{s^2}{(1+s)^2}. \qquad (5.288)$$

(incoherent scattering rate)

Thus, since the saturation parameter is proportional to the driving intensity, for small drive intensities the coherent component ''turns on'' linearly with the intensity, whereas the incoherent component turns on quadratically with intensity.



We thus see that the coherent component represents the linear response of the atom to the field, and thus the classical component. The incoherent part is the nonlinear part of the response of the two-level atom to the field, and is manifestly quantum.

5.7.3 Quantum Regression Theorem

In order to evaluate the frequency dependence of the scattered radiation, we must calculate the correlation functions $\langle\sigma^\dagger(t)\sigma(t+\tau)\rangle$. One method of calculating these correlation functions comes from the **quantum regression theorem**,[20] which we now describe. The upshot is that according to the quantum regression theorem, two-time correlation functions obey the same equations of motion as one-time averages, which considerably simplifies their calculation. In the spirit of the Church of the Larger Hilbert Space, we will regard the evolution of the atomic system according to a master equation as unitary evolution of the system coupled to an external ''reservoir.''

Before we consider the correlation functions, we will briefly review the calculation of single-time averages. Recall that if A is a system operator (i.e., it does not operate on the reservoir coupled to the system), then its time-averaged value is given in the Heisenberg picture by

$$\langle A(t)\rangle = {\rm Tr}[A(t)\,\rho_{\rm SR}], \qquad (5.289)$$

where Tr is a trace over both the system and reservoir variables, $\rho_{\rm SR}$ is the composite density operator, $\rho = {\rm Tr}_{\rm R}[\rho_{\rm SR}]$ is the reduced density operator for the system, and $\rho_{\rm R} = {\rm Tr}_{\rm S}[\rho_{\rm SR}]$ is the reduced density operator for the reservoir. Also, ${\rm Tr}_{\rm S}$ and ${\rm Tr}_{\rm R}$ are partial traces over the system and reservoir, respectively.

[20] Melvin Lax, ''Quantum Noise. IV. Quantum Theory of Noise Sources,'' *Physical Review* **145**, 110 (1966) (doi: 10.1103/PhysRev.145.110); S. Swain, ''Master equation derivation of quantum regression theorem,'' *J. Phys. A* **14**, 2577 (1981) (doi: 10.1088/0305-4470/14/10/013).

We can now change to the Schrödinger representation using $A(t) = U^\dagger(t,0)\, A\, U(t,0)$, where $U(t,0)$ is the unitary time-evolution operator from 0 to t, with the result

$$\begin{aligned} \langle A(t)\rangle &= \mathrm{Tr}[A\, U(t,0)\, \rho_{\mathrm{SR}}\, U^\dagger(t,0)] \\ &= \mathrm{Tr}_{\mathrm{S}}[A\, \rho(t)], \end{aligned} \tag{5.290}$$

where we have used the invariance of the trace operation under cyclic permutations and we have carried out the trace over the reservoir by setting as before

$$\rho(t) = \mathrm{Tr}_{\mathrm{R}}[U(t,0)\, \rho_{\mathrm{SR}}\, U^\dagger(t,0)]. \tag{5.291}$$

Then, if $\rho(t)$ satisfies the master equation

$$\partial_t \rho(t) = \mathcal{L}\rho(t), \tag{5.292}$$

with Liouvillian operator $\mathcal{L}$, the evolution of $\langle A(t)\rangle$ can be computed by solving Eq. (5.292) for the time evolution of $\rho(t)$.

To calculate the correlation function (or two-time average) $\langle A(t)B(t+\tau)\rangle$, where $A(t)$ and $B(t)$ are arbitrary Heisenberg operators, we proceed in a similar manner. Factoring out the explicit time dependence and using the composition and inversion properties of the evolution operator,

$$\begin{aligned} U(t,t')U(t',t'') &= U(t,t'') \\ U^\dagger(t,t') &= U(t',t), \end{aligned} \tag{5.293}$$

we find

$$\begin{aligned} \langle A(t)B(t+\tau)\rangle &= \mathrm{Tr}[A(t)B(t+\tau)\rho_{\mathrm{SR}}] \\ &= \mathrm{Tr}[A\, U^\dagger(t+\tau,t)\, B\, U(t+\tau,t)\, \{U(t,0)\, \rho_{\mathrm{SR}}\, U^\dagger(t,0)\}] \\ &= \mathrm{Tr}[B\, U(t+\tau,t)\, \rho_{\mathrm{SR}}(t)\, A\, U^\dagger(t+\tau,t)\,]. \end{aligned} \tag{5.294}$$

Then if we define the two-time operator

$$\Lambda(t+\tau,t) := \mathrm{Tr}_{\mathrm{R}}[U(t+\tau,t)\, \rho_{\mathrm{SR}}(t)\, A\, U^\dagger(t+\tau,t)], \tag{5.295}$$

the two-time correlation function becomes

$$\langle A(t)B(t+\tau)\rangle = \mathrm{Tr}_{\mathrm{S}}[B\, \Lambda(t+\tau,t)]. \tag{5.296}$$

This last expression looks much like an evolving expectation value, as in Eq. (5.290), with $\Lambda(t+\tau,t)$ replacing the reduced density operator $\rho(t)$. Similarly, comparing to Eq. (5.291), the definition (5.295) of Λ looks much like a density operator, which ρ_{SR} replaced by $\rho_{\mathrm{SR}}(t)\, A$. Thus, we see that $\Lambda(t+\tau,t)$ obeys the same equation of motion as $\rho(t)$, but as a function of τ,

$$\partial_\tau \Lambda(t+\tau,t) = \mathcal{L}\Lambda(t+\tau,t), \tag{5.297}$$

because the time evolution is governed by the same evolution operators in each case. This evolution is subject to the boundary condition

$$\Lambda(t,t) = \rho(t)A. \tag{5.298}$$

Hence, in the long-time limit, we may summarize the quantum regression theorem as

$$\lim_{t\to\infty} \langle A(t)B(t+\tau)\rangle = \mathrm{Tr}_{\mathrm{S}}[B\Lambda(\tau)] \tag{5.299}$$

(quantum regression theorem)

where

$$\partial_\tau \Lambda(\tau) = \mathcal{L}\Lambda(\tau), \tag{5.300}$$

(quantum regression theorem: evolution)

and

$$\Lambda(0) = \rho(t \to \infty)A$$

(quantum regression theorem: initial condition) (5.301)

is the initial condition for the evolution.

If we apply the quantum regression theorem to the emission correlation operator $\langle \sigma^\dagger(t)\sigma(t+\tau)\rangle$, we find that the operator $\Lambda(\tau)$, which satisfies the master equation, has the specific initial condition $\Lambda(0) = \tilde{\rho}(t \to \infty)\sigma^\dagger$. Writing out the matrix elements of the initial condition explicitly, these conditions become

$$\Lambda_{\alpha\beta}(0) = \delta_{\beta g}\tilde{\rho}_{\alpha e}(t \to \infty). \tag{5.302}$$

Even more explicitly, $\Lambda_{ee}(0) = \Lambda_{ge}(0) = 0$, while $\Lambda_{eg}(0) = \rho_{ee}(t \to \infty)$ and $\Lambda_{gg}(0) = \tilde{\rho}_{ge}(t \to \infty)$. Then, in terms of the solution $\Lambda(\tau)$ of the optical Bloch equations with these initial conditions, the correlation function that we need is given by

$$\langle \sigma^\dagger\sigma(\tau)\rangle = \mathrm{Tr}_S[\sigma\Lambda(\tau)] = \Lambda_{eg}(\tau). \tag{5.303}$$

Using these two relations, we can now use the solutions of the optical Bloch equations that we already obtained to evaluate the two-time correlation function for the resonance fluorescence spectrum.

5.7.3.1 Alternate Form

The quantum regression theorem can be written in a useful alternate form as follows. Suppose that the one-time average of an operator A can be written in the form

$$\langle B(t)\rangle = \sum_j g_j(t)\,\langle B_j(0)\rangle\,, \tag{5.304}$$

where $g_j(t)$ are functions representing the solution in terms of initial conditions $\langle A_j(0)\rangle$ of some set of operators, then the two-time average may be written

$$\langle A(t)B(t+\tau)\rangle_{t\to\infty} \equiv \langle A(0)B(\tau)\rangle = \sum_j g_j(\tau)\,\langle AB_j\rangle_{t\to\infty}\,.$$

(quantum regression theorem) (5.305)

To show this, we can see that

$$\begin{aligned}
\langle A(0)B(\tau)\rangle &= \mathrm{Tr}_S\big[B\Lambda(\tau)\big] \\
&= \sum_j g_j(\tau)\,\mathrm{Tr}_S\big[B_j\Lambda(0)\big] \\
&= \sum_j g_j(\tau)\,\mathrm{Tr}_S\big[B_j\rho(t \longrightarrow \infty)A\big] \\
&= \sum_j g_j(\tau)\,\langle AB_j\rangle_{t\to\infty}\,,
\end{aligned} \tag{5.306}$$

where we used Eq. (5.304) in the second step, recalling that $\Lambda(\tau)$ is formally equivalent to a density operator, and we used the initial condition $\Lambda(0) = \rho(t \longrightarrow \infty)A$ in the third step.

This form of the quantum regression theorem can also be generalized a bit to read

$$\langle A(t)B(t+\tau)C(t)\rangle_{t\to\infty} \equiv \langle A(0)B(\tau)C(0)\rangle = \sum_j g_j(\tau)\,\langle AB_jC\,\rangle_{t\to\infty}\,.$$

(quantum regression theorem) (5.307)

This form is useful in computing the second-order coherence $g^{(2)}(\tau)$ with normally-ordered operators. We leave the proof of this form as an exercise (Problem 5.22).

5.7.4 Mollow Triplet

The above recipe for computing the spectrum is good for a numerical computation. However, we can make our lives a bit easier by modifying the equations for the analytical calculation. The resulting spectrum was first computed by Mollow,[21] and is now called the **Mollow spectrum** or **Mollow triplet**.

Given that the initial conditions for $\Lambda(\tau)$ are zero for the (ge) component but not for the (eg) component, it is not so convenient to use Torrey's solutions from Section 5.5.2.2, since they are best when real values are expected for $\langle\sigma_x\rangle$ and $\langle\sigma_y\rangle$. Instead, let us cast the optical Bloch equations in the form

$$\begin{aligned}
\partial_t\tilde{\rho}_{\rm eg} &= -(\gamma_\perp - i\Delta)\tilde{\rho}_{\rm eg} + i\frac{\Omega}{2}\langle\sigma_z\rangle \\
\partial_t\tilde{\rho}_{\rm ge} &= -(\gamma_\perp + i\Delta)\tilde{\rho}_{\rm ge} - i\frac{\Omega}{2}\langle\sigma_z\rangle \\
\partial_t\langle\sigma_z\rangle &= i\Omega(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}) - \Gamma\Big(\langle\sigma_z\rangle + 1\Big),
\end{aligned} \tag{5.308}$$

which is similar to the Bloch-vector form of Eqs. (5.127), but keeps the complex coherences instead of the transverse Bloch-vector components. Since we need only compute the incoherent part of the spectrum, we need only treat the fluctutation parts of the atomic operators. Thus, we wish to compute

$$\left\langle \delta\sigma^\dagger(t)\delta\sigma(t+\tau)\right\rangle_{t\to\infty}, \tag{5.309}$$

where

$$\delta\sigma = \sigma - \langle\sigma\rangle_{t\to\infty}. \tag{5.310}$$

When we subtract off the steady-state components of the Bloch equations (5.308), we obtain the matrix form

$$\partial_t\begin{bmatrix}\delta\tilde{\rho}_{\rm eg}\\ \delta\tilde{\rho}_{\rm ge}\\ \langle\delta\sigma_z\rangle\end{bmatrix} = \begin{bmatrix}-\frac{\Gamma}{2}+i\Delta & 0 & i\frac{\Omega}{2}\\ 0 & -\frac{\Gamma}{2}-i\Delta & -i\frac{\Omega}{2}\\ i\Omega & -i\Omega & -\Gamma\end{bmatrix}\begin{bmatrix}\delta\tilde{\rho}_{\rm eg}\\ \delta\tilde{\rho}_{\rm ge}\\ \langle\delta\sigma_z\rangle\end{bmatrix} =: \mathbf{P}\begin{bmatrix}\delta\tilde{\rho}_{\rm eg}\\ \delta\tilde{\rho}_{\rm ge}\\ \langle\delta\sigma_z\rangle\end{bmatrix}, \tag{5.311}$$

where we are considering only the case of homogeneous broadening ($\gamma_\perp = \Gamma/2$). Note that this is simpler than the form (5.127) that we used for Torrey's solutions, since there is no extra constant component. We thus have a purely linear system to solve.

We then need to work out a modified form of the quantum regression theorem, since we want to compute the fluctuation part of the correlation function. From Eq. (5.302), the initial conditions are given by subtracting the steady-state values from the previous initial condition:

$$\delta\Lambda(0) = \tilde{\rho}(t\longrightarrow\infty)\delta\sigma^\dagger. \tag{5.312}$$

[21] B. R. Mollow, "Power Spectrum of Light Scattered by Two-Level Systems," *Physical Review* **188**, 1969 (1969) (doi: 10.1103/PhysRev.188.1969); B. R. Mollow, "Absorption and Emission Line-Shape Functions for Driven Atoms," *Physical Review A*, **5**, 1522 (1972) (doi: 10.1103/PhysRevA.5.1522); B. R. Mollow, "Stimulated Emission and Absorption near Resonance for Driven Systems," *Physical Review A*, **5**, 2217 (1972) (doi: 10.1103/PhysRevA.5.2217). The calculation presented here is similar to that of Howard Carmichael, *An Open System Approach to Quantum Optics* (Springer-Verlag, 1993), section 3.3. For the experimental observation of the Mollow triplet, see F. Schuda, C. R. Stroud, Jr., and M. Hercher, "Observation of the resonant Stark effect at optical frequencies," *Journal of Physics B: Atomic and Molecular Physics* **7**, L198 (1974) (doi: 10.1088/0022-3700/7/7/002); F. Y. Wu, R. E. Grove, and S. Ezekiel, "Investigation of the Spectrum of Resonance Fluorescence Induced by a Monochromatic Field," *Physical Review Letters* **35**, 1426 (1975) (doi: 10.1103/PhysRevLett.35.1426); R. E. Grove, F. Y. Wu, and S. Ezekiel, "Measurement of the spectrum of resonance fluorescence from a two-level atom in an intense monochromatic field," *Physical Review A* **15**, 227 (1977) (doi: 10.1103/PhysRevA.15.227); W. Hartig, W. Rassmussen, R. Schieder, and H. Walther, "Study of the frequency distribution of the fluorescent light induced by monochromatic radiation," *Zeitschrift für Physik A* **278**, 205 (1976) (doi: 10.1007/BF01409169); J. L. Carlsten, A. Szöke, and M. G. Raymer, "Collisional redistribution and saturation of near-resonance scattered light," *Physical Review A* **15**, 1029 (1977) (doi: 10.1103/PhysRevA.15.1029); and J. Hunnekens and A. Gallagher, "Self-broadening of the sodium resonance lines and excitation transfer between the $3P_{3/2}$ and $3P_{1/2}$ levels," *Physical Review A* **27**, 1851 (1983) (doi: 10.1103/PhysRevA.27.1851).

In component form, this reads

$$\delta\Lambda_{\alpha\beta}(0) = \delta_{\beta \mathrm{g}}\tilde{\rho}_{\alpha \mathrm{e}}(t \longrightarrow \infty) - \tilde{\rho}_{\alpha\beta}(t \longrightarrow \infty)\tilde{\rho}_{\mathrm{ge}}(t \longrightarrow \infty), \tag{5.313}$$

which we can write out even more explicitly as

$$\delta\Lambda(0) = \left[\begin{array}{c} \rho_{\mathrm{ee}} - \tilde{\rho}_{\mathrm{eg}}\tilde{\rho}_{\mathrm{ge}} \\ -\tilde{\rho}_{\mathrm{ge}}\tilde{\rho}_{\mathrm{ge}} \\ -\rho_{\mathrm{ee}}\tilde{\rho}_{\mathrm{ge}} - \tilde{\rho}_{\mathrm{ge}} + \rho_{\mathrm{gg}}\tilde{\rho}_{\mathrm{ge}} \end{array} \right]_{t\to\infty} = \left[\begin{array}{c} \rho_{\mathrm{ee}} - |\tilde{\rho}_{\mathrm{eg}}|^2 \\ -\tilde{\rho}_{\mathrm{ge}}^2 \\ -\tilde{\rho}_{\mathrm{ge}}\left(1 + \langle\sigma_z\rangle\right) \end{array} \right]_{t\to\infty}, \tag{5.314}$$

where all the atomic expectation values here refer to the steady-state values.

If we now restrict our attention to the resonant ($\Delta = 0$) case, we have the steady-state values

$$\begin{aligned} \rho_{\mathrm{ee}}(t \longrightarrow \infty) &= \frac{1}{2}\frac{s}{1+s} \\ \langle\sigma_z(t \longrightarrow \infty)\rangle &= 2\rho_{\mathrm{ee}}(t \longrightarrow \infty) - 1 = -\frac{1}{1+s} \\ \tilde{\rho}_{\mathrm{ge}}(t \longrightarrow \infty) &= i\sqrt{\frac{s}{2}}\frac{1}{1+s} \\ [\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty)]^2 &= -\frac{1}{2}\frac{s}{(1+s)^2} \\ |\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty)|^2 &= \frac{1}{2}\frac{s}{(1+s)^2}, \end{aligned} \tag{5.315}$$

where

$$s = 2\frac{\Omega^2}{\Gamma^2}. \tag{5.316}$$

Thus, we can write the initial conditions as

$$\delta\Lambda(0) = \frac{1}{2}\frac{s}{(1+s)^2}\left[\begin{array}{c} s \\ 1 \\ -i\sqrt{2s} \end{array} \right]. \tag{5.317}$$

We also note that the on-resonance evolution matrix

$$\mathbf{P} = \begin{bmatrix} -\dfrac{\Gamma}{2} & 0 & i\dfrac{\Omega}{2} \\ 0 & -\dfrac{\Gamma}{2} & -i\dfrac{\Omega}{2} \\ i\Omega & -i\Omega & -\Gamma \end{bmatrix} \tag{5.318}$$

has eigenvalues

$$-\frac{\Gamma}{2}, \qquad -\frac{3\Gamma}{4} \pm i\Omega_\Gamma, \tag{5.319}$$

where as before

$$\Omega_\Gamma^2 := \Omega^2 - \left(\frac{\Gamma}{4}\right)^2. \tag{5.320}$$

The corresponding eigenvectors are

$$\left[\begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right], \qquad \left[\begin{array}{c} -\dfrac{1}{2\Omega}\left(i\dfrac{\Gamma}{4} \mp \Omega_\Gamma\right) \\ \dfrac{1}{2\Omega}\left(i\dfrac{\Gamma}{4} \mp \Omega_\Gamma\right) \\ 1 \end{array} \right]. \tag{5.321}$$

If we write these as the columns of the matrix

$$\mathbf{S} = \begin{bmatrix} 1 & -\frac{1}{2\Omega}\left(i\frac{\Gamma}{4} - \Omega_\Gamma\right) & -\frac{1}{2\Omega}\left(i\frac{\Gamma}{4} + \Omega_\Gamma\right) \\ 1 & \frac{1}{2\Omega}\left(i\frac{\Gamma}{4} - \Omega_\Gamma\right) & \frac{1}{2\Omega}\left(i\frac{\Gamma}{4} + \Omega_\Gamma\right) \\ 0 & 1 & 1 \end{bmatrix}, \tag{5.322}$$

then this matrix diagonalizes the evolution matrix:

$$\mathbf{S}^{-1}\mathbf{P}\mathbf{S} = \mathbf{D} = \begin{bmatrix} -\frac{\Gamma}{2} & 0 & 0 \\ 0 & -\frac{3\Gamma}{4} + i\Omega_\Gamma & 0 \\ 0 & 0 & -\frac{3\Gamma}{4} - i\Omega_\Gamma \end{bmatrix}. \tag{5.323}$$

Now that we have an evolution equation of the form

$$\partial_\tau \delta\Lambda(\tau) = \mathbf{P}\,\delta\Lambda(\tau), \tag{5.324}$$

we can use $\mathbf{P} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$ to obtain the solution

$$\delta\Lambda(\tau) = \exp(\mathbf{P}\tau)\delta\Lambda(0) = \mathbf{S}\exp(\mathbf{D}\tau)\mathbf{S}^{-1}\delta\Lambda(0). \tag{5.325}$$

The element of the operator Λ that we need is

$$\delta\Lambda_{\rm eg}(\tau) = \left\langle \delta\sigma^\dagger(t)\,\delta\sigma(t+\tau)\right\rangle_{t\to\infty}, \tag{5.326}$$

which is equivalent to the desired correlation function according to the quantum regression theorem. In the vector form of the solution (5.325), this is simply the first (topmost) component in the ordering we have used here. After multiplying everything out (a symbolic algebra package helps a great deal here), we obtain the result

$$\begin{aligned}\delta\Lambda_{\rm eg}(\tau) = \left\langle \delta\sigma^\dagger(0)\,\delta\sigma(\tau)\right\rangle = {}& \frac{s}{4(1+s)}e^{-(\Gamma/2)\tau} \\ &+ \frac{s}{8(1+s)^2}\left[s - 1 - i\frac{\Gamma}{4\Omega_\Gamma}(5s-1)\right]e^{-(3\Gamma/4)\tau}e^{i\Omega_\Gamma\tau} \\ &+ \frac{s}{8(1+s)^2}\left[s - 1 + i\frac{\Gamma}{4\Omega_\Gamma}(5s-1)\right]e^{-(3\Gamma/4)\tau}e^{-i\Omega_\Gamma\tau}.\end{aligned}$$

(atomic dipole correlation function) (5.327)

We can see that there are three components here: the first is a simple damped exponential and thus corresponds to a Lorentzian of width Γ (FWHM) centered on the resonant frequency; the other two are Lorentzian peaks shifted by $\pm\Omega_\Gamma$ from the resonance frequency (if Ω is large enough that Ω_Γ is real), each of width $3\Gamma/2$ (FWHM). In the weak-field limit, all component are centered at the resonance frequency, since Ω_Γ is imaginary. In the strong-field limit ($\Omega \gg \Gamma$ or $s \gg 1$), this correlation function reduces to

$$\left\langle \delta\sigma^\dagger(0)\,\delta\sigma(\tau)\right\rangle = \frac{1}{4}e^{-(\Gamma/2)\tau} + \frac{1}{8}e^{-(3\Gamma/4)\tau}e^{i\Omega_\Gamma\tau} + \frac{1}{8}e^{-(3\Gamma/4)\tau}e^{-i\Omega_\Gamma\tau}.$$

(dipole correlation, strong-field limit) (5.328)

Noting that the coherent part of the spectrum is negligible in this regime, we see that the spectrum consists of three well-separated Lorentzian peaks. Accounting for the extra factor of 2 that we pick up when calculating the one-sided spectrum, we note that the total integrated spectrum is given by the correlation function evaluated at $\tau = 0$, so we conclude that 1/2 of the total power is in the central lobe, and 1/4 of the total power is in each of the side lobes.

Computing the Fourier transform in Eq. (5.264) to find the explicit spectrum, we find [including the elastic component from Eq. (5.281)]

$$\begin{aligned} S(\omega_s) = {} & \frac{s}{(1+s)^2}\delta(\omega_s-\omega) + \frac{s}{8\pi(1+s)}\frac{\Gamma}{[(\omega_s-\omega)^2+(\Gamma/2)^2]} \\ & + \frac{s}{16\pi(1+s)^2}\frac{(s-1)(3\Gamma/2)+(5s-1)(\Gamma/2\Omega_\Gamma)[\omega_s-(\omega-\Omega_\Gamma)]}{\Big[[\omega_s-(\omega-\Omega_\Gamma)]^2+(3\Gamma/4)^2\Big]} \\ & + \frac{s}{16\pi(1+s)^2}\frac{(s-1)(3\Gamma/2)+(5s-1)(\Gamma/2\Omega_\Gamma)[\omega_s-(\omega+\Omega_\Gamma)]}{\Big[[\omega_s-(\omega+\Omega_\Gamma)]^2+(3\Gamma/4)^2\Big]} \end{aligned}$$

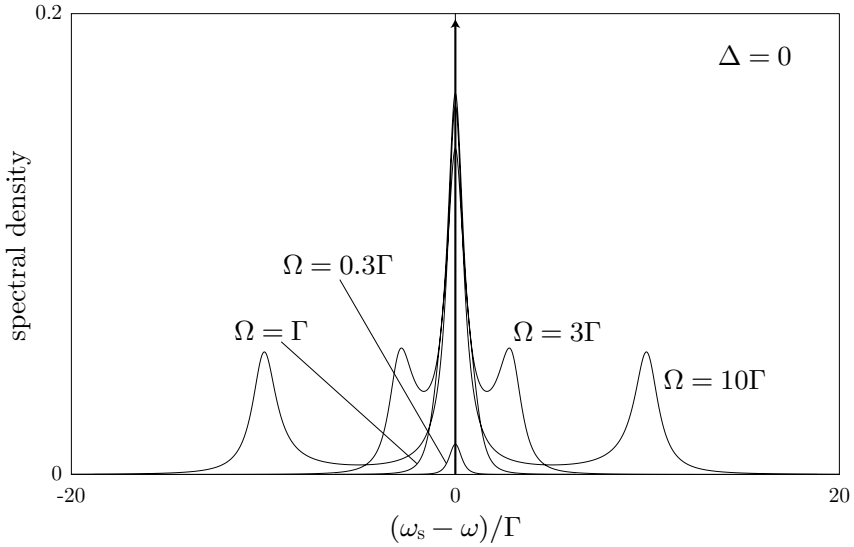
(Mollow triplet, strong-field limit) (5.329)

in the strong-field case where Ω_Γ is real ($\Omega > \Gamma/4$), and

$$\begin{aligned} S(\omega_s) = {} & \frac{s}{(1+s)^2}\delta(\omega_s-\omega) + \frac{s}{8\pi(1+s)}\frac{\Gamma}{[(\omega_s-\omega)^2+(\Gamma/2)^2]} \\ & + \frac{s}{4\pi(1+s)^2}\frac{(s-1)(3\Gamma/4+\Omega_\Gamma)}{[(\omega_s-\omega)^2+(3\Gamma/4+\Omega_\Gamma)^2]} \end{aligned}$$

(Mollow spectrum, weak-field limit) (5.330)

in the weak-field case where Ω_Γ is imaginary ($\Omega < \Gamma/4$). Several resonant spectra are plotted below.

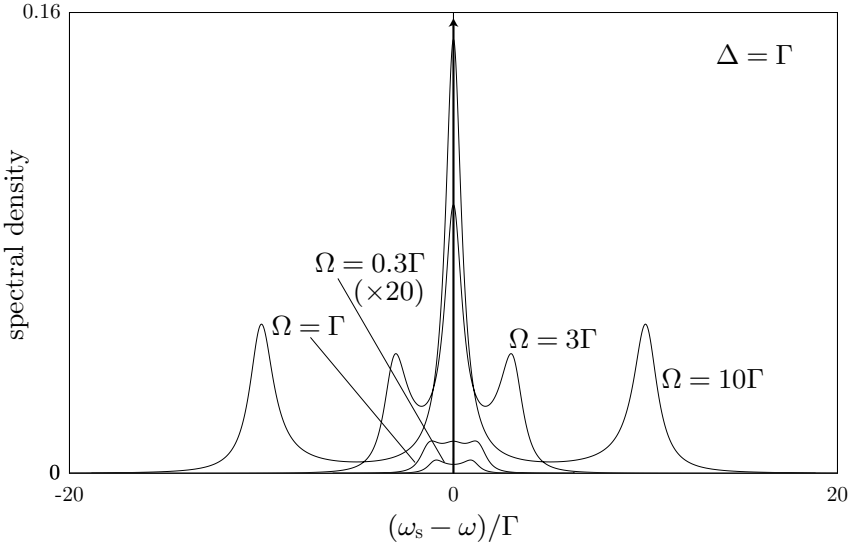


The elastic delta function of varying heights is schematically included. The spectrum is always symmetric about the laser (and hence atomic) frequency.

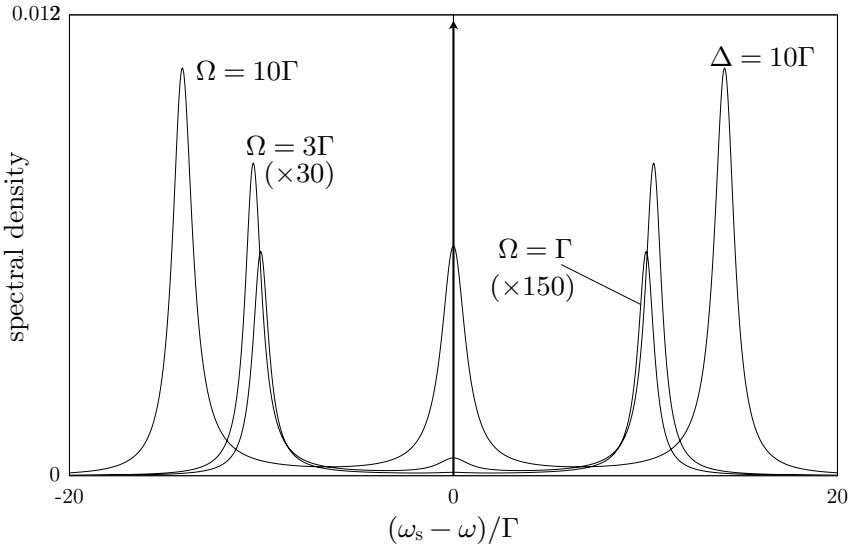
5.7.4.1 Off Resonance

In the off-resonant case, the eigenvalues do not have a simple form, and it is difficult to treat the problem analytically. However, in the case of very large detuning or driving, such that Γ is negligible, the eigenvalues of $\mathbf{P}$ are 0 and $\pm i\tilde{\Omega}$, so that the splitting reduces to the generalized Rabi frequency. In the case of large detuning and weak excitation ($\Gamma \ll \Omega \ll |\Delta|$), the splitting just reduces to Δ, and we expect a central peak at the laser frequency ω, a side peak at the atomic resonance ω_0, and the other side peak at $\omega+\Delta = 2\omega-\omega_0$.

The line shapes for detuning $\Delta = 1$ are quite different for small drive, but similar to the resonant line shapes for large drive, with the outer lobes being slightly larger and farther out.



For larger detunings, the central peak is suppressed compared to the resonant case, especially for small driving intensities. Again, the spectrum is always centered about the *laser* frequency, not the atomic resonance frequency.



5.7.4.2 Interpretations

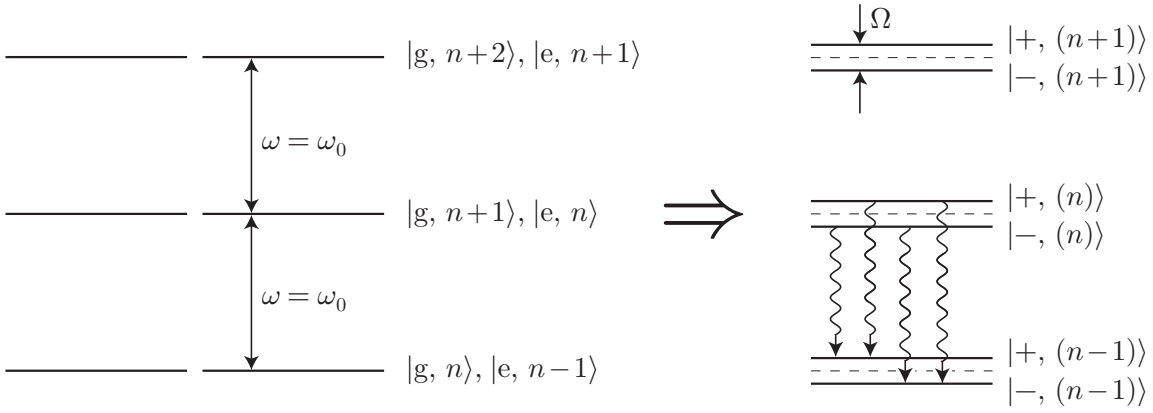
One nice physical picture of the Mollow triplet is an amplitude-modulated radiator. A source at frequency ω modulated in amplitude at frequency ω_m can be modeled as

$$\cos(\omega t)\cos^2(\omega_m t/2) = \frac{1}{2}\cos(\omega t)\left[1+\cos(\omega_m t)\right] = \frac{1}{2}\cos(\omega t) + \frac{1}{4}\cos[(\omega+\omega_m)t] + \frac{1}{4}\cos[(\omega-\omega_m)t]. \quad (5.331)$$

Amplitude modulation of a wave thus produces two sidebands with the original ''carrier,'' where the sideband splitting is just the modulation frequency. In the atom, the emission probability is proportional to ρ_{ee}, which is modulated by the Rabi oscillations, and thus we expect a similar triplet spectrum based on the modulation of the emission probability. The Mollow spectrum is thus a direct signature of Rabi oscillations. In the resonant case, the Rabi oscillations happen at the damped Rabi frequency Ω_Γ, and thus we expect the Mollow splitting to occur at this frequency. This argument extends to the far-off-resonant case, where

Rabi oscillations occur at the generalized Rabi frequency $\tilde{\Omega}$. However, this model is too simple to account for the spectral widths or relative heights of the peaks in the off-resonant case.

The other nice intepretation of the Mollow triplet comes from the dressed-atom picture. We worked out the dressed states for the atom interacting with a classical field before in Section 5.3, where we found that the new eigenstates of the combined atom-field system are split by $\tilde{\Omega}$, as opposed to the uncoupled states, which are split by the detuning Δ. Although we have not explicitly quantized the field, we have discussed the idea of photons, and so it is easy to extend the dressed-atom picture to the case of the atom coupled to the quantized field. The only difference from the previous analysis is that the two atomic levels are repeated in energy corresponding to the presence of different numbers of field quanta. The repetition occurs every $\hbar\omega$ in energy.

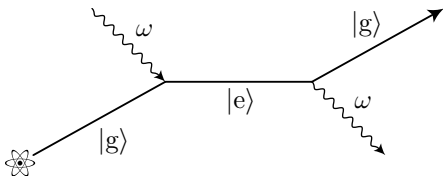


As we see, for the resonant case, the rotating-frame states $|\mathrm{g}, n+1\rangle$ and $|\mathrm{e}, n\rangle$ are degenerate, but they are coupled to each other by the field. They are thus split into a doublet, with the splitting again given by the Rabi frequency Ω. On the right-hand side of the diagram are shown four possible decay paths. Note that this analysis is only valid for $\Omega \gg \Gamma$ (otherwise the states are not resolved), and thus the decay here corresponds to the incoherent spectrum. The dressed-state splitting in this regime is likewise $\Omega \approx \Omega_\Gamma$.

We see that two possible paths give rise to the central peak, centered about the resonant frequency, while the two other decay paths are shifted in energy by $\pm\Omega$ from the resonance. Also, on resonance, each dressed state is an equal superposition of the atomic excited and ground states, and thus we expect each decay path to occur at the same rate. Thus, the central band has twice the integrated area as each of the side bands, as we have already seen in the $\Omega \gg \Gamma$ limit. Of course, in the off-resonant case, the uncoupled states are split by Δ, and the dressed states are split by $\tilde{\Omega}$, so the splittings of the Mollow triplet are just given by the generalized Rabi frequency $\tilde{\Omega}$. In addition to this simple physical picture, it is possible to compute quantitatively the general line widths and weights of the lines in the secular limit (of large driving or detuning).[22]

5.7.4.3 Energy Conservation

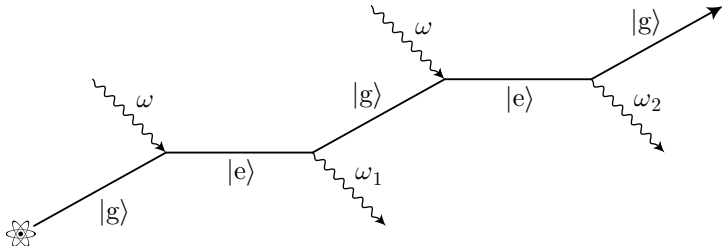
These scalings have a nice interpretation in terms of scattering processes in perturbation theory.[23] Consider the case of small saturation s (i.e., we take s to be the expansion parameter). Then the diagrammatic representation of the first-order atom-field interaction is as follows.



[22] Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Atom–Photon Interactions: Basic Processes and Applications* (Wiley, 1992), Section VI.E, p. 437.

[23] Claude Cohen-Tannoudji, "Atoms in Strong Resonant Fields," in *Frontiers in laser spectroscopy: Proceedings of the Les Houches Summer School, Session XXVII, 1975*, R. Balian, S. Haroche, and S. Liberman, Eds. (North-Holland, 1977), p. 1.

That is, we associate a factor of s with each absorption/emission cycle. Since there is only one emitted photon, and the atom ends in its initial state, the emitted photon must have the same frequency ω as the incident photon. This scattering process represents the elastic peak, the coherent delta-function spectrum from Section 5.7.2. The second-order diagram, on the other hand, is as follows.



Energy needs only to be conserved for the *entire* process here, so energy conservation imposes the requirement

$$2\omega = \omega_1 + \omega_2 \tag{5.332}$$

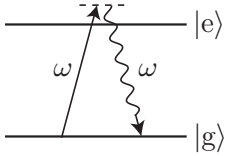
on the two emitted photons. This is how photons may be inelastically scattered, say into the Mollow side bands, while maintaining energy conservation: any inelastically scattered photon must be balanced by another one (or possibly more, in higher order) that maintains overall energy balance.

From before, we saw that for small s, the amplitude of the elastic peak scales as s, while the power of the inelastic component scales as s^2. This matches our intuition here from perturbation theory. At higher orders, perturbation theory becomes rapidly more complicated, and not so useful to consider, but we can get some powerful intuition in the weak-excitation limit, as the next example shows.

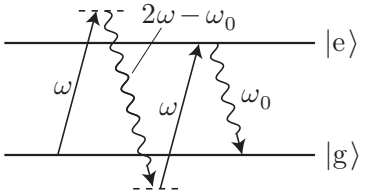
5.7.4.4 Nonclassical Correlations

The above analysis implies that for weak excitation, any photon emitted in one Mollow side band will be strongly correlated with another photon emitted in the other side band. This effect was observed in a beautiful experiment[24] on the fluorescence of a strontium beam. The atoms were excited very far off resonance (28 Å off the 460.7 nm resonance, corresponding to $\Delta/2\pi = 4$ THz, which was much larger than the Rabi frequency $\Omega/2\pi = 80$ GHz and the decay rate, which is specified in terms of the 4.7 ns lifetime). Again, the outer side bands are located approximately at ω_0 and $2\omega - \omega_0$, and photon coincidences in these two bands were monitored as a function of time delay.

The data showed a large correlation over what would be expected for random coincidences, with the peak correlation occuring for a time delay of about one lifetime. Again, this is not explained by the linear Rayleigh process,

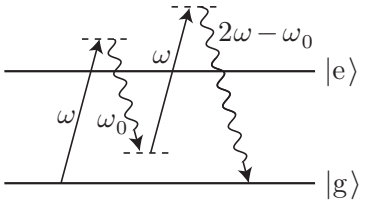


but rather by the nonlinear, second-order multiphoton process in the following diagram (shown for the blue detuning of the experiment).



[24] A. Aspect, G. Roger, S. Reynaud, J. Dalibard, and C. Cohen-Tannoudji, *Physical Review Letters* **45**, 617 (1980) (doi: 10.1103/PhysRevLett.45.617); Jean Dalibard and Serge Reynaud, ''Non-Classical Properties of Resonance Fluorescence Light,'' in *New trends in atomic physics: Proceedings of the Les Houches Summer School, Session XXXVIII, 1982*, G. Grynberg and R. Stora, Eds. (Elsevier, 1984), Seminar 2, p. 181.

Note the ordering here, as the photon of frequency ω_0 nearly always follows the emission of the photon of frequency $2\omega-\omega_0$. This is because the diagram for emission in the reverse order does not have an intermediate resonance with the excited state, as shown here.



Thus, the reverse process proceeds at a much smaller rate. (By contrast, for a resonantly driven atom, the photons in the two side bands can come in either order.) It is important to realize that this effect is manifestly quantum: it represents a correlation between the photons in two frequency modes, while no such correlation exists between photons in the same mode. The atom thus emits nonclassically correlated pairs of photons into the two frequency bands.

5.7.5 Antibunching of Resonance Fluorescence

The elastic part of the spectrum is first-order coherent, because it is monochromatic. It turns out not to be coherent at second order, as we will now show, and in fact it turns out to be **antibunched**.[25] To see this antibunching, we will need to examine the second-order coherence function. From our discussion of coherence in Section 2.6, Eq. (2.68), the normalized, second-order coherence function is

$$g^{(2)}(\tau) := \frac{\langle E^{(-)}(t)E^{(-)}(t+\tau)E^{(+)}(t+\tau)E^{(+)}(t)\rangle}{\langle E^{(-)}(t)E^{(+)}(t)\rangle^2}. \tag{5.333}$$

Again, we can replace the scattered-field operators with the atomic operators, with the result

$$g^{(2)}(\tau) := \frac{\langle \sigma^\dagger(t)\sigma^\dagger(t+\tau)\sigma(t+\tau)\sigma(t)\rangle}{\langle \sigma^\dagger\sigma\rangle^2}. \tag{5.334}$$

Recall that from our classical calculation, $g^{(2)}(\tau=0)\geq 1$ [Eq. (2.72)]. We will work with the unnormalized form

$$G^{(2)}(\tau) = \langle \sigma^\dagger(t)\sigma^\dagger(t+\tau)\sigma(t+\tau)\sigma(t)\rangle, \tag{5.335}$$

where we can see where things will become nonclassical: at $\tau=0$,

$$G^{(2)}(0) = \langle \sigma^\dagger\sigma^\dagger\sigma\sigma\rangle = 0, \tag{5.336}$$

since $\sigma^2\psi = 0$. This is impossible for a *classical* field, so resonance fluorescence from a two-level atom is *manifestly quantum*. The correlation function vanishes at $\tau=0$ because just after a photon is detected, the atom is known to be in the ground state, and cannot emit again until a time of order $2\pi/\Omega$ elapses. In other words, two photons cannot be detected simultaneously in the resonance fluorescence of a single atom.

[25] For the first theoretical proposals, see H. J. Carmichael and D. F. Walls, "Proposal for the measurement of the resonant Stark effect by photon correlation techniques," *Journal of Physics B: Atomic and Molecular Physics* **9**, L43 (1976) (doi: 10.1088/0022-3700/9/4/00); H. J. Kimble and L. Mandel, "Theory of resonance fluorescence," *Physical Review A* **13**, 2123 (1976) (doi: 10.1103/PhysRevA.13.2123); H. J. Carmichael and D. F. Walls, "A quantum-mechanical master equation treatment of the dynamical Stark effect," *Journal of Physics B: Atomic and Molecular Physics* **9**, 1199 (1976) (doi: 10.1088/0022-3700/9/8/007). For the experimental observation, see H. J. Kimble, M. Dagenais, and L. Mandel, "Photon Antibunching in Resonance Fluorescence," *Physical Review Letters* **39**, 691 (1977) (doi: 10.1103/PhysRevLett.39.691); M. Dagenais and L. Mandel, "Investigation of two-time correlations in photon emissions from a single atom," *Physical Review A* **18**, 2217 (1978) (doi: 10.1103/PhysRevA.18.2217).

To evaluate the correlation function, we will use a variation on the quantum regression theorem. Following along the lines of the above derivation,

$$\begin{aligned} G^{(2)}(\tau) &= \mathrm{Tr}\left[\sigma^\dagger(t)\sigma^\dagger(t+\tau)\sigma(t+\tau)\sigma(t)\rho_{\mathrm{SR}}\right] \\ &= \mathrm{Tr}\left[\sigma^\dagger U^\dagger(t+\tau,t)\sigma^\dagger\sigma U(t+\tau,t)\sigma\rho_{\mathrm{SR}}(t)\right] \\ &= \mathrm{Tr}\left[\sigma^\dagger\sigma U(t+\tau,t)\sigma\rho_{\mathrm{SR}}(t)\sigma^\dagger U^\dagger(t+\tau,t)\right] \\ &= \mathrm{Tr}_{\mathrm{S}}\left[\sigma^\dagger\sigma\Lambda(t+\tau,t)\right], \end{aligned} \tag{5.337}$$

where

$$\Lambda(t+\tau,t) = \mathrm{Tr}_{\mathrm{R}}\left[U(t+\tau,t)\sigma\rho_{\mathrm{SR}}(t)\sigma^\dagger U^\dagger(t+\tau,t)\right]. \tag{5.338}$$

Then in the $t \longrightarrow \infty$ limit, $\Lambda(\tau)$ satisfies the optical Bloch equations

$$\partial_\tau\Lambda(\tau) = \mathcal{L}\Lambda(\tau) \tag{5.339}$$

with initial condition

$$\Lambda(0) = \mathrm{Tr}_{\mathrm{R}}\left[\sigma\rho_{\mathrm{SR}}(t\longrightarrow\infty)\sigma^\dagger\right] = \sigma\rho(t\longrightarrow\infty)\sigma^\dagger. \tag{5.340}$$

In components, this means that $\Lambda_{\mathrm{eg}}(0) = \Lambda_{\mathrm{ge}}(0) = \Lambda_{\mathrm{ee}}(0) = 0$, while $\Lambda_{\mathrm{gg}}(0) = \rho_{\mathrm{ee}}(t\longrightarrow\infty)$. The component of the solution that we want is,

$$G^{(2)}(\tau) = \Lambda_{\mathrm{ee}}(\tau), \tag{5.341}$$

the excited-state ''population.''

Since the optical Bloch equations in terms of $\tilde{\rho}_{\alpha\beta}$ form a linear system, any scalar multiple of this solution is also a solution. In particular, rescaling the solution by $\rho_{\mathrm{ee}}(t\longrightarrow\infty)$ matches the initial conditions for the quantum regression theorem, and thus

$$G^{(2)}(\tau) = \rho_{\mathrm{ee}}(\tau)\Big|_{\rho(0)=|\mathrm{g}\rangle\langle\mathrm{g}|}\rho_{\mathrm{ee}}(t\longrightarrow\infty), \tag{5.342}$$

where $\rho_{\mathrm{ee}}(\tau)|_{\rho(0)=|\mathrm{g}\rangle\langle\mathrm{g}|}$ is the solution to the optical Bloch equations subject to the condition that the atom is intially in the ground state. Thus, for arbitrary excitation, the correlation function $G^{(2)}(\tau)$ starts off at zero, and for sufficiently large excitation shows damped oscillations towards steady state.

We already know the solution of the on-resonance optical Bloch equations in the case where $\rho_{\mathrm{gg}}(0) = 1$ and all other components are initially zero from Eq. (5.176):

$$\rho_{\mathrm{ee}}(t) = \frac{\langle\sigma_z(t)\rangle + 1}{2} = \frac{\Omega^2/2}{\Omega^2+\Gamma^2/2}\left[1 - e^{-(3\Gamma/4)t}\left(\cos\Omega_\Gamma t + \frac{3\Gamma}{4\Omega_\Gamma}\sin\Omega_\Gamma t\right)\right]. \tag{5.343}$$

Thus, on resonance, the correlation function is

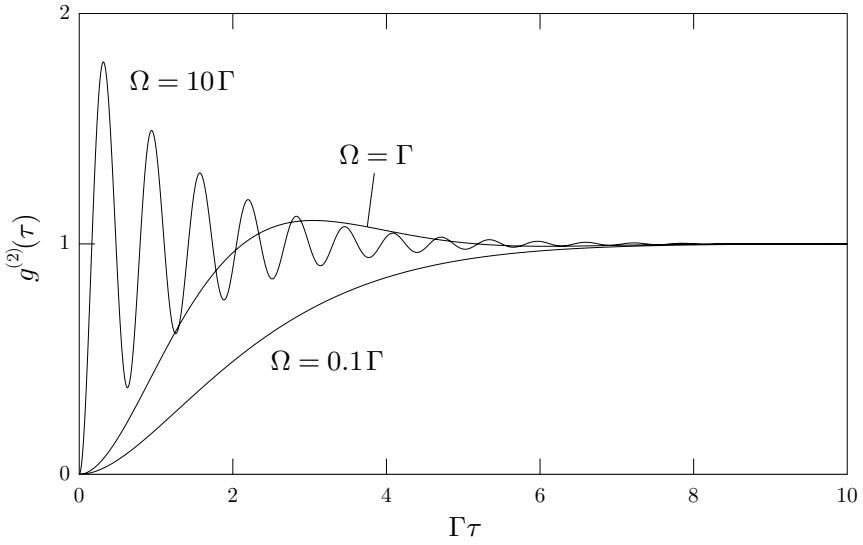
$$G^{(2)}(\tau) = \left(\frac{\Omega^2/2}{\Omega^2+\Gamma^2/2}\right)^2\left[1 - e^{-(3\Gamma/4)\tau}\left(\cos\Omega_\Gamma\tau + \frac{3\Gamma}{4\Omega_\Gamma}\sin\Omega_\Gamma\tau\right)\right]. \tag{5.344}$$

We can find the normalized correlation function by dividing by $\rho_{\mathrm{ee}}^2(t\longrightarrow\infty)$ (or by requiring that the correlation function settle to unity), with the result

$$g^{(2)}(\tau) = \left[1 - e^{-(3\Gamma/4)\tau}\left(\cos\Omega_\Gamma\tau + \frac{3\Gamma}{4\Omega_\Gamma}\sin\Omega_\Gamma\tau\right)\right]. \tag{5.345}$$

(second-order coherence for resonance fluorescence)

Again, for small excitation, the correlation function decays smoothly to unity. For larger excitations, the behavior is similar, but with Rabi oscillations along the way.



5.7.6 Probe Absorption

Some interesting features come up when we consider the effect of a driven, two-level atom on a second, weak probe field. Namely, we will now derive the **probe-absorption spectrum** for the auxiliary probe field.[26] Following Mollow's argument, we start by computing the lowest-order perturbation to the atomic density operator due to the coupling to the probe field. Let $H_{\rm p}(t)$ denote the atom–probe coupling Hamiltonian. Then in time-dependent perturbation theory in the interaction picture, the perturbation to the state is

$$\delta\rho(t) = -\frac{i}{\hbar}\int_{-\infty}^{t} dt'[H_{\rm p}(t'),\rho]. \tag{5.346}$$

Here, ρ is the Heisenberg-picture density operator with respect to the probe-free evolution, and thus has no time dependence. The effect of the Rabi oscillations must be incorporated into the time-dependence of $H_{\rm p}(t)$. The rate at which energy is absorbed from the probe is the rate at which the probe field does work on the system, which is given by

$$\begin{aligned} P_{\rm abs}(t) &= \langle \partial_t H_{\rm p}(t)\rangle \\ &= {\rm Tr}\left[\partial_t H_{\rm p}(t)\,\delta\rho(t)\right] \\ &= -\frac{i}{\hbar}\int_{-\infty}^{t} dt'\,{\rm Tr}\left\{[\partial_t H_{\rm p}(t), H_{\rm p}(t')]\,\rho\right\}. \end{aligned} \tag{5.347}$$

Note that there is no contribution from the unperturbed state ρ, since it does not include the effect of the probe. Assuming a monochromatic probe of frequency $\omega_{\rm p}$, we can write the atom–probe coupling Hamiltonian in the usual form

$$H_{\rm p}(t) = \frac{\hbar\Omega_{\rm p}}{2}\left(\sigma(t)e^{i\omega_{\rm p}t} + \sigma^\dagger(t)e^{-i\omega_{\rm p}t}\right), \tag{5.348}$$

[26] B. R. Mollow, *op. cit.*; see also N. Bloembergen and Y. R. Shen, ''Quantum-Theoretical Comparison of Nonlinear Susceptibilities in Parametric Media, Lasers, and Raman Lasers,'' *Physical Review* **133**, A37 (1964) (doi: 10.1103/PhysRev.133.A37); Murray Sargent III, ''Spectroscopic Techniques Based on Lamb's Laser Theory,'' *Physics Reports (Section C of Physics Letters)* **43**, 223 (1978) (doi: 10.1016/0370-1573(78)90163-1); Donald J. Harter and Robert W. Boyd, ''Nearly Degenerate Four-Wave Mixing Enhanced by the ac Stark Effect,'' *IEEE Journal of Quantum Electronics*, **QE-16**, 1126 (1980); Robert W. Boyd, Michael G. Raymer, Paul Narum, and Donald J. Harter, ''Four-wave parametric interactions in a strongly driven two-level system,'' *Physical Review A* **24**, 411 (1981) (doi: 10.1103/PhysRevA.24.411). For the experimental observation of the Mollow absorption spectra, see F. Y. Wu, S. Ezekiel, M. Ducloy, and B. R. Mollow, ''Observation of Amplification in a Strongly Driven Two-Level Atomic System at Optical Frequencies,'' *Physical Review Letters* **38**, 1077 (1977) (doi: 10.1103/PhysRevLett.38.1077).

where the probe Rabi frequency is given by

$$\Omega_{\rm p} := -\frac{\langle {\rm g}|\hat{\varepsilon}\cdot\mathbf{d}|{\rm e}\rangle E_{0{\rm p}}}{\hbar}, \tag{5.349}$$

where $E_{0{\rm p}}$ is the real amplitude of the probe field. Here, $\sigma(t)$ is the Heisenberg-picture (with respect to the probeless atom) atomic lowering operator, which we assume to be slowly varying on the scale of $\omega_{\rm p}$ if the pump and probe are nearly resonant. Then we can evaluate the derivative

$$\partial_t H_{\rm p}(t) \approx \frac{i\hbar\omega_{\rm p}\Omega_{\rm p}}{2}\left(\sigma(t)e^{i\omega_{\rm p}t} - \sigma^\dagger(t)e^{-i\omega_{\rm p}t}\right). \tag{5.350}$$

Putting these Hamiltonian expressions into Eq. (5.347), we find

$$P_{\rm abs}(t;\omega_{\rm p}) = \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{t} dt'\,{\rm Tr}\left\{\left[\sigma(t)\sigma^\dagger(t')e^{-i\omega_{\rm p}(t'-t)} - \sigma^\dagger(t)\sigma(t')e^{i\omega_{\rm p}(t'-t)}\right]\rho\right\} + {\rm c.c.}, \tag{5.351}$$

where we have dropped terms of the form $\sigma(t)\sigma(t')e^{i\omega_{\rm p}(t+t')}$, which will vanish under the subsequent time average. Notice that this spectrum is automatically one-sided since $\omega_{\rm p}$ is the frequency of the real probe field, which has a time dependence of the form $\cos(\omega_{\rm p}t)$. Taking $t' = t+\tau$, and assuming a stationary process so that we can drop the explicit t dependence and perform a time average,

$$\begin{aligned}
P_{\rm abs}(\omega_{\rm p}) &= \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{0} d\tau\left\langle\sigma(t)\sigma^\dagger(t+\tau)e^{-i\omega_{\rm p}\tau} - \sigma^\dagger(t)\sigma(t+\tau)e^{i\omega_{\rm p}\tau}\right\rangle + {\rm c.c.}\\
&= \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{0} d\tau\left\langle\sigma(t+\tau)\sigma^\dagger(t) - \sigma^\dagger(t)\sigma(t+\tau)\right\rangle e^{i\omega_{\rm p}\tau} + {\rm c.c.}\\
&= \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{0} d\tau\left\langle\left[\sigma(t+\tau),\sigma^\dagger(t)\right]\right\rangle e^{i\omega_{\rm p}\tau} + {\rm c.c.}\\
&= \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{0}^{\infty} d\tau\left\langle\left[\sigma(t+\tau),\sigma^\dagger(t)\right]\right\rangle e^{i\omega_{\rm p}\tau} + {\rm c.c.}\\
&= \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{\infty} d\tau\left\langle\left[\sigma(t+\tau),\sigma^\dagger(t)\right]\right\rangle e^{i\omega_{\rm p}\tau}.
\end{aligned} \tag{5.352}$$

We may thus write the absorption spectrum in terms of a new correlation function,

$$P_{\rm abs}(\omega_{\rm p}) = \frac{\hbar\omega_{\rm p}\Omega_{\rm p}^2}{4}\int_{-\infty}^{\infty} d\tau\, g_{\rm a}(\tau)e^{i\omega_{\rm p}\tau}, \tag{5.353}$$
(probe absorption)

where

$$g_{\rm a}(\tau) := \left\langle\left[\sigma(t+\tau),\sigma^\dagger(t)\right]\right\rangle. \tag{5.354}$$
(probe absorption correlation function)

The $t\longrightarrow\infty$ limit is implied here. This correlation function allows us to define the (unnormalized) probe-absorption spectrum

$$S_{\rm a}(\omega_{\rm p}) = \int_{-\infty}^{\infty} g_{\rm a}(\tau)e^{i\omega_{\rm p}\tau}d\tau. \tag{5.355}$$

[Note that there is an arbitrary but different normalization coefficient here compared to the emission spectrum of Eq. (5.264).] Again, the Heisenberg operator $\sigma(t)$ only has slow time dependence in the rotating frame of the laser field, and if we evaluate the correlation function in this frame, the spectrum becomes

$$S_{\rm a}(\omega_{\rm p}) = \int_{-\infty}^{\infty} g_{\rm a}(\tau)e^{i(\omega_{\rm p}-\omega)\tau}d\tau, \tag{5.356}$$
(probe-absorption spectrum, rotating frame)

and is implicitly centered about the laser frequency.

When we integrate over all frequencies, we find the total absorbed power,

$$\begin{aligned} P_{\rm abs} &= \int_0^\infty d\omega_{\rm p}\, P_{\rm abs}(\omega_{\rm p}) \\ &\approx \frac{\hbar\omega_0\Omega_{\rm p}^2}{4}\int_0^\infty d\omega_{\rm p}\, S_{\rm a}(\omega_{\rm p}) \\ &= \frac{\hbar\omega_0\Omega_{\rm p}^2}{4}\int_{-\infty}^\infty d\tau\, g_{\rm a}(\tau)\int_0^\infty d\omega_{\rm p}\, e^{i(\omega_{\rm p}-\omega)\tau} \\ &\approx \frac{\pi\hbar\omega_0\Omega_{\rm p}^2}{2}\int_{-\infty}^\infty d\tau\, g_{\rm a}(\tau)\delta(\tau) \\ &\approx \frac{\pi\hbar\omega_0\Omega_{\rm p}^2}{2} g_{\rm a}(0), \end{aligned} \tag{5.357}$$

where we have assumed that the absorption spectrum is peaked near the atomic resonance. The undelayed correlation function is

$$g_{\rm a}(0) = \left\langle [\sigma, \sigma^\dagger] \right\rangle = \rho_{\rm gg} - \rho_{\rm ee}, \tag{5.358}$$

and is thus related only to the population inversion. This absorbed power implies an absorption coefficient for a vapor of number density N of

$$a = \frac{P_{\rm abs} N}{I} = \frac{\pi\hbar\omega_0\Omega_{\rm p}^2}{2I}(\rho_{\rm gg} - \rho_{\rm ee})N = \frac{\pi\hbar\omega_0\Omega_{\rm p}^2}{2I}(N_{\rm g} - N_{\rm e}). \tag{5.359}$$

Noting that the saturation intensity is defined such that $\Omega_{\rm p}^2/I = \Gamma^2/2I_{\rm sat}$ [from Eq. (5.248)], where $I_{\rm sat} = \hbar\omega_0^3\Gamma/4\pi c^2$ [from Eq. (5.253)], the integrated absorption coefficient becomes

$$a = \frac{\pi^2 c^2 \Gamma}{\omega_0^2}(N_{\rm g} - N_{\rm e}). \tag{5.360}$$

This turns out to match the result we get directly from the rate equations: recall that the rate-equation absorption coefficient from Eq. (3.25) is

$$a(\omega) = \sigma(\omega)\left[N_{\rm g} - N_{\rm e}\right], \tag{5.361}$$

where the laser cross section is

$$\sigma(\omega) = \Gamma\frac{\lambda_0^2}{4} s(\omega). \tag{5.362}$$

Integration over all frequencies gives

$$\int_0^\infty d\omega\, \sigma(\omega) = \Gamma\frac{\lambda_0^2}{4} = \frac{\pi^2 c^2\Gamma}{\omega_0^2}, \tag{5.363}$$

which gives precisely the same integrated absorption coefficient. We can thus be confident that we are more or less on the right track.

The absorption correlation function can be written as

$$g_{\rm a}(\tau) = \left\langle \sigma(t+\tau)\sigma^\dagger(t)\right\rangle - \left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle =: g_{\rm d}(\tau) - g_{\rm e}(\tau), \tag{5.364}$$

where $g_{\rm e}(\tau)$ is the correlation function that we already evaluated in deriving the emission spectrum. Hence, it remains to compute the correlation function

$$g_{\rm d}(\tau) := \left\langle \sigma(t+\tau)\sigma^\dagger(t)\right\rangle. \tag{5.365}$$

The usual business of the quantum regression theorem (see Problem 5.21) tells us to solve the optical Bloch equations, with initial condition

$$\Lambda(0) = \sigma^\dagger\rho(t \longrightarrow \infty), \tag{5.366}$$

so that $\Lambda_{\rm eg}(0) = \rho_{\rm gg}(t \longrightarrow \infty)$, $\Lambda_{\rm ee}(0) = \tilde{\rho}_{\rm ge}(t \longrightarrow \infty)$, and the other components are initially zero. We can then use the component

$$g_{\rm d}(\tau) = \Lambda_{\rm eg}(\tau) \tag{5.367}$$

of the solution, whose Fourier transform then gives the probe absorption spectrum.

This problem can also be solved by explicitly including the coupling to the weak probe field in the optical Bloch equations, essentially by making the substitution

$$\Omega\sigma e^{i\omega t} \longrightarrow \Omega\sigma e^{i\omega t} + \Omega_{\rm p}\sigma e^{i\omega_{\rm p} t} \tag{5.368}$$

in the laboratory (nonrotating) frame, and then finding the (oscillating) equilibrium solution. Using this procedure, Mollow[27] gave an analytic form for the absorption spectrum (Problem 5.20):

$$S_{\rm a}(\omega_{\rm p}) = \frac{\Big[\rho_{\rm gg}(t \longrightarrow \infty) - \rho_{\rm ee}(t \longrightarrow \infty)\Big]\left[(\Gamma - i\Delta_{\rm p})[\gamma_\perp + i(\Delta - \Delta_{\rm p})] + \dfrac{i\Omega^2\Delta_{\rm p}}{2(\gamma_\perp + i\Delta)}\right]}{(\Gamma - i\Delta_{\rm p})[\gamma_\perp + i(\Delta - \Delta_{\rm p})][\gamma_\perp - i(\Delta + \Delta_{\rm p})] + \Omega^2(\gamma_\perp - i\Delta_{\rm p})} + \text{c.c.}$$

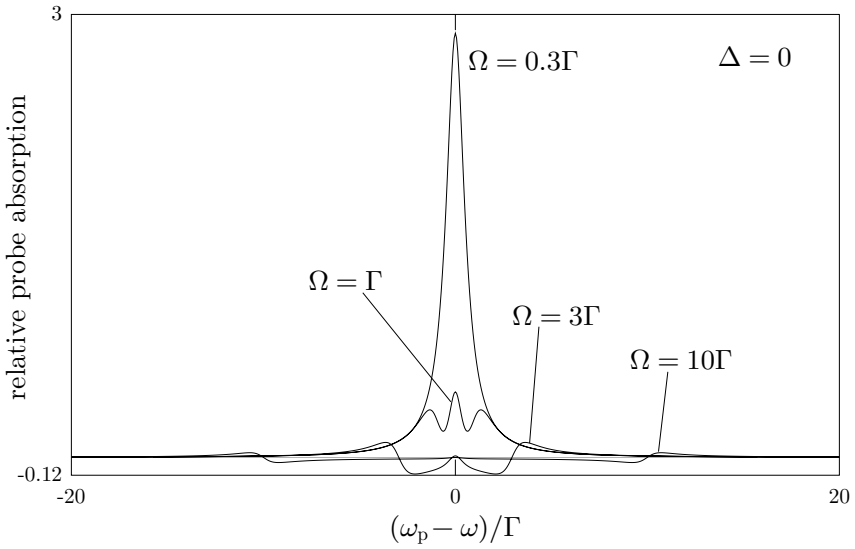
(Mollow probe-absorption spectrum) (5.369)

Here,

$$\Delta_{\rm p} := \omega_{\rm p} - \omega \tag{5.370}$$

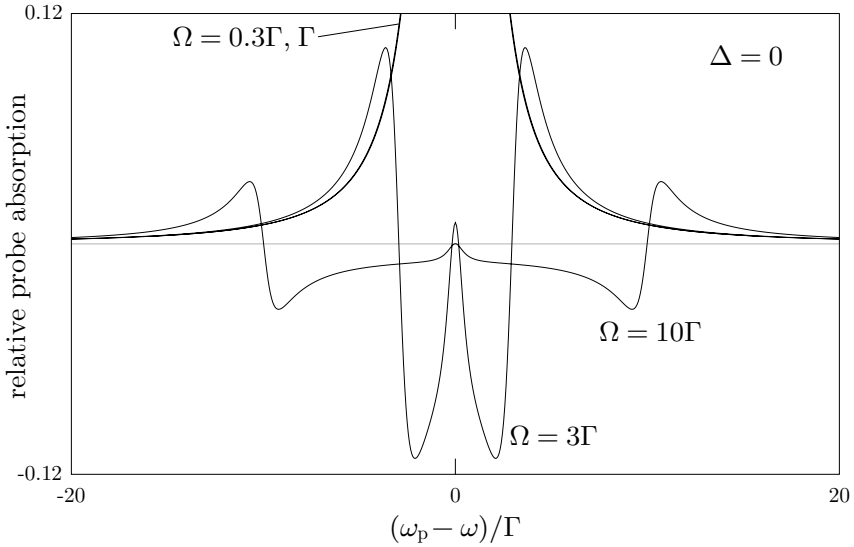
is the probe detuning from the pump frequency (so that $\Delta + \Delta_{\rm p} = \omega_{\rm p} - \omega_0$ is the probe detuning from resonance, and $\Delta - \Delta_{\rm p} = 2\omega - \omega_0 - \omega_{\rm p}$.. This spectrum is valid off resonance and for general damping (with or without collisions).

Some on-resonance absorption line shapes are plotted below in the absence of collisions.



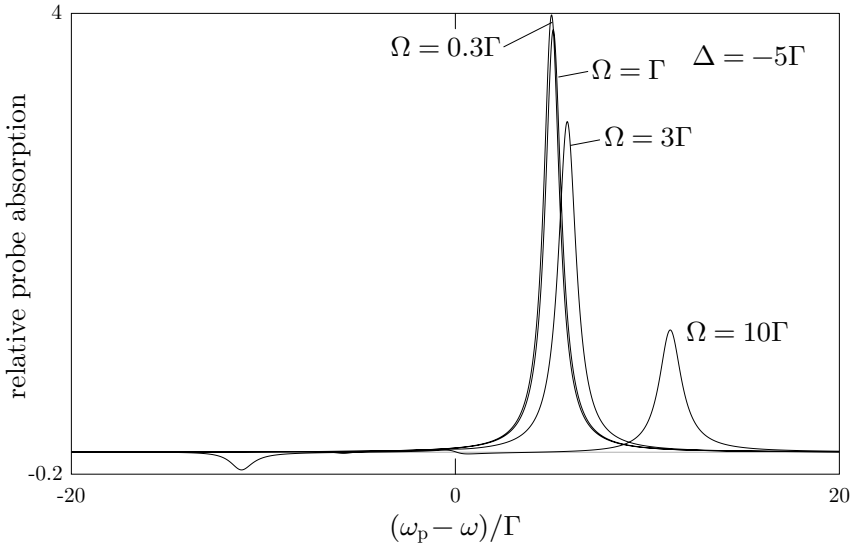
The absorption line is strongest and Lorentzian for small pump intensities. As the pump intensity gets larger, the absorption line shape becomes more complicated, showing multiple peaks. A vertically zoomed version of the same plot is shown below.

[27] B. R. Mollow, ''Stimulated Emission and Absorption near Resonance for Driven Systems,'' *Physical Review A*, **5**, 2217 (1972) (doi: 10.1103/PhysRevA.5.2217). See in particular Eqs. (3.8) and (3.11a).

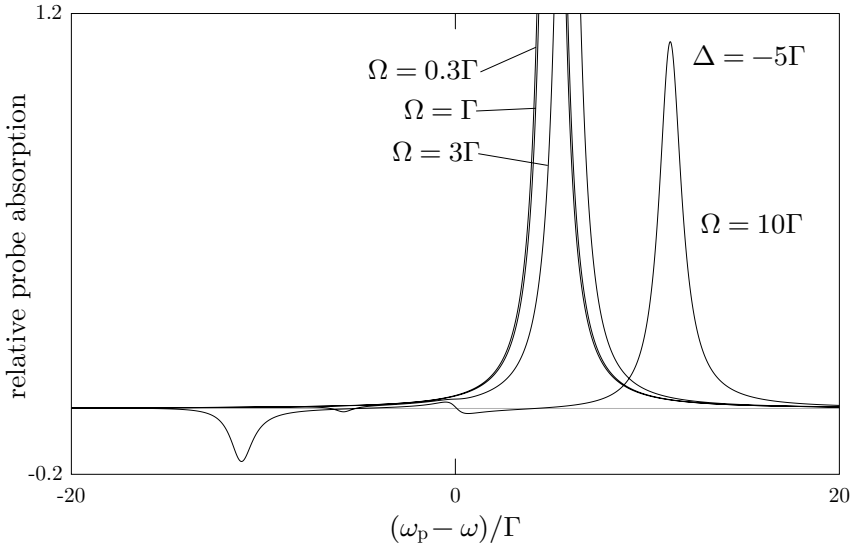


For saturating pump intensities, the absorption line shape crosses through zero and is negative for certain frequency ranges. These regions of *negative absorption* correspond, of course, to *stimulated emission* of the probe. For large pump intensities, the outer zero crossings occur at $\pm\Omega$ from the pump laser frequency. Right on resonance, the probe is absorbed, but just off of resonance stimulated emission dominates.

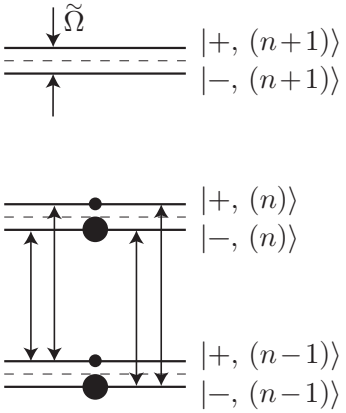
For larger detunings (and still with no collisions), the line shape is pretty much the same for small pump intensities: a single Lorentzian peak at the atomic resonance. But as the pump intensity gets larger, the line shape becomes dramatically different from the on-resonance case.



Side peaks develop near $\pm\tilde{\Omega}$ from the pump laser frequency, one positive and one negative. Also, as we can see from the zoomed version of the same plot, there is also a central dispersive structure in the line shape.



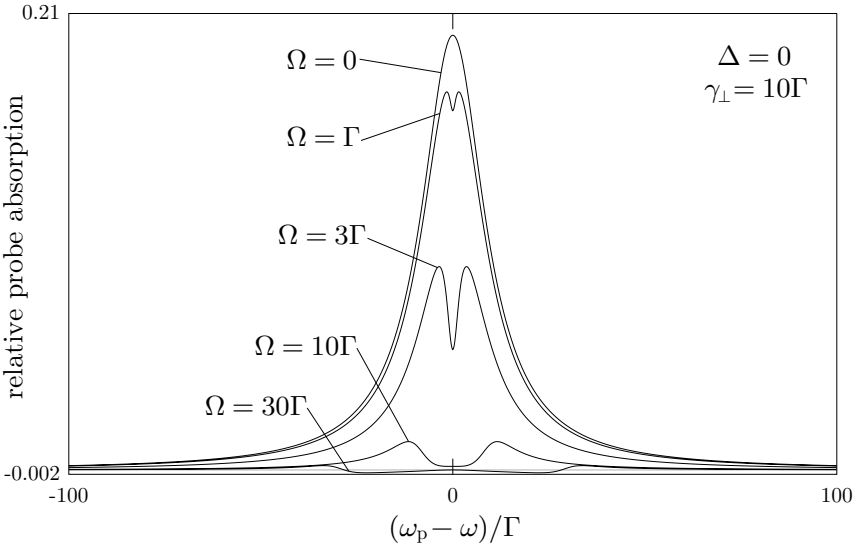
While the center structure can be understood qualitatively,[28] its interpretation is relatively subtle. The side peaks, however, are reasonably easy to understand in the dressed-state picture (which again applies if $\tilde{\Omega} \gg \Gamma$).



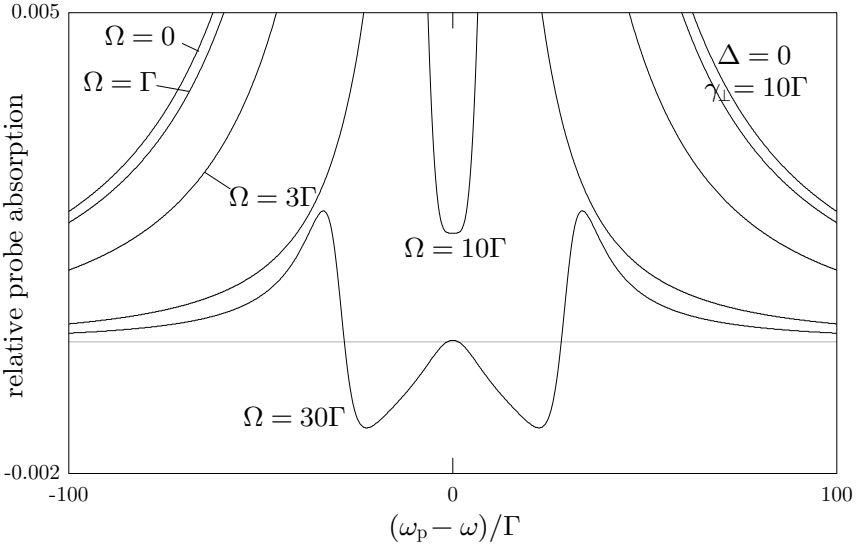
The absorption spectrum is essentially a probe of the dressed levels, which off resonance are split approximately by the generalized Rabi frequency $\tilde{\Omega}$, and hence the location of the side peaks. Recall also that for large negative detuning, the $|-\rangle$ dressed state is essentially the same as $|g\rangle$, and hence should be more populated in steady state. Then the more energetic (blue) side peak probes a transition where the ground state is more populated than the excited state, hence giving rise to an absorptive peak. The less energetic (red) side peak probes a transition *with a population inversion*, and thus gives rise to a negative-absorption peak. The central structure is due to two transitions that probe levels with no population difference, and again has a more subtle interpretation.

The absorption line shapes in the regime of strong collisional damping are also qualitatively quite different. We will show the case where $\gamma_\perp = 10\Gamma$ and the pump is on resonance ($\Delta = 0$). For small pump intensities, the line shape is a broadened Lorentzian, corresponding to the collisional line width. Interestingly, as the pump intensity is increased, a narrow dip appears in the line shape.

[28] Gilbert Grynberg and Claude Cohen-Tannoudji, "Central resonance of the Mollow absorption spectrum: physical origin of gain without population inversion," *Optics Communications* **96**, 150 (1993) (doi: 10.1016/0030-4018(93)90538-G).



As the pump intensity is increased to large values, the dip becomes a large ''hole'' in the line shape, and the line shape eventually takes on a form much like the homogeneously broadened case, with regions of stimulated emission.



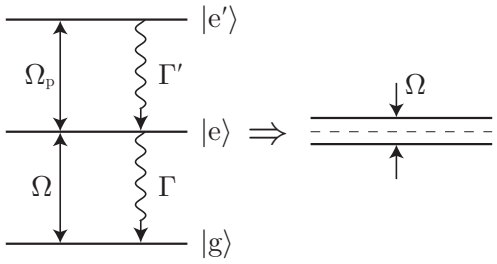
Notice that the collisional damping overall suppresses the absorption, as the absorption lines are much weaker than in the homogeneously broadened cases.

5.7.6.1 Autler–Townes Doublet

We saw above how the peaks in the emission and probe absorption spectra of the driven two-level atom can be explained in the strongly driven limit in terms of the splitting of the dressed states of the atom. The **Autler–Townes doublet**[29] is an even more direct manifestation of the dressed-state splittings.

Consider the usual two-level atom, driven by a resonant field of Rabi frequency Ω. Now consider a third, auxiliary level $|e'\rangle$, an energy $\hbar\omega_0'$ above the usual excited state $|e\rangle$. We will assume this state decays to $|e\rangle$ at a rate Γ'. We will also assume a weak probe field of frequency ω_p coupling $|e\rangle \longrightarrow |e'\rangle$ (such that $\omega_p \approx \omega_0'$) with Rabi frequency Ω_p.

[29] S. H. Autler and C. H. Townes, ''Stark Effect in Rapidly Varying Fields,'' *Physical Review* **100**, 703 (1955) (doi: 10.1103/PhysRev.100.703).



In the presence of a strong drive (large Ω), the excited state splits into a doublet of splitting Ω due to mixing with the ground state. Thus, we expect the probe-absorption spectrum to have two peaks, corresponding to resonance of $|\mathrm{e}'\rangle$ with each of the dressed states. In the limit of large Ω, we expect the absorption spectrum to be a sum of two Lorentzian peaks. However, we need a formalism for dealing with this more quantitatively. Fortunately, it is easy to extend the formalism we have already developed to handle this problem. To compute the probe-absorption spectrum, we will treat the probe perturbatively. Thus, we will again need the correlation function

$$g_{\mathrm{a}}(\tau) = \left\langle \left[\sigma'(t+\tau), \sigma'^{\dagger}(t)\right]\right\rangle, \tag{5.371}$$

where now $\sigma' = |\mathrm{e}\rangle\langle\mathrm{e}'|$ is the lowering operator for the probe transition. The first term in this correlation function is

$$g_{\mathrm{d}}(\tau) = \left\langle \sigma'(t+\tau)\sigma'^{\dagger}(t)\right\rangle. \tag{5.372}$$

Working out the quantum regression theorem, we solve the master equation for this three-level atom, with initial condition

$$\Lambda(0) = \sigma'^{\dagger}\rho(t \longrightarrow \infty), \tag{5.373}$$

which becomes in components

$$\Lambda_{\alpha\beta}(0) = \delta_{\alpha\mathrm{e}'}\rho_{\mathrm{e}\beta}(t \longrightarrow \infty). \tag{5.374}$$

Then the correlation function corresponds to the component

$$g_{\mathrm{d}}(\tau) = \mathrm{Tr}_{\mathrm{S}}[\sigma'\Lambda(\tau)] = \Lambda_{\mathrm{e}'\mathrm{e}}(\tau). \tag{5.375}$$

The other part of the absorption correlation function is the emission correlation function

$$g_{\mathrm{e}}(\tau) = \left\langle \sigma'^{\dagger}(t)\sigma'(t+\tau)\right\rangle. \tag{5.376}$$

This correlation function satisfies the atomic master equation with initial condition

$$\Lambda(0) = \rho(t \longrightarrow \infty)\sigma'^{\dagger}, \tag{5.377}$$

which becomes in components

$$\Lambda_{\alpha\beta}(0) = \delta_{\beta\mathrm{e}}\rho_{\alpha\mathrm{e}'}(t \longrightarrow \infty). \tag{5.378}$$

This the correlation function corresponds to the same component as $g_{\mathrm{d}}(\tau)$:

$$g_{\mathrm{e}}(\tau) = \mathrm{Tr}_{\mathrm{S}}[\sigma'\Lambda(\tau)] = \Lambda_{\mathrm{e}'\mathrm{e}}(\tau). \tag{5.379}$$

Note that the initial conditions for the components involving the $|\mathrm{e}'\rangle$ are $\Lambda_{\mathrm{e}'\mathrm{e}'}(0) = 0$ and $\Lambda_{\mathrm{e}'\mathrm{e}}(0) = 0$, and because there is no field coupling any state to $|\mathrm{e}'\rangle$, these components remain zero for all time. Thus, we can see that $g_{\mathrm{e}}(\tau) = 0$. This of course means that there is no spontaneous emission on the $\mathrm{e}' \longrightarrow \mathrm{e}$ transition, as we expect since it is not pumped by any strong field.

The master equation for the atom is the same as for the usual optical Bloch equations, with an extra dissipation term for the second decay channel,

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \Gamma\mathcal{D}[\sigma]\tilde{\rho} + \Gamma'\mathcal{D}[\sigma']\tilde{\rho}, \tag{5.380}$$

where the atomic Hamiltonian is

$$\tilde{H}_{\rm A} = -\hbar\Delta|{\rm e}\rangle\langle{\rm e}| - \hbar\Delta|{\rm e}'\rangle\langle{\rm e}'|, \tag{5.381}$$

in the rotating frame where the $|{\rm e}'\rangle$ state is degenerate with the $|{\rm e}\rangle$ state. The interaction is still

$$\tilde{H}_{\rm AF} = \frac{\hbar\Omega}{2}\left(\sigma + \sigma^\dagger\right), \tag{5.382}$$

since we are neglecting the effect of the probe field on the atom.

Written out as a set of coupled equations for the density-matrix elements, we find

$$\begin{aligned}
\partial_t\rho_{\rm gg} &= -i\frac{\Omega}{2}(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}) + \Gamma\rho_{\rm ee} \\
\partial_t\tilde{\rho}_{\rm ge} &= -(\Gamma/2 + i\Delta)\tilde{\rho}_{\rm ge} - i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}) \\
\partial_t\tilde{\rho}_{\rm ge'} &= -(\Gamma'/2 + i\Delta)\tilde{\rho}_{\rm ge'} - i\frac{\Omega}{2}\tilde{\rho}_{\rm ee'} \\
\partial_t\tilde{\rho}_{\rm eg} &= -(\Gamma/2 - i\Delta)\tilde{\rho}_{\rm eg} + i\frac{\Omega}{2}(\rho_{\rm ee} - \rho_{\rm gg}) \\
\partial_t\rho_{\rm ee} &= i\frac{\Omega}{2}(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}) - \Gamma\rho_{\rm ee} + \Gamma'\rho_{\rm e'e'} \\
\partial_t\tilde{\rho}_{\rm ee'} &= -[(\Gamma + \Gamma')/2]\tilde{\rho}_{\rm ee'} - i\frac{\Omega}{2}\tilde{\rho}_{\rm ge'} \\
\partial_t\tilde{\rho}_{\rm e'g} &= -(\Gamma'/2 - i\Delta)\tilde{\rho}_{\rm e'g} + i\frac{\Omega}{2}\tilde{\rho}_{\rm e'e} \\
\partial_t\tilde{\rho}_{\rm e'e} &= -[(\Gamma + \Gamma')/2]\tilde{\rho}_{\rm e'e} + i\frac{\Omega}{2}\tilde{\rho}_{\rm e'g} \\
\partial_t\rho_{\rm e'e'} &= -\Gamma'\rho_{\rm e'e'}.
\end{aligned} \tag{5.383}$$

These equations are easy to solve numerically. The normalized spectrum is then again the Fourier transform of the correlation function

$$S_{\rm a}(\omega_{\rm p}) = \int_{-\infty}^{\infty} g_{\rm a}(\tau)e^{i\omega_{\rm p}\tau}d\tau, \tag{5.384}$$

which is also readily computed numerically. However, we can actually get a closed-form expression for the spectrum: the correlation function becomes

$$\begin{aligned}
g_{\rm a}(\tau) = &\frac{\Omega\tilde{\rho}_{\rm eg}(t\longrightarrow\infty)}{2\xi}e^{-(\Gamma/4+\Gamma'/2)\tau}e^{i\Delta\tau/2}\left(e^{i\xi\tau/2} - e^{-i\xi\tau/2}\right) \\
&+\frac{\rho_{\rm ee}(t\longrightarrow\infty)}{2}e^{-(\Gamma/4+\Gamma'/2)\tau}e^{i\Delta\tau/2}\left\{\left[1 + \frac{i}{\xi}\left(\frac{\Gamma}{2} + i\Delta\right)\right]e^{i\xi\tau/2} + \left[1 - \frac{i}{\xi}\left(\frac{\Gamma}{2} + i\Delta\right)\right]e^{-i\xi\tau/2}\right\},
\end{aligned} \tag{5.385}$$

where

$$\xi = \sqrt{\Omega^2 + \Delta^2 - \frac{\Gamma^2}{4} - i\Delta\Gamma}. \tag{5.386}$$

Notice that on resonance ($\Delta = 0$), $\xi = \Omega_{2\Gamma} = \sqrt{\Omega^2 - \Gamma^2/4}$, and

$$\begin{aligned}
g_{\rm a}(\tau) = &\frac{\Omega\tilde{\rho}_{\rm eg}(t\longrightarrow\infty)}{2\Omega_{2\Gamma}}e^{-(\Gamma/4+\Gamma'/2)\tau}\left(e^{i\Omega_{2\Gamma}\tau/2} - e^{-i\Omega_{2\Gamma}\tau/2}\right) \\
&\frac{\rho_{\rm ee}(t\longrightarrow\infty)}{2}e^{-(\Gamma/4+\Gamma'/2)\tau}\left\{\left[1 + \frac{i\Gamma}{2\Omega_{2\Gamma}}\right]e^{i\Omega_{2\Gamma}\tau/2} + \left[1 - \frac{i\Gamma}{2\Omega_{2\Gamma}}\right]e^{-i\Omega_{2\Gamma}\tau/2}\right\}.
\end{aligned} \tag{5.387}$$

In this resonant case, for $\Omega^2 \geq \Gamma^2/4$, the absorption spectrum consists of something like a pair of lines of width $\Gamma/2 + \Gamma'$, split by $\Omega_{2\Gamma}$ (but with dispersive components),

$$S_\mathrm{a}(\omega_\mathrm{p}) = \frac{(\Gamma + 2\Gamma')\rho_{\mathrm{ee}}(t \longrightarrow \infty) - \dfrac{2}{\Omega_{2\Gamma}}\left\{2\Omega \mathrm{Im}[\tilde{\rho}_{\mathrm{eg}}(t \longrightarrow \infty)] + \Gamma\rho_{\mathrm{ee}}(t \longrightarrow \infty)\right\}\left(\omega_\mathrm{p} + \dfrac{\Omega_{2\Gamma}}{2}\right)}{4\left[\left(\omega_\mathrm{p} + \dfrac{\Omega_{2\Gamma}}{2}\right)^2 + \dfrac{1}{4}\left(\dfrac{\Gamma}{2} + \Gamma'\right)^2\right]} + (\Omega_{2\Gamma} \longrightarrow -\Omega_{2\Gamma}),$$

(Autler–Townes absorption spectrum) (5.388)

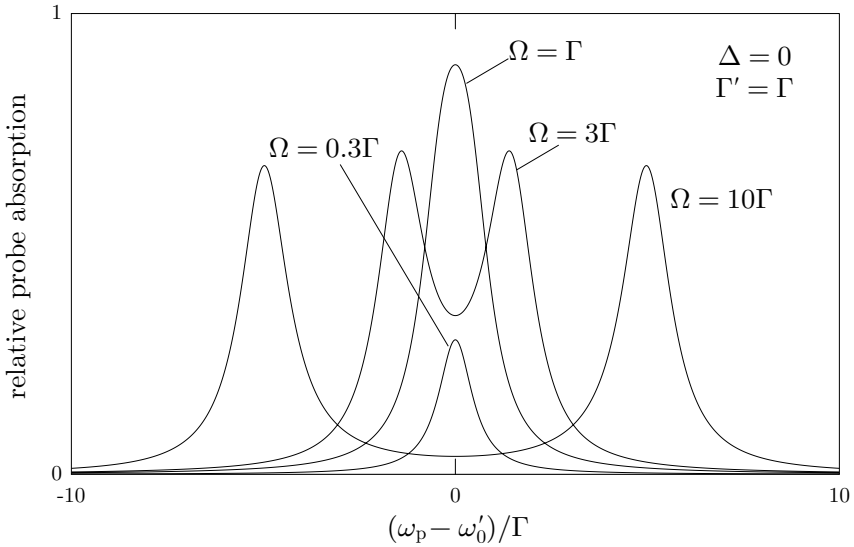
assuming Ω is large enough that $\Omega_{2\Gamma}$ is real. As the pump intensity becomes large, the dispersive components become unimportant near each resonance, and at the same time $\Omega_{2\Gamma} \longrightarrow \Omega$. The dispersive components thus cease to shift the spectral peaks in this limit—although they are still important in principle for the wings, since they fall off more slowly than the absorptive part—and the peaks are then separated by Ω, as we expected from the dressed-atom picture. Of course, we can get the absorption spectrum in the general case, but it's complicated enough that it's not very illuminating.

Why the line width of $\Gamma/2 + \Gamma'$? In general, the weak-probe absorption width of two states of total decay rates Γ_1 and Γ_2 is simply $\Gamma_1 + \Gamma_2$, because the convolution of two Lorentzians of these widths is a Lorentzian of width $\Gamma_1 + \Gamma_2$. The dressed states on resonance are equal superpositions of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, which are states of respective decay rates 0 and Γ. Thus, each dressed state should only decay at rate $\Gamma/2$, which sets the width of each state. In general, from Eq. (5.385), we can see that the line widths are given by the total exponential decaying part, and so

$$\Delta\omega = \frac{\Gamma}{2} + \Gamma' \pm \mathrm{Im}[\xi]. \tag{5.389}$$

Again, in the limit of large Ω, we have $\Delta\omega = \Gamma/2 + \Gamma'$. On the other hand, in the limit of small Ω and resonant pump ($\Delta = 0$), $\xi \approx i\Gamma/2$, and thus $\Delta\omega$ takes on the values Γ', which is what we expect for probe absorption on $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}'\rangle$, and $\Gamma + \Gamma'$, which is what we expect for probe absorption on $|\mathrm{e}\rangle \longrightarrow |\mathrm{e}'\rangle$. For weak drive ($\Omega \ll \Gamma$) and large detuning, $\mathrm{Im}[\xi] = \pm\Gamma/2$ for $\Delta \lesseqgtr 0$, which also gives the same two line widths. Interestingly, the on-resonance Autler–Townes spectrum has a minimum width for $\Omega = \Gamma/2$, when $\mathrm{Im}[\xi]$ vanishes.

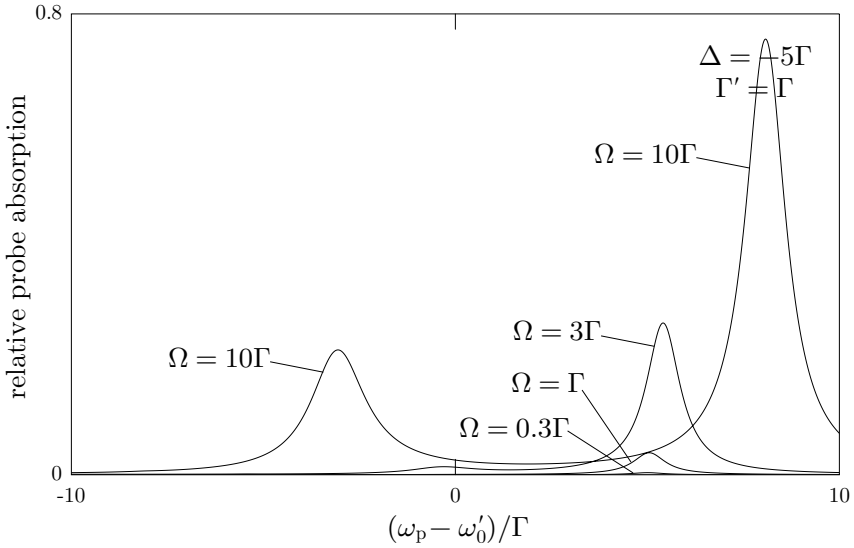
The absorption spectra for $\Gamma' = \Gamma$ for several different values of the pumping rate Ω are shown below.



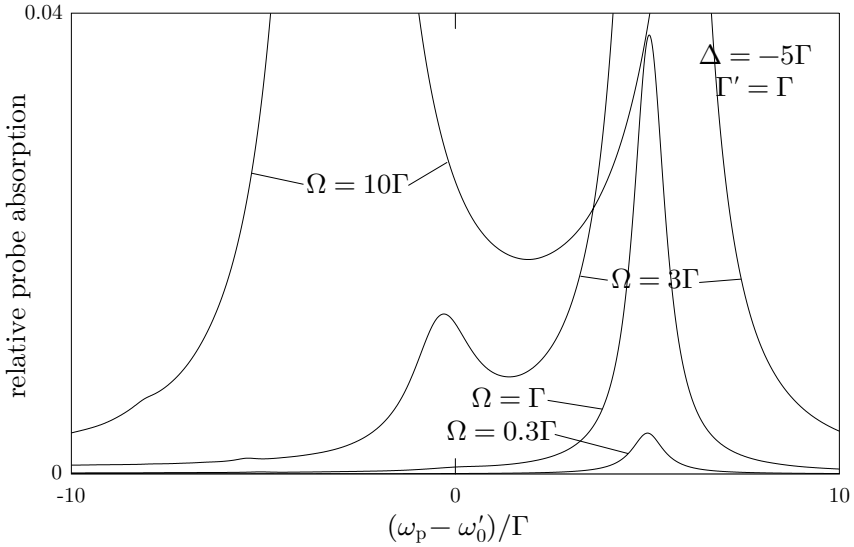
As we expected, we see two absorption lines, which become resolved as the pumping rate becomes large.

In the case where the pump field is detuned from resonance, the doublet lines have asymmetric amplitudes. The dressed states are an asymmetric superposition of the bare states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, which turns out

to be more important than the asymmetric steady-state populations of the dressed states. Also, the doublet center shifts, as we expect from the shift of the center of the bare states. It is a useful exercise to understand the placement and weights of the two lines here.



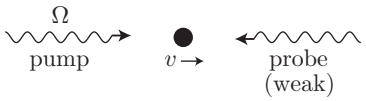
In a magnified view of the same spectrum, we can see the behavior of the cases of small pump intensity, where the displaced line is dominant, because there is little mixing of the pump-transition states.



Minor secondary absorption peaks are also barely visible in the spectra in this magnified plot.

5.7.6.2 Lamb Dip

One difficulty in precision spectroscopy of an atomic vapor is the Doppler broadening of the atomic transitions. For example, to use ^{87}Rb as an absoute frequency reference, the natural linewidth is about 5 MHz, and the center of the transition can be determined far better than this width. However, the Doppler-broadened width at room temperature is about 500 MHz, making the precision of measurements much worse (and in fact blending several hyperfine levels into a single Doppler line). Of course, one can now use laser-cooled atoms, where the Doppler effect is unimportant. However, there is a much easier trick for sub-Doppler spectroscopy: the **Lamb dip** or **saturated-absorption spectroscopy**. The basic idea is as follows. Consider a vapor of atoms where the Doppler width is larger than the natural line width. Now illuminate the atoms with *two* counterpropagating lasers.



The *pump* laser functions simply to saturate the atoms, while we monitor the absorption of the *probe* laser due to the atomic vapor. The two fields are generally produced from the same source, and thus their frequencies are swept together. There may be a frequency offset between them, but to simplify our discussion for the moment, assume they have the same frequency. Atoms at a particular velocity v experience the two fields with equal but opposite Doppler shifts, and thus a moving atom experiences them as having different frequencies. Moving atoms can thus only be resonant with at most one of the fields. Atoms at rest, however, see two fields at the same frequency, and can be resonant with both fields simultaneously. Because these atoms are effectively pumped with higher intensity than other atoms, the saturation reduces the resonant absorption coefficient. This reduction happens again only for atoms (nearly) at rest, and thus this effect is not Doppler broadened. The saturation produces a ''dip'' (the Lamb dip) in the Doppler absorption profile, centered at the atomic resonance, which has a width that can be on the order of the natural linewidth. Of course, if the pump and probe are not degenerate, the dip still occurs, but is displaced from the atomic resonance by an amount proportional to the pump-probe detuning.

To analyze this problem quantitatively, we first note that in the interest of precision spectroscopy, the pump and especially the probe will have low intensity to avoid power broadening of the transition. We will thus treat the probe field perturbatively, and only explicitly include the effect of the pump field on the atom. We can then use Mollow's formula (5.369) for the probe-absorption spectrum, writing $\rho_{\rm ee}(t \longrightarrow \infty)$ explicitly using Eq. (5.132):

$$S_{\rm a}(\omega_{\rm p}) = \left\langle \left(1 - \frac{\Omega^2/\gamma_\perp\Gamma}{1 + \Delta^2/\gamma_\perp^2 + \Omega^2/\gamma_\perp\Gamma}\right) \frac{\left[(\Gamma - i\Delta_{\rm p})[\gamma_\perp + i(\Delta - \Delta_{\rm p})] + \dfrac{i\Omega^2\Delta_{\rm p}}{2(\gamma_\perp + i\Delta)}\right]}{(\Gamma - i\Delta_{\rm p})[\gamma_\perp + i(\Delta - \Delta_{\rm p})][\gamma_\perp - i(\Delta + \Delta_{\rm p})] + \Omega^2(\gamma_\perp - i\Delta_{\rm p})} + \text{c.c.} \right\rangle_{\rm v} .$$

(saturated-absorption spectrum) (5.390)

The angle brackets denote an average over the atomic velocity distribution, and recall that $\Delta = \omega - \omega_0$ is the pump detuning from resonance, and $\Delta_{\rm p} = \omega_{\rm p} - \omega$ is the probe detuning from the pump. Both of these frequencies now depend on the atomic velocity due to the Doppler shift, and the average ''smears out'' these frequencies.

Recall that the Doppler shift of a field of wave vector $\mathbf{k}$, as seen by an atom of velocity $\mathbf{v}$, is $\delta\omega = -\mathbf{k}\cdot\mathbf{v}$. Why is this? A plane wave has the form

$$E_0 \cos(\mathbf{k}\cdot\mathbf{x} - \omega t), \tag{5.391}$$

and the position

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t dt'\, \mathbf{v}(t') \tag{5.392}$$

of the atom is time-dependent. Thus, we can write the field as

$$E_0 \cos\left[\mathbf{k}\cdot\left(\mathbf{x}_0 + \int_0^t dt'\, \mathbf{v}(t')\right) - \omega t\right] = E_0 \cos\left[\mathbf{k}\cdot\mathbf{x}_0 - \left(\omega t - \mathbf{k}\cdot\int_0^t dt'\, \mathbf{v}(t')\right)\right]. \tag{5.393}$$

In the moving frame, the effective frequency is the time derivative of the plane-wave phase:

$$\omega_{\rm v} = -\partial_t\phi = \omega - \mathbf{k}\cdot\mathbf{v}. \tag{5.394}$$

This establishes the Doppler shift, even for a time-varying velocity.

We can thus implement the Doppler shift of the pump field by the replacement

$$\Delta \longrightarrow \Delta - \mathbf{k}_{\rm pump}\cdot\mathbf{v} = \Delta - \mathbf{k}\cdot\mathbf{v}. \tag{5.395}$$

In the case of the probe field, recall that the detuning Δ_p is relative to the pump field, and thus we need to include the Doppler shift twice, once for the pump and once for the probe:

$$\Delta_\mathrm{p} \longrightarrow \Delta_\mathrm{p} + \mathbf{k}_\mathrm{pump}\cdot\mathbf{v} - \mathbf{k}_\mathrm{probe}\cdot\mathbf{v} \approx \Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v}. \tag{5.396}$$

Here, we do not distinguish between the pump and probe wave vectors (i.e., $|\Delta_\mathrm{p}| \ll \omega_0$), and we assume exactly counterpropagating pump and probe fields. Again, we will want the probe absorption as a function of the probe frequency, but the pump and probe are scanned together. Thus, Δ_p (after the velocity replacments) is a fixed parameter, while the spectrum is as a function of Δ. Also, in general, since we want the spectrum location relative to the atomic resonance, we will make the plot of $\Delta_\mathrm{p} + \Delta = \omega_\mathrm{p} - \omega_0$.

Thus, in principle we have our result in Eq. (5.390), after making the replacements (5.395) and (5.396). Unfortunately, this result is rather complicated to interpret. Thus, we will expand the spectrum to lowest order in Ω^2 as

$$S_\mathrm{a}(\omega_\mathrm{p}) = S_\mathrm{a}^{(0)}(\omega_\mathrm{p}) + S_\mathrm{a}^{(2)}(\omega_\mathrm{p})\Omega^2 + \ldots, \tag{5.397}$$

since it is reasonable to assume a weak pump. The zeroth-order spectrum,

$$\begin{aligned} S_\mathrm{a}^{(0)}(\omega_\mathrm{p}) &= \left\langle \frac{2\gamma_\perp}{\gamma_\perp^2 + (\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})^2} \right\rangle_\mathbf{v} \\ &= \left\langle \frac{2\gamma_\perp}{\gamma_\perp^2 + (\omega_\mathrm{p} - \omega_0 + \mathbf{k}\cdot\mathbf{v})^2} \right\rangle_\mathbf{v}, \end{aligned} \tag{5.398}$$

is simply the Doppler-broadened absorption line (the Doppler velocity profile convolved with the atomic Lorentzian line). The next order is much more complicated, but includes the Lamb dip:

$$\begin{aligned} &S_\mathrm{a}^{(2)}(\omega_\mathrm{p}) = \\ &-\left\langle \frac{2\gamma_\perp^2}{\Gamma[\gamma_\perp^2 + (\Delta - \mathbf{k}\cdot\mathbf{v})^2][\gamma_\perp^2 + (\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})^2]} \right\rangle_\mathbf{v} \\ &+\left\langle \frac{i(\Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v})}{2[\gamma_\perp + i(\Delta - \mathbf{k}\cdot\mathbf{v})][\Gamma - i(\Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v})][\gamma_\perp + i(\Delta - \Delta_\mathrm{p} - 3\mathbf{k}\cdot\mathbf{v})][\gamma_\perp - i(\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})]} + \mathrm{c.c.} \right\rangle_\mathbf{v}, \\ &-\left\langle \frac{[\gamma_\perp - i(\Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v})]}{[\Gamma - i(\Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v})][\gamma_\perp + i(\Delta - \Delta_\mathrm{p} - 3\mathbf{k}\cdot\mathbf{v})][\gamma_\perp - i(\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})]^2} + \mathrm{c.c.} \right\rangle_\mathbf{v}. \end{aligned} \tag{5.399}$$

To interpret this equation, we can focus on the resonant factor in the denominators. There are several different factors to discuss here:

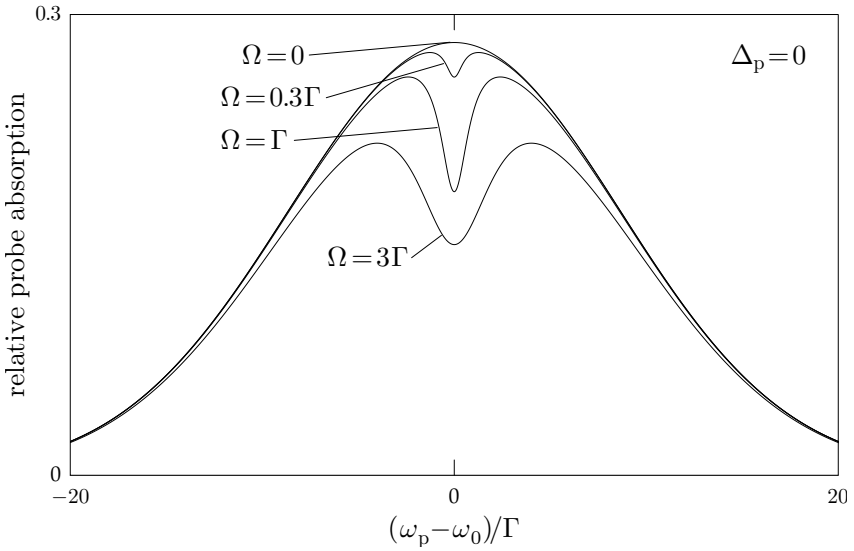
1. $[\gamma_\perp^2 + (\Delta - \mathbf{k}\cdot\mathbf{v})^2]$ or $[\gamma_\perp + i(\Delta - \mathbf{k}\cdot\mathbf{v})]$: simply says that the pump beam at detuning Δ is on resonance with atoms of velocity Δ/k along the direction of the pump beam. This factor is simply the overall Doppler profile. Of course, a Lamb dip can only occur inside the Doppler-broadened line.

2. $[\gamma_\perp^2 + (\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})^2]$ or $[\gamma_\perp - i(\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v})]$: has a resonance at $\Delta + \Delta_\mathrm{p} + \mathbf{k}\cdot\mathbf{v} = \omega_\mathrm{p} - \omega_0 + \mathbf{k}\cdot\mathbf{v} = 0$. This is the same overall Doppler profile, but for probe resonance. Thus, the first term in $S_\mathrm{a}^{(2)}(\omega_\mathrm{p})$ does not contribute to the Lamb dip, since it contains this and the former factor. This term simply represents lowest-order overall saturation of the Doppler transition, without any coherence-type effects.

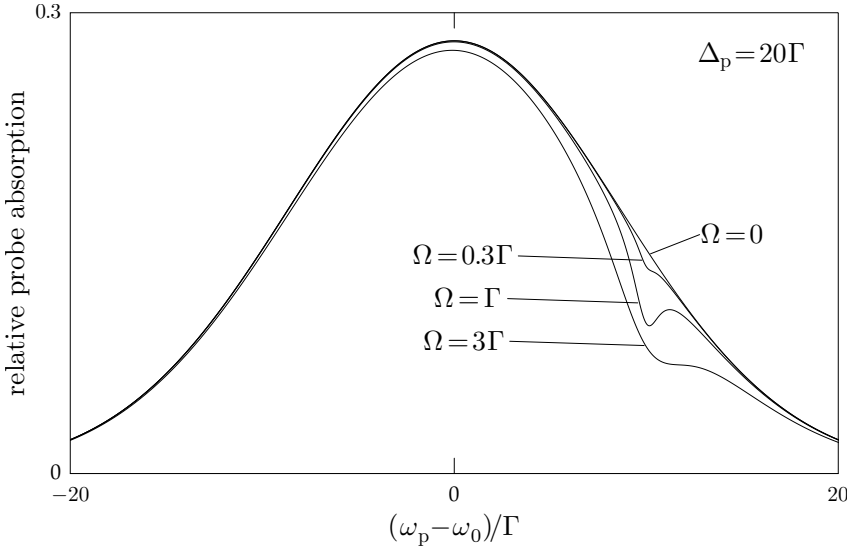
3. $[\Gamma - i(\Delta_\mathrm{p} + 2\mathbf{k}\cdot\mathbf{v})]$: resonance occurs for the velocity class $v = -\Delta_\mathrm{p}/2k = -(\omega_\mathrm{p} - \omega)/2k$ along the pump direction, which gives $v = 0$ for degenerate pump and probe beams ($\Delta_\mathrm{p} = 0$), picking out the atoms at rest. For nondegenerate pump-probe pairs, a moving velocity class is selected, effectively moving the Lamb dip. Notice that the last two terms seems to compete; but while the second term only has dispersive behavior, the last term has the only absorptive-type structure, which we identify as the dip.

4. $[\gamma_\perp + i(\Delta - \Delta_\mathrm{p} - 3\mathbf{k}\cdot\mathbf{v})]$: this factor gives rise to a secondary resonance at $2\omega - \omega_0 - \omega_\mathrm{p} - 3\mathbf{k}\cdot\mathbf{v} = 0$, or for degenerate pump-probe pairs, this selects out the velocity class $v = (\omega_\mathrm{p} - \omega_0)/3k$. This is the lowest-order "Doppleron" resonance (a higher-order treatment reveals more resonances of this form).[30] Since the velocity class is swept with the probe frequency, we do not see structures due to this term in the spectra.

The saturation spectrum for a degenerate pump-probe pair ($\Delta_\mathrm{p} = 0$) is shown here. We have chosen a fairly narrow Doppler width of 10Γ for clarity of the plots, with no collisions ($\gamma_\perp = \Gamma/2$). Without a pump ($\Omega = 0$), we simply get the Doppler-broadened line. For small pump powers, the Lamb dip appears at the atomic resonance, and it has a width of order Γ (though in reality somewhat larger than Γ). As the pump becomes stronger, the dip becomes more pronounced, but for high pump intensities, the dip power broadens.



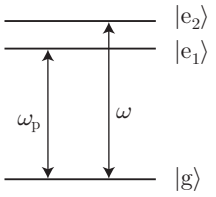
In the nondegenerate pump-probe case, we see essentially the same behavior, but now the Lamb dip is shifted by half the pump-probe detuning $\Delta_\mathrm{p} = 20\Gamma$. When the probe is at $+10\Gamma$, the pump is at -10Γ, so the average frequency matches the atomic resonance. In this case, the dip is due to atoms moving at $v = -10\Gamma/2k$ along the pump direction, which sees both beams as resonant.



[30] Stig Stenhom, *Foundations of Laser Spectroscopy* (Wiley, 1984), section 4.2, p. 149.

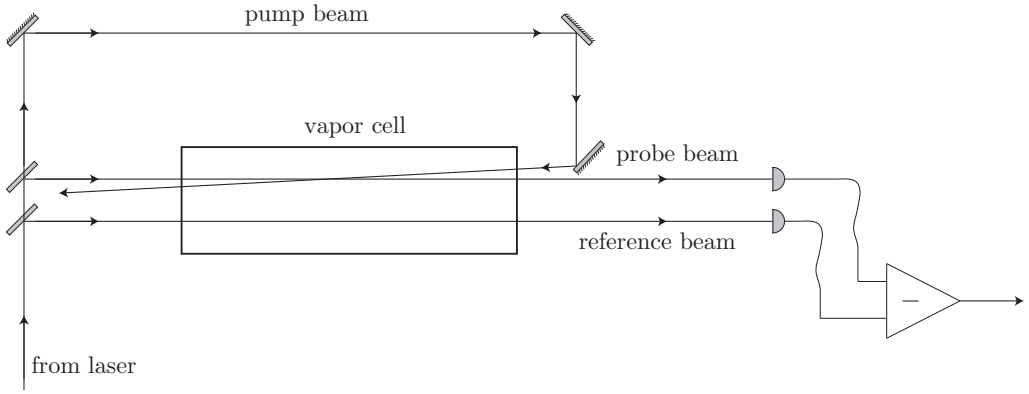
As we mentioned above, one of the main applications for saturation spectroscopy is for precision, absolute frequency references, such as for active frequency stabilization and absolute referencing of lasers. Saturation spectroscopy is the primary method used for stabilizing lasers for laser cooling and trapping, for example. Saturation spectroscopy can also be used to ''automatically,'' or passively, stabilize a laser system. One notable example is the methane-stabilized He–Ne laser, where a CH_4 cell is placed inside the linear resonator of a He-Ne laser. This system relies on the coincidence of a He-Ne laser line with a methane absorption line at 3.39 μm. The lifetime of the methane transition is around 10 ms, so that the resulting Lamb dip is very narrow (about 400 kHz). The Doppler-broadened methane absorption line causes extra loss of the He-Ne laser, which is minimized inside the Lamb dip. The laser thus naturally oscillates in the narrow frequency band of the dip. In the first study,[31] two He-Ne lasers were locked by this method to within 1 kHz of each other, for an absolute reproducibility of a part in 10^{11}. The frequency of the methane-stabilized He-Ne is very precisely known, with a modern measurement[32] giving a frequency of 88 376 182 599 976(10) Hz.

Finally, one other phenomenon that comes up in saturation spectroscopy is the *crossover resonance*. If multiple transitions lie within the Doppler width, of course you would expect the saturation spectrum to have a Lamb dip for each transition. However, one also finds an extra resonance for each pair of transitions. For a degenerate pump-probe pair, the two fields can be resonant with different transitions for a particular group of moving atoms.



This will result in another Lamb-type dip in the absorption spectrum, located halfway in between the Lamb dips for the individual transitions. The effect is somewhat different, though: the crossover dips are due to *depletion* of the transition rather than added saturation. Depletion tends to be more effective than saturation at reducing the absorption coefficient, and so the crossover dips tend to dominate the saturation spectra.

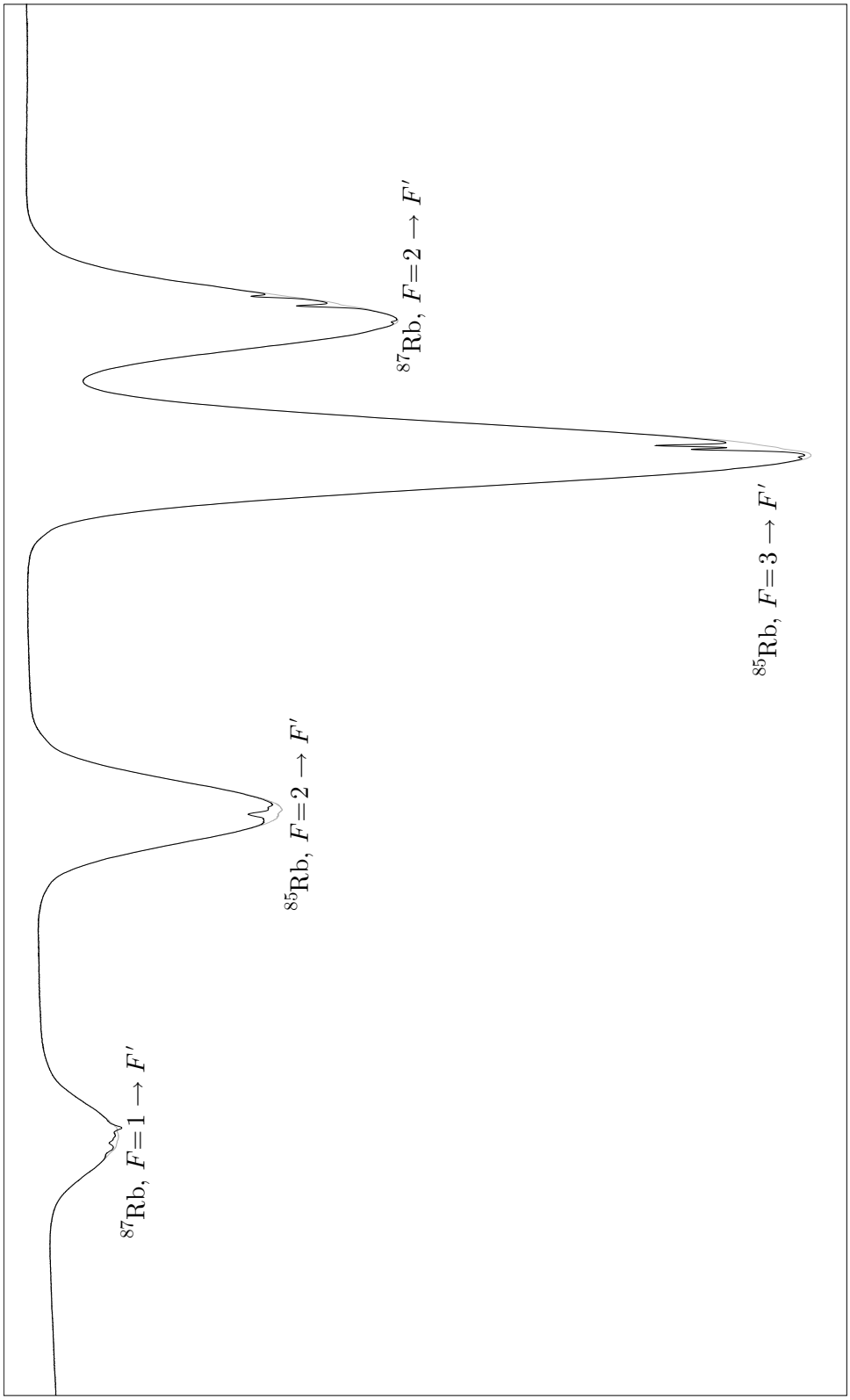
To illustrate Lamb-dip spectroscopy in a real system, we will look at a saturation spectrum of ^{87}Rb and ^{85}Rb (the two naturally abundant isotopes) in a room-temperature vapor cell.



Three beams are split from a common laser source. Two probes are sent through the cell. The third beam acts as a pump, and nearly counterpropagates with one of the probe beams, producing the Lamb dips. The other probe acts as an unsaturated reference, so that the Doppler profile can be subtracted from the Lamb-dip signal. The saturation spectrum is shown in the following plot.

[31] R. L. Barger and J. L. Hall, ''Pressure Shift and Broadening of Methane Line at 3.39 μ Studied by Laser-Saturated Molecular Absorption,'' *Physical Review Letters* **22**, 4 (1969) (doi: http://link.aps.org/abstract/PRL/v22/p4).

[32] P. V. Pokasov, R. Holzwarth, Th. Udem, J. Reichert, M. Niering, M. Zimmermann, M. Weitz, T. W. Hänsch, A. K. Dmitriev, S. N. Bagayev, P. Lemonde, G. Santarelli, P. Laurent, M. Abgrall, A. Clairon, and C. Salomon, ''Absolute Frequency Measurements of a Methane-Stabilized Transportable He-Ne Laser at 3.39 μm,'' in *Proceedings of the Sixth Symposium on Frequency Standards and Metrology*, P. Gill, Ed. (World Scientific, 2002), p. 510.



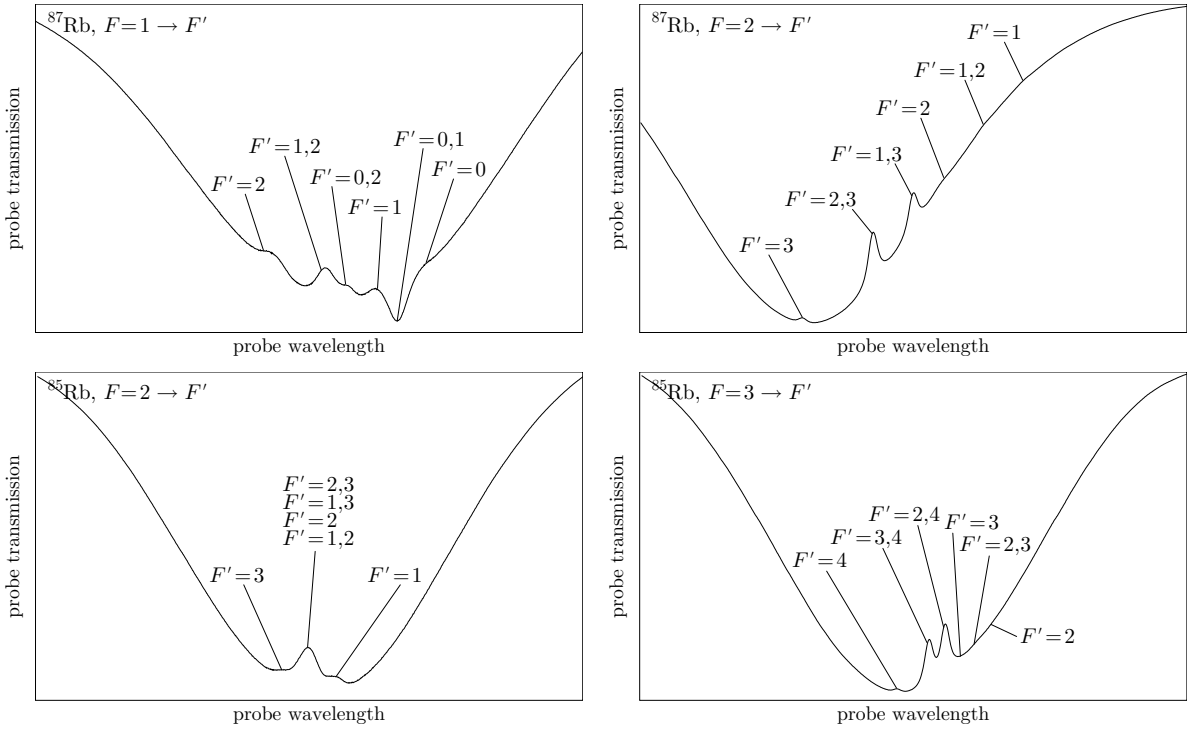
For this plot, there is no subtraction of the Doppler background. There are four Doppler-broadened ab-

sorption lines, two lines for each isotope of rubidium. Each pair of absorption lines corresponds to the two hyperfine ground states of each rubidium isotope, and each line represents a triplet of hyperfine transitions merged together within the Doppler width. Thus, we expect three Lamb dips and three crossover dips in each absorption line. For a sense of scale, the splitting between the two ^{87}Rb multiplets is 6.8 GHz, the splitting between the two ^{85}Rb multiplets is 3.0 GHz, the Doppler width is about 500 MHz at room temperature, and the natural line widths are about 5 MHz. Along with the saturation spectrum, the unsaturated spectrum (measured from the same probe beam, but with the pump beam blocked). Also shown here are zoomed versions of each of the absorption lines, with each of the saturation and crossover dips marked explicitly by the excited-state hyperfine number F'.

For the laser cooling transition of ^{87}Rb (the top right plot), the dips are well-resolved, though not necessarily strong. For the repumping transition of ^{87}Rb (the top left plot), the dips are more closely spaced, but still resolved. This is because the hyperfine splittings between the excited states is relatively large: 267 MHz between $F' = 2$ and 3, 157 MHz between $F' = 1$ and 2, and 72 MHz between $F' = 1$ and 2, compared to the 6 MHz line width. (Only 3 excited states can couple to each ground state, because F, being an angular momentum interacting with a spin-1 photon, can only change by 1, at least to leading order.) Oddly, one of the dips (the $F' = 0, 1$ crossover) is a ''negative dip.'' Reversed peaks such as this are sometimes possible,[33] and reflect the complicated degenerate Zeeman-level substructure of the hyperfine levels.

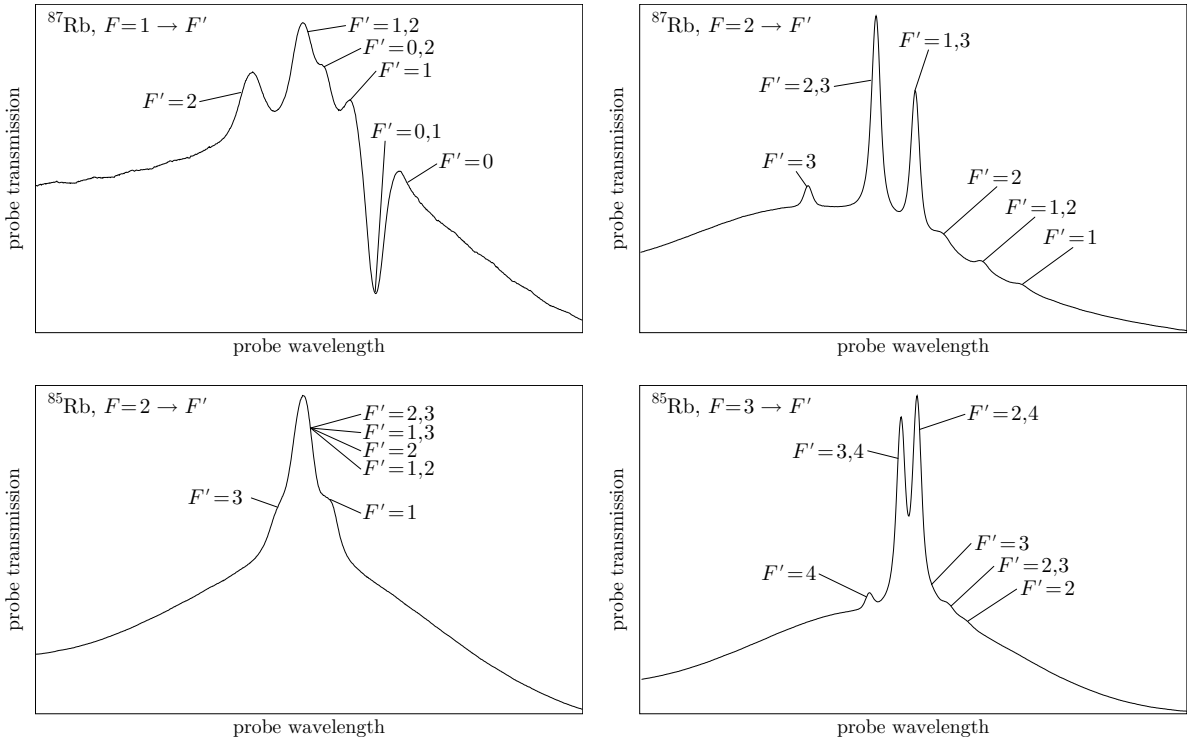
In ^{85}Rb, the states are much less well-resolved, because the hyperfine splittings are smaller: 39 MHz between $F' = 2$ and 3, 21 MHz between $F' = 1$ and 2, and 10 MHz between $F' = 1$ and 2, compared to the 6 MHz line width. (It is precisely this reason that ^{85}Rb is a difficult isotope to use for laser cooling and trapping.) In particular, for the ''repumping'' transition (lower left plot), four of the lines are merged together, and the other two lines are only marginally well resolved.

Especially in the case of the laser cooling transitions (right two plots), the dips are centered to the right of the Doppler line center. This is not because of a frequency offset between the pump and probe lasers. This is because the state with the largest hyperfine number is the strongest transition, contributing the most to the Doppler absorption profile. The Doppler line center is thus pulled towards these states.



[33] O. Schmidt, K.-M. Knaak, R. Wynands, and D. Meschede, ''Cesium saturation spectroscopy revisited: How to reverse peaks and observe narrow resonances,'' *Applied Physics B: Lasers and Optics* **59**, 167 (1994) (doi: 10.1007/BF01081167).

The following four plots are the same as the previous four zoomed plots, but here the Doppler background is (mostly) subtracted away by using the reference beam in the experimental setup. The dips are somewhat easier to see here, although their placement within the Doppler lines is of course not apparent



5.8 Mechanical Effects of Light on Two-Level Atoms

We considered the dipole force in our classical treatment of the atom, but now it is time for a proper quantum derivation. In particular, the internal and external atomic dynamics are coupled and we will show that under suitable conditions, they decouple to good approximation.

We will begin with a perturbative treatment, the basic conclusion being that under the proper conditions, it is possible to ignore the internal electronic structure of the atom, and treat the atom as a point particle. Furthermore, the ''reduced'' atom moves under the influence of the effective center-of-mass Hamiltonian

$$H_{\text{eff}} = \frac{p^2}{2m} + V_{\text{eff}}(x), \tag{5.400}$$

where m is the atomic mass and the potential V_{eff} is proportional to the laser intensity and inversely proportional to the detuning from the (nearest) atomic resonance. We will then examine things more generally in terms of dressed states and look at corrections to the simple dipole-force picture.

5.8.1 Atom-Field Interaction

We must now redo the atom-field interaction, this time including the center-of-mass motion of the atom. Considering a linearly polarized field,

$$\begin{aligned} \mathbf{E}(x,t) &= \hat{\varepsilon}\left(E^{(+)}(x)e^{-i\omega t} + E^{(-)}(x)e^{i\omega t}\right) \\ &= \mathbf{E}^{(+)}(x,t) + \mathbf{E}^{(-)}(x,t), \end{aligned} \tag{5.401}$$

where again $\mathbf{E}^{(+)}$ and $\mathbf{E}^{(-)}$ are the positive- and negative-rotating components of the field, respectively, and $E^{(\pm)}(x)$ is the space-dependent amplitude of the field.

The atomic free-evolution Hamiltonian is then given by

$$H_{\text{A}} = \frac{p^2}{2m} + \hbar\omega_0 |\text{e}\rangle\langle\text{e}|, \tag{5.402}$$

which is the same as before except for the inclusion of the kinetic energy. The atom-field interaction Hamiltonian is still given (in the dipole approximation) by

$$H_{\text{AF}} = -\mathbf{d}\cdot\mathbf{E}, \tag{5.403}$$

where $\mathbf{d}$ is the atomic dipole operator. In the rotating-wave approximation, this becomes

$$H_{\text{AF}} = \frac{\hbar}{2}\left[\Omega^*(x)\,\sigma e^{i\omega t} + \Omega(x)\,\sigma^\dagger e^{-i\omega t}\right], \tag{5.404}$$

where

$$\Omega(x) := -\frac{2\langle\text{g}|\hat{\varepsilon}\cdot\mathbf{d}|\text{e}\rangle E^{(+)}(x)}{\hbar} \tag{5.405}$$

is the space-dependent Rabi frequency, which is no longer a real number in general (due to the e^{ikx}-type dependence of the field). In the rotating frame, the interaction becomes

$$\tilde{H}_{\text{AF}} = \frac{\hbar}{2}\left[\Omega^*(x)\,\sigma + \Omega(x)\,\sigma^\dagger\right], \tag{5.406}$$
(atom–field interaction)

and the free Hamiltonian becomes

$$\tilde{H}_{\text{A}} = \frac{p^2}{2m} - \hbar\Delta|\text{e}\rangle\langle\text{e}|, \tag{5.407}$$
(free atomic evolution)

so that the electronic states are nearly degenerate.

5.8.2 Schrödinger Equation

We assume that the detuning from resonance is large ($|\Delta| \gg \Gamma$), we will neglect spontaneous emission and use the Schrödinger equation

$$(\tilde{H}_{\text{A}} + \tilde{H}_{\text{AF}})|\psi\rangle = i\hbar\partial_t|\psi\rangle \tag{5.408}$$

to describe the atomic evolution. It is convenient to decompose the state vector $|\psi\rangle$ into a product of internal and external states,

$$|\psi\rangle = |\psi_{\text{e}}(t)\rangle\,|\text{e}\rangle + |\psi_{\text{g}}(t)\rangle\,|\text{g}\rangle \tag{5.409}$$

where the $|\psi_\alpha(t)\rangle$ are state vectors in the center-of-mass space of the atom. In the following, we will associate all time dependence of the atomic state with the center-of-mass components of the state vector. Defining the coefficients $\psi_\alpha(x,t) := \langle x|\psi_\alpha(t)\rangle$, the equation of motion for the wave function $\langle x|\psi\rangle$ becomes

$$i\hbar\left(\partial_t\psi_{\text{e}}|\text{e}\rangle + \partial_t\psi_{\text{g}}|\text{g}\rangle\right) = \frac{p^2}{2m}\left(\psi_{\text{e}}|\text{e}\rangle + \psi_{\text{g}}|\text{g}\rangle\right) - \hbar\Delta\psi_{\text{e}}|\text{e}\rangle + \frac{\hbar}{2}\left[\Omega^*(x)\psi_{\text{e}}|\text{g}\rangle + \Omega(x)\psi_{\text{g}}|\text{e}\rangle\right]. \tag{5.410}$$

Separating the coefficients of $|\text{e}\rangle$ and $|\text{g}\rangle$, we obtain the coupled pair of equations

$$\begin{aligned} i\hbar\partial_t\psi_{\text{e}} &= \frac{p^2}{2m}\psi_{\text{e}} + \frac{\hbar\Omega(x)}{2}\psi_{\text{g}} - \hbar\Delta\psi_{\text{e}} \\ i\hbar\partial_t\psi_{\text{g}} &= \frac{p^2}{2m}\psi_{\text{g}} + \frac{\hbar\Omega^*(x)}{2}\psi_{\text{e}}. \end{aligned} \tag{5.411}$$

for the wave functions $\psi_\alpha(x,t)$.

5.8.3 Adiabatic Approximation

The equations of motion (5.411) are greatly simplified by using the *adiabatic approximation*, which we have seen a couple of times thus far. We can motivate this approximation by examining the various time scales in the evolution of ψ_e and ψ_g. The kinetic-energy terms in Eqs. (5.411) induce variations on time scales corresponding to kHz frequencies for ultracold atoms. However, the pump-field terms induce motion on a time scale corresponding to the Rabi frequency—typically from zero to several hundred MHz—and the free evolution term induces motion of ψ_e on a time scale corresponding to Δ, typically several to many GHz. Together, these terms induce internal atomic oscillations at the generalized Rabi frequency $\tilde{\Omega}(x) := \sqrt{\Omega^2(x) + \Delta^2} \approx |\Delta|$. Furthermore, in between these long and short time scales of external and internal atomic motion lies the damping time scale due to coupling with the vacuum, which corresponds to the natural decay rate Γ, and $\Gamma/2\pi$ is typically on the order of a few MHz for alkali atoms. Because we are primarily interested in the slow center-of-mass atomic motion, and the internal atomic dynamics take place over times much shorter than the damping time, it is a good approximation to assume that the internal motion is damped instantaneously to equilibrium, when compared to the external motion. Thus, $\partial_t \psi_e = 0$, because ψ_e is the variable that carries the natural internal free-evolution time dependence at frequency Δ, whereas ψ_g has no natural internal oscillation, because the state $|g\rangle$ is at zero energy. This approximation then gives a relation between ψ_e and ψ_g:

$$\left(\hbar\Delta - \frac{p^2}{2m}\right)\psi_e \approx \left(\frac{\hbar\Omega(x)}{2}\right)\psi_g. \tag{5.412}$$

Noting that the kinetic energy $p^2/2m$ is negligible compared to $\hbar\Delta$, we can then use this constraint to eliminate ψ_e in the second of Eqs. (5.411), with the result

$$i\hbar\partial_t\psi_g = \left(\frac{p^2}{2m}\right)\psi_g + \frac{\hbar|\Omega(x)|^2}{4\Delta}\psi_g. \tag{5.413}$$

Since the detuning is large, nearly all the population is contained in $|g\rangle$, so the excited state completely drops out of the problem. Hence, the atom obeys the Schrödinger equation with the effective center-of-mass Hamiltonian

$$H_{\text{eff}} = \frac{p^2}{2m} + V_{\text{eff}}(x), \tag{5.414}$$

where

$$V_{\text{eff}}(x) = \frac{\hbar|\Omega(x)|^2}{4\Delta}, \tag{5.415}$$

(effective optical potential)

and the atom behaves like a point particle in an effective potential, where the strength of the potential is given by (5.415).

5.8.3.1 Master-Equation Approach

It is also instructive to make the adiabatic approximation from the viewpoint of a master equation, where we can more explicitly see the effects of damping on the atomic motion. The idea will follow that of the adiabatic approximation for obtaining the rate equations from the optical Bloch equations from before (Section 5.6.2). The master equation for the atomic evolution (i.e., the optical Bloch equations generalized to include center-of-mass motion) has the general form in the absence of collisions

$$\partial_t\tilde{\rho}(t) = -\frac{i}{\hbar}[\tilde{H}_A + \tilde{H}_{AF}, \tilde{\rho}(t)] + \Gamma\int d\Omega\, f_{\hat{\varepsilon}}(\theta,\phi)\mathcal{D}[\sigma e^{-i\mathbf{k}_L\cdot\mathbf{r}}]\tilde{\rho}(t), \tag{5.416}$$

where $\mathbf{k}_L$ is the wave vector of the emitted photon (and $d\Omega$ is the angular integration element, not to be confused with the Rabi frequency). This is the same master equation as for the optical Bloch equations, except for three modifications: (1) the atomic kinetic energy is now included in $\tilde{H}_A$, (2) the spatial dependence of the field is now included in $\tilde{H}_{AF}$, and (3) we have made the replacement $\sigma \longrightarrow \sigma e^{-i\mathbf{k}_L\cdot\mathbf{r}}$ in

the Lindblad superoperator, since any spontaneous emission must be accompanied by a **photon recoil** to conserve total momentum, and then we have integrated over all possible emission directions, weighted by the dipole radiation pattern $f_{\hat{\varepsilon}}(\theta,\phi)$. We can write out the effect of the dissipation operator more explicitly, with the result

$$\partial_t\tilde{\rho}(t) = -\frac{i}{\hbar}[\tilde{H}_{\rm A}+\tilde{H}_{\rm AF},\tilde{\rho}(t)] - \frac{\Gamma}{2}\left[\sigma^\dagger\sigma,\tilde{\rho}\right]_+ + \Gamma\int d\Omega f_{\hat{\varepsilon}}(\theta,\phi)e^{ik_{\rm L}x\sin\theta\cos\phi}\sigma\rho\sigma^\dagger e^{-ik_{\rm L}x\sin\theta\cos\phi}, \tag{5.417}$$

where for simplicity we now restrict out attention to one dimension. Note that in writing down the master equation (5.417), we have assumed purely radiative damping. We can then write out the equations of motion for the density matrix elements $\tilde{\rho}_{\alpha\beta}(x,x',t) := \langle x|\langle\alpha|\tilde{\rho}|\beta\rangle|x'\rangle$ as

$$\begin{aligned}
\partial_t\rho_{\rm gg} &= -\frac{i}{\hbar}\left[\frac{p^2}{2m},\rho_{\rm gg}\right] - \frac{i}{2}\left(\Omega^*(x)\tilde{\rho}_{\rm eg}-\tilde{\rho}_{\rm ge}\Omega(x)\right) + \Gamma\int d\Omega f_{\hat{\varepsilon}}(\theta,\phi)e^{ik_{\rm L}x\sin\theta\cos\phi}\rho_{\rm ee}e^{-ik_{\rm L}x\sin\theta\cos\phi}\\
\partial_t\rho_{\rm ee} &= -\frac{i}{\hbar}\left[\frac{p^2}{2m},\rho_{\rm ee}\right] + \frac{i}{2}\left(\Omega^*(x)\tilde{\rho}_{\rm eg}-\tilde{\rho}_{\rm ge}\Omega(x)\right) - \Gamma\rho_{\rm ee}\\
\partial_t\tilde{\rho}_{\rm ge} &= -\frac{i}{\hbar}\left[\frac{p^2}{2m},\tilde{\rho}_{\rm ge}\right] - \left(\frac{\Gamma}{2}+i\Delta\right)\tilde{\rho}_{\rm ge} - \frac{i}{2}\left[\Omega^*(x)\rho_{\rm ee}-\rho_{\rm gg}\Omega^*(x)\right]\\
\partial_t\tilde{\rho}_{\rm eg} &= -\frac{i}{\hbar}\left[\frac{p^2}{2m},\tilde{\rho}_{\rm eg}\right] - \left(\frac{\Gamma}{2}-i\Delta\right)\tilde{\rho}_{\rm eg} - \frac{i}{2}\left[\Omega(x)\rho_{\rm gg}-\rho_{\rm ee}\Omega(x)\right].
\end{aligned} \tag{5.418}$$

We again assume that $|\Delta|\gg\Gamma$ and note that the equations have fast internal driving terms (with frequencies comparable to or greater than Γ) and slow center-of-mass terms; this time, however, the equations of motion for the coherences (which are responsible for the population oscillations) have explicit damping terms. Since we are interested in the slow external motion, we can use the fact that the steady-state solution for $\tilde{\rho}_{\rm ee}$ is of order $(\Gamma/\Delta)^2$, whereas the steady state solutions for the coherences $\tilde{\rho}_{\rm eg}$ and $\tilde{\rho}_{\rm ge}$ are of order $|\Gamma/\Delta|$, so that we can neglect the $\tilde{\rho}_{\rm ee}$ terms on the right-hand sides of these equations. Now, we will assume that the quickly rotating coherences are damped to equilibrium on a time scale short compared to the external motion of interest, and hence set $\partial_t\tilde{\rho}_{\rm ge}\approx\partial_t\tilde{\rho}_{\rm eg}\approx 0$. Doing so leads to the adiabatic relations

$$\begin{aligned}
\tilde{\rho}_{\rm ge} &= \rho_{\rm gg}\frac{\Omega^*(x)}{2\Delta}\\
\tilde{\rho}_{\rm eg} &= \frac{\Omega(x)}{2\Delta}\rho_{\rm gg},
\end{aligned} \tag{5.419}$$

where we have neglected the momentum and Γ terms in comparison to the Δ term. Substituting Eqs. (5.419) into the equation of motion for $\tilde{\rho}_{\rm gg}$ (and neglecting the $\tilde{\rho}_{\rm ee}$ term), we find

$$\partial_t\tilde{\rho}_{\rm gg} = -\frac{i}{\hbar}\left[\frac{p^2}{2m}+\frac{\hbar|\Omega(x)|^2}{4\Delta},\rho_{\rm gg}\right]. \tag{5.420}$$

This equation is simply the equation of motion for $\tilde{\rho}_{\rm gg}$ under the Hamiltonian

$$H_{\tilde{\rho}} = \frac{p^2}{2m}+\frac{\hbar|\Omega(x)|^2}{4\Delta}, \tag{5.421}$$

which is just the effective Hamiltonian (5.414). Note that we have ultimately discarded the spontaneous emission effects, which lead to extra diffusive terms in the reduced evolution equations here, as we discuss below.

From this approach, it is clear that the adiabatic approximation is good after a time on the order of $1/\Gamma$, when the coherences have damped away. After this initial transient, the adiabatic approximation remains good as long as any modulations of the optical potential take place over a time long compared to $1/\tilde{\Omega}$. This is clear from the dressed-state analysis below, because such modulations will not excite transitions between the dressed states and thus cause the adiabatic approximation to break down.

5.8.3.2 Bragg Scattering in an Optical Standing Wave

As an example to gain some physical insight into the nature of the dipole force, we will consider the problem of Bragg scattering of a two-level atom in a weak, optical standing wave. The potential is, of course, sinusoidal in space with period $\lambda/2$, and so this setup is equivalent to the quantum pendulum, but in a regime where the atoms have enough energy that they are not bound to the **optical lattice** of potential wells (rotational pendulum motion). The quantum-pendulum dynamics show a feature that is distinctly nonclassical: the momentum transferred from the potential to the atoms is quantized. To see this directly, we consider The following argument. For a standing wave composed of two equal but counterpropagating traveling waves, with a field of the form

$$\begin{aligned}\mathbf{E}(x,t) &= \hat{z}E_0[\cos(kx-\omega t)+\cos(kx+\omega t)] \\ &= \hat{z}E_0\cos(kx)\left(e^{-i\omega t}+e^{i\omega t}\right) \\ &= \mathbf{E}^{(+)}(x,t)+\mathbf{E}^{(-)}(x,t),\end{aligned} \tag{5.422}$$

the Rabi frequency is simply $\Omega(x)=\Omega_0\cos(kx)$. Thus, the effective dipole potential is (dropping an overall constant)

$$V_{\text{eff}}(x)=V_0\cos(2kx), \tag{5.423}$$

where the potential amplitude is

$$V_0=\frac{\hbar|\Omega_0|^2}{8\Delta}. \tag{5.424}$$

The Schrödinger equation is

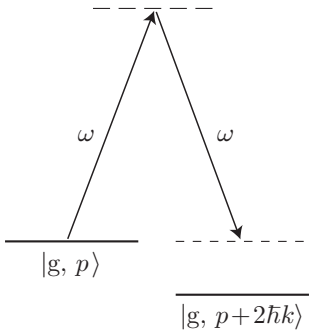
$$\begin{aligned}i\hbar\partial_t|\psi\rangle &= \left[\frac{p^2}{2m}+V_0\cos(2kx)\right]|\psi\rangle \\ &= \left[\frac{p^2}{2m}+\frac{V_0}{2}\left(e^{i2kx}+e^{-i2kx}\right)\right]|\psi\rangle,\end{aligned} \tag{5.425}$$

which can be written in the momentum representation as

$$i\hbar\partial_t\psi(p)=\frac{p^2}{2m}\psi(p)+\frac{V_0}{2}\left[\psi(p+2\hbar k)+\psi(p-2\hbar k)\right], \tag{5.426}$$

where $\psi(p):=\langle p|\psi\rangle$. This form follows from either recognizing $\exp(ikx)$ as a momentum-displacement operator, or by carrying out an explicit Fourier transform of the equation from the position to the momentum representation, and the explicit proof is left as an exercise. So, the evolution in the standing wave imposes a ''ladder'' structure in momentum, such that an atom beginning in a plane-wave state $|p\rangle$ can only subsequently occupy the states $|p+n2\hbar k\rangle$ for integer n. This momentum quantization has a clear interpretation in terms of the stimulated scattering of lattice photons: if the atom absorbs a photon that was traveling in one direction and then re-emits it into the counterpropagating mode, the atom will recoil, changing its momentum by twice the photon momentum, or by $2\hbar k$. Of course, the argument that we just considered was based on a classical treatment of the field, so it is the spatial periodicity of the potential that imposes the ladder structure in this model.

However, the momentum transfer to the atoms can be viewed as a **stimulated Raman transition**—a two-photon transition from one ground state to an excited state and back to another ground state—between different motional states, say $|\mathrm{g},p\rangle$ and $|\mathrm{g},p+2\hbar k\rangle$.



We can use Eq. (5.426) to write down coupled equations for the two states. Assuming that couplings to other states are negligible, we have

$$\begin{aligned} i\hbar\partial_t\psi(p) &= \frac{p^2}{2m}\psi(p) + \frac{V_0}{2}\psi(p+2\hbar k) \\ i\hbar\partial_t\psi(p+2\hbar k) &= \frac{(p+2\hbar k)^2}{2m}\psi(p) + \frac{V_0}{2}\psi(p)., \end{aligned} \tag{5.427}$$

These are the equations of motion for a two-level system, coupled with Rabi frequency

$$\Omega_{\rm R} = \frac{V_0}{\hbar} = \frac{|\Omega_0|^2}{8\Delta}, \tag{5.428}$$

with a dc interaction between states of energy difference

$$\Delta E = \frac{(p+2\hbar k)^2}{2m} - \frac{p^2}{2m} = \frac{2\hbar kp}{m} + 4\hbar\omega_{\rm r}, \tag{5.429}$$

where the recoil energy

$$\hbar\omega_{\rm r} := \frac{\hbar^2k^2}{2m} \tag{5.430}$$

(recoil energy)

is the atomic kinetic energy associated with a single photon recoil. Thus, atomic population oscillates between the two momentum states at the Raman Rabi frequency $\Omega_{\rm R}$: since we have adiabatically eliminated the intermediate (excited) state, the three-level system behaves approximately as an effective two-level system.

More generally speaking, the coupling between these two levels is described by a Raman Rabi frequency (as in the two-level atom), given by

$$\Omega_{\rm R} = \frac{\Omega_1\Omega_2}{2\Delta}, \tag{5.431}$$

(Raman Rabi frequency)

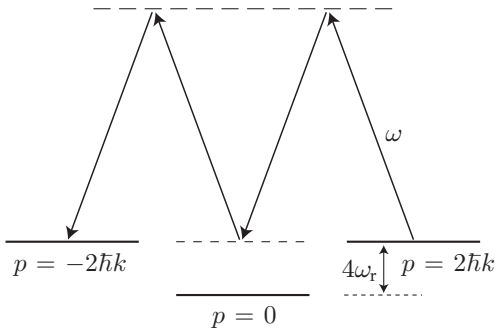
where $\Omega_{1,2}$ are the Rabi frequencies associated separately with each traveling-wave component of the standing wave, and Δ is the mutual detuning to the atomic excited state (the relative frequency difference is constrained by energy conservation to be the splitting between the motional states). To connect with the notation that we have already used, $\Omega_1 = \Omega_2 = \Omega_0/2$ for the case of identical traveling waves, so that $\hbar\Omega_{\rm R} = V_0$, and thus again V_0 also represents the strength of the Raman couplings.

The two-photon, stimulated Raman transition is an example of a **Bragg scattering** process.[34] In fact, it is the simplest (''first-order'') form of Bragg scattering; in general, nth-order Bragg scattering is a

[34]For experiments and more details about Bragg scattering in atom optics, see Peter J. Martin, Bruce G. Oldaker, Andrew H. Miklich, and David E. Pritchard, ''Bragg scattering of atoms from a standing light wave,'' *Physical Review Letters* **60**, 515 (1988) (doi: 10.1103/PhysRevLett.60.515); David M. Giltner, Roger W. McGowan, and Siu Au Lee, ''Theoretical and experimental study of the Bragg scattering of atoms from a standing light wave,'' *Physical Review A* **52**, 3966 (1995) (doi: 10.1103/PhysRevA.52.3966), and M. Kozuma, L. Deng, E. W. Hagley, J. Wen, R. Lutwak, K. Helmerson, S. L. Rolston, and W. D. Phillips, ''Coherent Splitting of Bose-Einstein Condensed Atoms with Optically Induced Bragg Diffraction,'' *Physical Review Letters* **82**, 871 (1999) (doi: 10.1103/PhysRevLett.82.871).

$2n$-photon transition spanning an interval of $2n\hbar k$ in momentum between the $|\pm n\hbar k\rangle$ states.. The term ''Bragg scattering'' applies to the weakly coupled regime, where the intermediate states are not appreciably populated, and so the transition between the two distant momentum states can be treated as a two-level problem. In this regime, classical transport between these distinct momentum regions is forbidden, as the classical potential is not sufficiently strong to cause a correspondingly large change in the classical momentum. As such, Bragg scattering is an example of **dynamical tunneling**, which is quantum tunneling between regions in phase space between which classical transport is forbidden, but by the dynamics (here, the nature of asymptotically free-particle motion) rather than by a potential barrier.

Although the potential has a small amplitude, quantum coherence can build up as the atoms sample the potential and cause the atoms to significantly change their motion. We will illustrate this process by considering the relatively simple case of second-order Bragg scattering, and then we will generalize our results to the nth-order case. We consider the case where the standing wave is stationary, so that only the states $|-2\hbar k\rangle$ and $|2\hbar k\rangle$ are resonantly coupled in the limit of small $\Omega_{\rm R}$. No other states will be substantially coupled by these fields, unless the Raman Rabi frequency is large enough to power-broaden the off-resonant transitions, which would not correspond to the Bragg regime. The relevant energy-level diagram is shown below, which shows that the detuning from the $|p=0\rangle$ motional state is simply the kinetic-energy shift.



Neglecting couplings to other states (which are even further detuned than the $|p=0\rangle$ state), the Schrödinger equation for the three coupled momentum states then becomes

$$\begin{aligned} i\hbar\partial_t\psi(-2\hbar k,t) &= \frac{(-2\hbar k)^2}{2m}\psi(-2\hbar k,t)+\frac{\hbar\Omega_{\rm R}}{2}\psi(0,t)\\ i\hbar\partial_t\psi(0,t) &= \frac{\hbar\Omega_{\rm R}}{2}\big[\psi(-2\hbar k,t)+\psi(2\hbar k,t)\big]\\ i\hbar\partial_t\psi(2\hbar k,t) &= \frac{(2\hbar k)^2}{2m}\psi(2\hbar k,t)+\frac{\hbar\Omega_{\rm R}}{2}\psi(0,t). \end{aligned} \tag{5.432}$$

Adding an energy offset of $-4\hbar\omega_{\rm r}$, the equations become

$$\begin{aligned} i\hbar\partial_t\psi(\pm 2\hbar k,t) &= \frac{\hbar\Omega_{\rm R}}{2}\psi(0,t)\\ i\hbar\partial_t\psi(0,t) &= \frac{\hbar\Omega_{\rm R}}{2}\big[\psi(-2\hbar k,t)+\psi(2\hbar k,t)\big]-4\hbar\omega_{\rm r}\psi(0,t). \end{aligned} \tag{5.433}$$

Now we assume that $\Omega_{\rm R}\ll 4\omega_{\rm r}$, so that the population in the $|p=0\rangle$ state is $O(\Omega_{\rm R}^2/\omega_{\rm r}^2)$ and hence negligible. Additionally, we can make an adiabatic approximation for the evolution of the $|p=0\rangle$ state, by formally setting $\partial_t\psi(0,t)=0$, as we did in Section 5.8.3. Again, though, this is a shortcut for considering the density-matrix picture and replacing the rapidly-varying coherences with their locally average value (although this procedure is a result of coarse-graining here, rather than radiative damping as in the previous treatment). Doing so leads to the adiabatic relation

$$4\omega_{\rm r}\psi(0,t)=\frac{\Omega_{\rm R}}{2}\big[\psi(-2\hbar k,t)+\psi(2\hbar k,t)\big], \tag{5.434}$$

which can be used to eliminate the intermediate state, resulting in a two-level evolution:

$$i\hbar\partial_t\psi(\pm 2\hbar k,t)=\frac{\hbar\Omega_{\rm R}^2}{16\omega_{\rm r}}\big[\psi(\pm 2\hbar k,t)+\psi(\mp 2\hbar k,t)\big]. \tag{5.435}$$

Hence, we see explicitly the Raman-Rabi oscillations between the two motional states (which is *not* the classical pendulum oscillation of the momentum), and the second-order Bragg Rabi frequency is $\Omega_{\rm B,2}=\Omega_{\rm R}^2/8\omega_{\rm r}$. The first term represents a Stark shift of $\Omega_{\rm B,2}/2$, due to scattering processes where the absorbed and emitted photons have the same $\mathbf{k}$, while the second term represents the Rabi-type coupling, where the absorbed and emitted photons have opposite $\mathbf{k}$. Comparing this expression to the form (5.431) for the two-photon Rabi frequency, we see that this second-order Bragg process can be viewed also as a Raman process of two Raman transitions, where the detuning to the intermediate state Δ is identified as $4\omega_{\rm r}$.

Continuing in this manner, the Bragg rate for nth-order scattering from $n\hbar k$ to $-n\hbar k$ is given by

$$\Omega_{{\rm B},n}=\frac{\Omega_{\rm R}^n}{2^{n-1}\displaystyle\prod_{k=1}^{n-1}\delta_k}, \tag{5.436}$$

where δ_k is the detuning of the kth intermediate motional state. Notice that the intermediate detunings are given by $[n^2-(n-2)^2]\omega_{\rm r}$, $[n^2-(n-4)^2]\omega_{\rm r}$, ..., $[n^2-(2-n)^2]\omega_{\rm r}$, so that this Bragg frequency can be written as

$$\Omega_{{\rm B},n}=\frac{\Omega_{\rm R}^n}{(8\omega_{\rm r})^{n-1}[(n-1)!]^2} \tag{5.437}$$

(Bragg transition rate)

The transition frequency obviously becomes small for high-order Bragg processes, as the Rabi frequency decreases exponentially with the order. Nevertheless, Bragg oscillations of up to sixth[35] and eighth[36] order have been observed experimentally for an atomic beam crossing an optical standing wave.

5.8.4 Nonperturbative Analysis

The above analysis of the dipole force was a perturbative treatment in Ω/Δ for the ground-state energy shift. We will now perform a better analysis that gets the (adiabatic) potential correct even for strong excitation, as well as the radiation-pressure force.[37] We start with the Heisenberg-picture force,

$$\mathbf{F}=\partial_t\mathbf{p}=\frac{i}{\hbar}[H,\mathbf{p}]=-\nabla\tilde{H}_{\rm AF}, \tag{5.438}$$

where the last equality follows because $\mathbf{p}=-i\hbar\nabla$. Here, the atomic position $\mathbf{r}$ is in principle an operator, but we will take on the semiclassical view that it refers to the mean atomic position to simplify this treatment, in contrast the preceding perturbative treatment. Again, the rotating-frame interaction Hamiltonian is given by

$$\tilde{H}_{\rm AF}=\frac{\hbar}{2}\left[\Omega^*(\mathbf{r})\,\sigma+\Omega(\mathbf{r})\,\sigma^\dagger\right]. \tag{5.439}$$

Here, we have written the Rabi frequency again as

$$\Omega(\mathbf{r})=-\frac{2\langle {\rm g}|\hat{\varepsilon}\cdot\mathbf{d}|{\rm e}\rangle E_0^{(+)}(\mathbf{r})}{\hbar}=|\Omega(\mathbf{r})|e^{i\phi(\mathbf{r})}, \tag{5.440}$$

where $E_0^{(+)}(\mathbf{r})$ is the positive-rotating part of the field. The spatial dependence includes both any phase rotation as well as slow envelope variations.

[35] David M. Giltner *et al.*, *op. cit.*

[36] Armand Eugéne Albert Koolen, *Dissipative Atom Optics with Cold Metastable Helium Atoms*, Ph.D. thesis, Technische Universiteit Eindhoven (2000).

[37] J. P. Gordon and A. Ashkin, "Motion of atoms in a radiation trap," *Physical Review A* **21**, 1606 (1980) (doi: 10.1103/PhysRevA.21.1606).

The force then depends on the gradient of the Rabi frequency according to

$$\mathbf{F} = -\frac{\hbar}{2}\left[\nabla\Omega^*(\mathbf{r})\,\sigma + \nabla\Omega(\mathbf{r})\,\sigma^\dagger\right]. \tag{5.441}$$

The gradient is given by

$$\nabla\Omega(\mathbf{r}) = (\nabla|\Omega|)\,e^{i\phi} + i(\nabla\phi)|\Omega|e^{i\phi} = \Omega\left(\frac{\nabla|\Omega|}{|\Omega|} + i\nabla\phi\right), \tag{5.442}$$

so we can write

$$\nabla\log\Omega(\mathbf{r}) = \frac{\nabla\Omega(\mathbf{r})}{\Omega(\mathbf{r})} = \nabla\log|\Omega(\mathbf{r})| + i\nabla\phi(\mathbf{r}) \tag{5.443}$$

for the gradient. The mean force then becomes

$$\langle\mathbf{F}\rangle = -\frac{\hbar\Omega^*(\mathbf{r})}{2}\Big(\nabla\log|\Omega(\mathbf{r})| - i\nabla\phi(\mathbf{r})\Big)\langle\sigma\rangle + \text{c.c.} \tag{5.444}$$

There are two terms here. Both go as the interaction energy

$$\hbar\Omega^*(\mathbf{r})\,\langle\sigma\rangle = \left\langle\mathbf{d}^{(+)}\cdot\mathbf{E}^{(-)}\right\rangle \tag{5.445}$$

for the dipole in the external field, but only the first depends on gradients of the field *amplitude*, and this term corresponds to the dipole force. The second term is due to absorption, since it is 90° out of phase with respect to to the dipole force. This is the radiation-pressure force.

In the case where the atom is at rest or moves slowly on time scales of order Γ^{-1}, we can use the steady-state coherence $\langle\sigma\rangle = \tilde{\rho}_{\text{eg}}$. From the solution of the optical Bloch equations [Eq. (5.138)], now accounting for the fact that Ω is no longer necessarily real, the steady-state coherences are

$$\begin{aligned}\tilde{\rho}_{\text{eg}} &= -\frac{i\Omega}{\Gamma}\frac{1+\dfrac{i2\Delta}{\Gamma}}{1+\left(\dfrac{2\Delta}{\Gamma}\right)^2+2\dfrac{|\Omega|^2}{\Gamma^2}} = -\frac{i\Omega}{2(\Gamma/2-i\Delta)(1+s)}\\ \tilde{\rho}_{\text{ge}} &= \frac{i\Omega^*}{2(\Gamma/2+i\Delta)(1+s)},\end{aligned} \tag{5.446}$$

where the saturation parameter is

$$s(\mathbf{r}) = \frac{|\Omega(\mathbf{r})|^2}{2[(\Gamma/2)^2+\Delta^2]}. \tag{5.447}$$

With these relations, the adiabatic mean force is

$$\begin{aligned}\langle\mathbf{F}\rangle &= \frac{i\hbar|\Omega(\mathbf{r})|^2}{4(\Gamma/2-i\Delta)(1+s)}\Big(\nabla\log|\Omega(\mathbf{r})| - i\nabla\phi(\mathbf{r})\Big) + \text{c.c.}\\ &= \frac{\hbar s(\mathbf{r})}{1+s(\mathbf{r})}\left(-\Delta\nabla\log|\Omega(\mathbf{r})| + \frac{\Gamma}{2}\nabla\phi(\mathbf{r})\right)\\ &= \langle\mathbf{F}_{\text{dip}}\rangle + \langle\mathbf{F}_{\text{rad}}\rangle.\end{aligned} \tag{5.448}$$

We can write the second term, the mean radiation pressure, as

$$\langle\mathbf{F}_{\text{rad}}\rangle = \Gamma\rho_{\text{ee}}(\mathbf{r}, t\longrightarrow\infty)\hbar\nabla\phi(\mathbf{r}), \tag{5.449}$$

(radiation-pressure force)

where we used $\rho_{\text{ee}}(t\longrightarrow\infty) = (s/2)/(1+s)$. For a plane wave, where $\phi(\mathbf{r}) = \mathbf{k}\cdot\mathbf{r}$, the radiation-pressure force becomes

$$\langle\mathbf{F}_{\text{rad}}\rangle = \Gamma\rho_{\text{ee}}(\mathbf{r}, t\longrightarrow\infty)\hbar\mathbf{k}. \tag{5.450}$$

The radiation pressure thus has the physically reasonable interpretation of being the photon scattering rate multiplied by the photon recoil momentum. The force is in the direction of the wave vector $\mathbf{k}$. On the other hand, for a standing wave composed of two equal but counterpropagating traveling waves, with a field of the form

$$\begin{aligned}\mathbf{E}(x,t) &= \hat{z}E_0[\cos(kx-\omega t)+\cos(kx+\omega t)] \\ &= \hat{z}E_0\cos(kx)\left(e^{-i\omega t}+e^{i\omega t}\right),\end{aligned} \tag{5.451}$$

the phase $\phi(\mathbf{r})$ is a constant, and the (mean) radiation-pressure force vanishes.

The first term, the mean dipole force, is

$$\langle \mathbf{F}_{\rm dip}\rangle = -\frac{\hbar\Delta\, s(\mathbf{r})}{1+s(\mathbf{r})}\mathrm{Re}\big[\nabla\log|\Omega(\mathbf{r})|\big]. \tag{5.452}$$

This force depends on Δ and the field intensity (via s), and thus on the phase between the applied and dipole radiated fields, giving a dispersive frequency dependence. The force represents a change in the field momentum due to the interference of the radiated and (outgoing) applied fields. The quantum-mechanical interpretation is that the force occurs via coherent scattering processes of absorption and stimulated emission, where the absorbed and emitted photons have different $\mathbf{k}$ vector orientations. The atom therefore recoils to conserve the total atom-field momentum. In a plane wave, there is only one $\mathbf{k}$, and so there is no possibility for changing $\mathbf{k}$ on scattering. There is thus no dipole force in a plane wave: intensity gradients, which are connected to uncertainty in the direction of $\mathbf{k}$, are necessary to produce a dipole force.

Since the gradient of the saturation parameter is

$$\nabla s = \frac{2|\Omega|\nabla|\Omega|}{2[(\Gamma/2)^2+\Delta^2]} = 2s\frac{\nabla|\Omega|}{|\Omega|} = 2s\nabla\log|\Omega|, \tag{5.453}$$

we can write the dipole force in the form

$$\langle \mathbf{F}_{\rm dip}\rangle = -\frac{\hbar\Delta}{2}\frac{\nabla s}{1+s} = -\frac{\hbar\Delta}{2}\nabla\log(1+s) = -\nabla V_{\rm dip}, \tag{5.454}$$

where

$$V_{\rm dip} = \frac{\hbar\Delta}{2}\log[1+s(\mathbf{r})] = \frac{\hbar\Delta}{2}\log\left[1+\frac{|\Omega(\mathbf{r})|^2}{2[(\Gamma/2)^2+\Delta^2]}\right] = \frac{\hbar\Delta}{2}\log\left[1+\frac{I(\mathbf{r})/I_{\rm sat}}{1+4\Delta^2/\Gamma^2}\right]. \tag{5.455}$$

(dipole potential)

This is the main result of this section. This gives the dipole potential for any field intensity, for a stationary atom, or for an atom moving very slowly (as we have used the steady-state solutions of the optical Bloch equations). Notice that while the radiation pressure force saturates for large s, the dipole force can continue to increase without bound, though for large intensities it only does so logarithmically. Also, the dipole potential is a negative shift for red detuning and a positive shift for blue detuning. Exactly on resonance, the dipole force vanishes. Additionally, note that there is no counter-rotating term of the form $(\omega+\omega_0)^{-1}$ as there was in the classical result of Eq. (1.76). The solution here is based on the optical Bloch equations, which assumed the rotating-wave approximation.

Far off resonance, we can expand the logarithm to first order in $s(\mathbf{r})$:

$$V_{\rm dip} \approx \frac{\hbar\Delta}{2}s(\mathbf{r}) \approx \frac{\hbar|\Omega(\mathbf{r})|^2}{4\Delta}. \tag{5.456}$$

(dipole potential, far off resonance)

Thus, we recover our effective dipole potential from our previous perturbative treatment. In terms of the saturation intensity, using $I/I_{\rm sat} = 2\Omega^2/\Gamma^2$ we find the perturbative result in standard form

$$V_{\rm dip} \approx \frac{\hbar\Gamma^2}{8\Delta}\frac{I(\mathbf{r})}{I_{\rm sat}}, \tag{5.457}$$

(dipole potential, far off resonance)

in agreement with the classical expression (1.77) in the same regime.

5.8.5 Dressed-State Interpretation

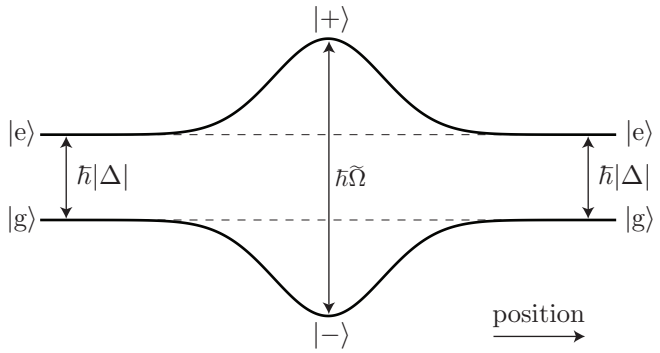
Using the dressed-atom picture, we can obtain more generally valid results for the dipole potential.[38] The dressed-state energies from Eq. (5.64) are

$$E_\pm = -\frac{\hbar\Delta}{2} \pm \frac{\hbar\tilde{\Omega}}{2}, \tag{5.458}$$

where the generalized Rabi frequency is

$$\tilde{\Omega} = \sqrt{|\Omega|^2 + \Delta^2} \tag{5.459}$$

in the case of a complex Rabi frequency. The energies of the dressed states relative to the mean level energy $-\hbar\Delta/2$ are thus $\pm\hbar\tilde{\Omega}/2$. For a Gaussian laser beam, the dressed states shift in opposite directions. For concreteness, assume a large red detuning ($\Delta < 0$), where we can approximately identify $|\mathrm{g}\rangle \approx |-\rangle$. Then a ground-state atom sees a potential well, while an excited-state atom sees a potential barrier.



The ground-state shift is then

$$V_{\mathrm{gnd}} = -\frac{\hbar\tilde{\Omega}}{2} = -\frac{\hbar}{2}|\Delta|\sqrt{1+\frac{|\Omega|^2}{\Delta^2}} \approx -\frac{\hbar}{2}|\Delta|\left(1+\frac{|\Omega|^2}{2\Delta^2}\right) \approx \frac{\hbar\Delta}{2} + \frac{\hbar|\Omega|^2}{4\Delta}. \tag{5.460}$$

The first term is the bare-state energy, while the second term is the lowest-order dipole potential. Of course, this is only valid in the far-detuned limit ($|\Delta| \gg \Gamma, |\Omega|$), where we can assume the atom is approximately in the ground state. For larger excitation, the atom is in a mixture of the two dressed states in steady state, and since the shifts are opposite, the total potential shift is less than this perturbative result indicates. This motivates the logarithmic saturation of the dipole potential with intensity (but see the homework for a more precise interpretation of the dipole potential in terms of dressed states). The sign of the dipole force, which again depends only on the sign of the detuning, is thus explained by which dressed state has the most population. Far to the red of resonance, $|\mathrm{g}\rangle \approx |-\rangle$, and so the shift is negative, while far to the blue, $|\mathrm{g}\rangle \approx |+\rangle$, and so the dipole potential shift is positive.

5.8.6 Fluctuations of the Optical Force

The optical force on the atoms in the standing wave can also lead to momentum diffusion. Part of this diffusion is due to spontaneous emission. The dipole moment of the atom fluctuates due to spontaneous emission, and this fluctuating dipole interacts with the field gradients in the standing wave to produce momentum diffusion. Alternately, you can think of it this way: the atom occasionally jumps to the ground state whenever a photon is emitted (seen by a fictitious photodetector), and then relaxes towards equilibrium. This means that the atom is changing its weight stochastically between the dressed states, which have shifts of opposite signs. Thus, the dipole force the atoms experience is also stochastic.

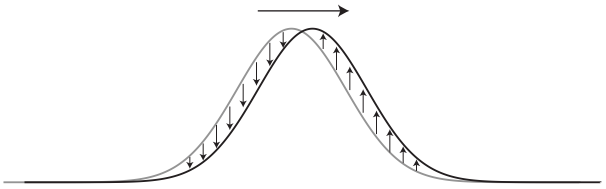
[38] J. Dalibard and C. Cohen-Tannoudji, "Dressed-atom approach to atomic motion in laser light: the dipole force revisited," *Journal of the Optical Society of America B* **2**, 1707 (1985).

5.8.6.1 Fokker–Planck Equation

To handle the effect of fluctuations on the atoms, we will take the semiclassical view of atoms as localized particles on the scale of the potential, and treat the momentum probability density $f(p,t)$ for an ensemble of atoms. In treating the mean force, what we have derived is the **drift coefficient** A for the *advection equation*

$$\partial_t f(p,t) = -A\partial_p f(p,t). \tag{5.461}$$

The solution to this equation is simply $f(p - At)$, and thus this equation simply represents *translation* of the momentum distribution by At. Thus, the drift coefficient A is the mean force on the atoms. We can see the effect of the right-hand-side on the distribution by visualizing its derivatives on a simple distribution.

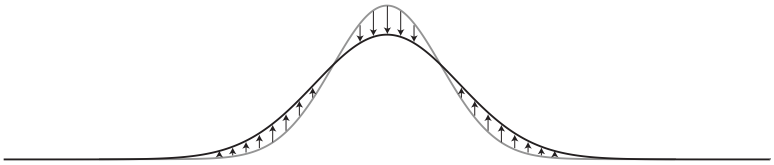


On the left-hand side, where the derivative is positive, the advection term causes the function to decrease, assuming $A > 0$. On the right-hand side, the opposite is true. The net effect is motion of the distribution to the right.

To treat diffusion, we use a second-order term to obtain the **diffusion equation**

$$\partial_t f(p,t) = \frac{D}{2}\partial_p^2 f(p,t) \tag{5.462}$$

with **diffusion coefficient** D. The diffusion term causes the distribution to spread, which we can again see by visualizing the derivatives for a smooth, single-peaked distribution.



For a positive diffusion coefficient, the second derivative is negative at the peak but positive in the wings, resulting in a net spreading. The diffusion equation has the Gaussian solution, assuming an initial condition of $\delta(x - x_0)$, of

$$f(p,t) = \frac{1}{\sqrt{2\pi Dt}} \exp\left[-\frac{1}{2}\frac{(x-x_0)^2}{Dt}\right], \tag{5.463}$$

which has variance Dt and width $\sqrt{Dt}$. Note that under this evolution, even non-Gaussian initial conditions become asymptotically Gaussian, as we expect for a random-walk process. This is because the Gaussian solution is the solution for the delta-function initial condition, and thus the solution at time t for a general initial condition is the convolution of the initial condition with the above Gaussian. At late times, when the Gaussian solution is much broader than the initial condition, the contribution of the initial condition to the solution is negligible.

Combining these two effects, we arrive at a simple advection-diffusion equation

$$\partial_t f(p,t) = -A\partial_p f(p,t) + \frac{D}{2}\partial_p^2 f(p,t). \tag{5.464}$$

In the more general, one-dimensional case, the advection and diffusion coefficients can depend on the momentum itself

$$\partial_t f(p,t) = -\partial_p A(p) f(p,t) + \frac{1}{2}\partial_p^2 D(p) f(p,t). \tag{5.465}$$

(Fokker–Planck equation)

This equation is the one-dimensional **Fokker–Planck equation**. Again, note that we have computed the advection coefficient for the optical force, although thus far we have ignored its momentum dependence. In

the following sections, we will be concerned with computing the diffusion coefficient, also ignoring its velocity dependence. The simplest velocity dependent case that we would like to consider for laser cooling is the linear case of $A(p) = Ap$ and $D(p) = D$, where we can write down the Gaussian solution[39]

$$f(p,t) = \frac{1}{\sqrt{2\pi(D/2A)(e^{2At}-1)}} \exp\left[-\frac{1}{2}\frac{(p-p_0e^{At})^2}{(D/2A)(e^{2At}-1)}\right]. \tag{5.466}$$

For $A > 0$, the system is unstable and runs away, while for $A < 0$, the solution settles down to the steady state Gaussian centered at $p = 0$ and width $\sqrt{D/2|A|}$:

$$f(p, t \longrightarrow \infty) = \frac{1}{\sqrt{2\pi D/2|A|}} \exp\left[-\frac{1}{2}\frac{p^2}{D/2|A|}\right]. \tag{5.467}$$

(steady-state, linear solution)

This, of course, is the problem of laser cooling with intrinsic noise, as we discussed when deriving the Doppler limit of Section 1.4.3.1.

In the most general case, the Fokker–Planck equation in three dimensions is (note the implied summations)

$$\partial_t f(\mathbf{p},t) = -\frac{\partial}{\partial p_\alpha} A_\alpha(\mathbf{p}) f(\mathbf{p},t) + \frac{1}{2}\frac{\partial^2}{\partial p_\alpha \partial p_\beta} D_{\alpha\beta}(\mathbf{p}) f(\mathbf{p},t). \tag{5.468}$$

However, we will only be concerned with the total diffusion rate $D_{\alpha\alpha}$, rather than with the anisotropic components of the diffusion tensor $D_{\alpha\beta}$.

5.8.6.2 Diffusion Coefficient

For the calculation of the diffusion coefficient due to optical forces, we will again assume an atom at rest (or slowly moving), and we will also assume that the atom is spatially localized. Based on our above discussion, we may take the diffusion coefficient (henceforth denoted by D_p) to be defined by the rate at which the momentum variance increases:

$$\begin{aligned} D_p &= \partial_t V_p \\ &= \partial_t \left(\langle p^2 \rangle - \langle p \rangle^2 \right) \\ &= \langle \mathbf{p} \cdot \mathbf{F} \rangle + \langle \mathbf{F} \cdot \mathbf{p} \rangle - 2 \langle \mathbf{p} \rangle \cdot \langle \mathbf{F} \rangle , \end{aligned} \tag{5.469}$$

where we have used $\mathbf{F} = \partial_t \mathbf{p}$. Then expressing $\mathbf{p}$ as the time integral of $\mathbf{F}$,

$$\begin{aligned} D_p &= \int_{-\infty}^{t} dt' \left[\langle \mathbf{F}(t') \cdot \mathbf{F}(t) \rangle + \langle \mathbf{F}(t) \cdot \mathbf{F}(t') \rangle - 2 \langle \mathbf{F}(t') \rangle \cdot \langle \mathbf{F}(t) \rangle \right] \\ &= \int_{-\infty}^{0} d\tau \left[\langle \mathbf{F}(t+\tau) \cdot \mathbf{F}(t) \rangle + \langle \mathbf{F}(t) \cdot \mathbf{F}(t+\tau) \rangle - 2 \langle \mathbf{F}(t) \rangle \cdot \langle \mathbf{F}(t+\tau) \rangle \right], \end{aligned} \tag{5.470}$$

where we have set $t' = t + \tau$ and assumed stationarity of the force. Thus, we can write

$$D_p = \int_{-\infty}^{\infty} d\tau \left[\langle \mathbf{F}(t) \cdot \mathbf{F}(t+\tau) \rangle - \langle \mathbf{F}(t) \rangle^2 \right]. \tag{5.471}$$

(diffusion coefficient)

Now recalling from Eq. (5.441) that

$$\mathbf{F} = -\frac{\hbar}{2} \left[\nabla\Omega^*(\mathbf{r})\, \sigma + \nabla\Omega(\mathbf{r})\, \sigma^\dagger \right], \tag{5.472}$$

[39] H. J. Carmichael, *Statistical Methods in Quantum Optics 1: Master Equations and Fokker–Planck Equations* (Springer, 1999), p. 148.

we can write

$$\langle \mathbf{F}(t)\cdot\mathbf{F}(t+\tau)\rangle = \frac{\hbar^2}{4}\Big\{|\nabla\Omega(\mathrm{r})|^2\Big[\left\langle\sigma^\dagger(t)\sigma(t+\tau)\right\rangle + \left\langle\sigma(t)\sigma^\dagger(t+\tau)\right\rangle\Big] + [\nabla\Omega(\mathbf{r})]^2\left\langle\sigma^\dagger(t)\sigma^\dagger(t+\tau)\right\rangle + [\nabla\Omega^*(\mathbf{r})]^2\left\langle\sigma(t)\sigma(t+\tau)\right\rangle\Big\}. \tag{5.473}$$

Had we treated the field quantum-mechanically here, there would be an extra term describing diffusion due to spontaneous emission.[40] However, this is missing in our semiclassical treatment, and we will treat that effect separately below.

5.8.6.3 Quantum Regression Theorem

Thus, we will need integrals of two-time averages of the form

$$\int_{-\infty}^{\infty} d\tau\left[\left\langle\sigma^\dagger(t)\sigma(t+\tau)\right\rangle - |\left\langle\sigma\right\rangle|^2\right]. \tag{5.474}$$

To do this, we will use the alternate form of the quantum regression theorem (Section 5.7.3.1): given that one-time average $\langle\sigma(t)\rangle$ has a solution of the form

$$\langle\sigma(t)\rangle = g_0(t) + g_1(t)\langle\sigma(0)\rangle + g_2(t)\left\langle\sigma^\dagger(0)\right\rangle + g_3(t)\langle\sigma_z(0)\rangle, \tag{5.475}$$

which it must have to be completely determined by the initial quantum state, it follows from the quantum regression theorem that the two-time average has the similar solution

$$\left\langle\sigma^\dagger(t)\sigma(t+\tau)\right\rangle = g_0(\tau)\left\langle\sigma^\dagger\right\rangle + g_1(\tau)\left\langle\sigma^\dagger\sigma\right\rangle + g_2(\tau)\left\langle\sigma^\dagger\sigma^\dagger\right\rangle + g_3(\tau)\left\langle\sigma^\dagger\sigma_z\right\rangle, \tag{5.476}$$

where the expectation values are taken in steady state. Using $(\sigma^\dagger)^2 = 0$ and $\sigma^\dagger\sigma_z = -\sigma^\dagger$,

$$\left\langle\sigma^\dagger(t)\sigma(t+\tau)\right\rangle = [g_0(\tau) - g_3(\tau)]\tilde{\rho}_{\mathrm{ge}}(t\longrightarrow\infty) + g_1(\tau)\rho_{\mathrm{ee}}(t\longrightarrow\infty). \tag{5.477}$$

Similarly, we will need the correlation function $\left\langle\sigma(t)\sigma^\dagger(t+\tau)\right\rangle$, which should have a similar solution as

$$\left\langle\sigma^\dagger(t)\right\rangle = g_0^*(t) + g_1^*(t)\left\langle\sigma^\dagger(0)\right\rangle + g_2^*(t)\langle\sigma(0)\rangle + g_3^*(t)\langle\sigma_z(0)\rangle, \tag{5.478}$$

in which case we note that to get the right steady state, we replace $g_0^*(t)\longrightarrow g_0^*(\tau)\langle\sigma\rangle$, and for the other coefficients we replace $\langle C(0)\rangle\longrightarrow\langle\sigma C\rangle$ to obtain

$$\begin{aligned}\left\langle\sigma(t)\sigma^\dagger(t+\tau)\right\rangle &= g_0^*(\tau)\left\langle\sigma^\dagger\right\rangle + g_1^*(\tau)\left\langle\sigma\sigma^\dagger\right\rangle + g_2^*(\tau)\langle\sigma\sigma\rangle + g_3^*(\tau)\langle\sigma\sigma_z\rangle\\ &= [g_0^*(\tau) + g_3^*(\tau)]\tilde{\rho}_{\mathrm{eg}}(t\longrightarrow\infty) + g_1^*(\tau)\rho_{\mathrm{gg}}(t\longrightarrow\infty).\end{aligned} \tag{5.479}$$

Finally, for the two remaining correlation functions, we have

$$\begin{aligned}\left\langle\sigma^\dagger(t)\sigma^\dagger(t+\tau)\right\rangle &= g_0^*(\tau)\left\langle\sigma^\dagger\right\rangle + g_1^*(\tau)\left\langle\sigma^\dagger\sigma^\dagger\right\rangle + g_2^*(\tau)\left\langle\sigma^\dagger\sigma\right\rangle + g_3^*(\tau)\left\langle\sigma^\dagger\sigma_z\right\rangle\\ &= [g_0^*(\tau) - g_3^*(\tau)]\tilde{\rho}_{\mathrm{ge}}(t\longrightarrow\infty) + g_2^*(\tau)\rho_{\mathrm{ee}}(t\longrightarrow\infty)\end{aligned} \tag{5.480}$$

and

$$\begin{aligned}\langle\sigma(t)\sigma(t+\tau)\rangle &= g_0(\tau)\langle\sigma\rangle + g_1(\tau)\langle\sigma\sigma\rangle + g_2(\tau)\left\langle\sigma\sigma^\dagger\right\rangle + g_3(\tau)\langle\sigma\sigma_z\rangle\\ &= [g_0(\tau) + g_3(\tau)]\tilde{\rho}_{\mathrm{eg}}(t\longrightarrow\infty) + g_2(\tau)\rho_{\mathrm{gg}}(t\longrightarrow\infty).\end{aligned} \tag{5.481}$$

[40] J. P. Gordon and A. Ashkin, *op. cit.*

Now, to carry out the appropriate integrals over the correlation functions, we can write

$$\begin{aligned}
\int_{-\infty}^{\infty} d\tau \left[\left\langle \sigma^\dagger(t)\sigma(t+\tau)\right\rangle - |\left\langle\sigma\right\rangle|^2\right] &= (G_0 - G_3)\tilde{\rho}_{\rm ge}(t\longrightarrow\infty) + G_1\rho_{\rm ee}(t\longrightarrow\infty) + \text{c.c.} \\
\int_{-\infty}^{\infty} d\tau \left[\left\langle \sigma(t)\sigma^\dagger(t+\tau)\right\rangle - |\left\langle\sigma\right\rangle|^2\right] &= (G_0^* + G_3^*)\tilde{\rho}_{\rm eg}(t\longrightarrow\infty) + G_1^*\rho_{\rm gg}(t\longrightarrow\infty) + \text{c.c.} \\
\int_{-\infty}^{\infty} d\tau \left[\left\langle \sigma^\dagger(t)\sigma^\dagger(t+\tau)\right\rangle - \left\langle\sigma^\dagger\right\rangle^2\right] &= (G_0^* - G_3^*)\tilde{\rho}_{\rm ge}(t\longrightarrow\infty) + G_2^*\rho_{\rm ee}(t\longrightarrow\infty) + \text{c.c.} \\
\int_{-\infty}^{\infty} d\tau \left[\left\langle \sigma(t)\sigma(t+\tau)\right\rangle - \left\langle\sigma\right\rangle^2\right] &= (G_0 + G_3)\tilde{\rho}_{\rm eg}(t\longrightarrow\infty) + G_2\rho_{\rm gg}(t\longrightarrow\infty) + \text{c.c.}
\end{aligned} \tag{5.482}$$

where

$$G_\alpha := \int_0^\infty d\tau \left[g_\alpha(\tau) - \left\langle\sigma\right\rangle \delta_{\alpha 0}\right], \tag{5.483}$$

and noting that $g_\alpha(-\tau) = g_\alpha^*(\tau)$, so that the G_α are real. Note that we subtract the dc amplitude in G_0, which implements the dc subtraction in Eqs. (5.474). In terms of these integrals, we can use Eqs. (5.482) in Eqs. (5.471) and (5.473) to write

$$\begin{aligned}
D_p = \frac{\hbar^2}{4}\Bigg\{ &|\nabla\Omega(\mathrm{r})|^2\Big[(G_0\tilde{\rho}_{\rm ge} + \text{c.c.}) - (G_3\tilde{\rho}_{\rm ge} - \text{c.c.}) + G_1\Big] \\
&+ [\nabla\Omega(\mathbf{r})]^2\left[(G_0^* - G_3^*)\tilde{\rho}_{\rm ge} + G_2^*\rho_{\rm ee}\right] + [\nabla\Omega^*(\mathbf{r})]^2\left[(G_0 + G_3)\tilde{\rho}_{\rm eg} + G_2\rho_{\rm gg}\right]\Bigg\} + \text{c.c.}
\end{aligned} \tag{5.484}$$

But D_p is real by construction, so we can explicitly drop any imaginary terms with the result

$$D_p = \frac{\hbar^2}{2}\mathrm{Re}\Big\{|\nabla\Omega(\mathrm{r})|^2[2G_0\tilde{\rho}_{\rm ge}(t\longrightarrow\infty) + G_1] + [\nabla\Omega^*(\mathbf{r})]^2\left[2G_0\tilde{\rho}_{\rm eg}(t\longrightarrow\infty) + G_2\right]\Big\}, \tag{5.485}$$

so that all that remains is to evaluate the integrals G_α.

To do these integrals, note that given a function $f(t)$ and its Laplace transform $\mathscr{L}[f](s)$, we have

$$\int_0^\infty dt\, f(t) = \mathscr{L}[f](0), \tag{5.486}$$

so that all we have to do is to compute Laplace transforms and evaluate them at $s = 0$ (s being the dummy variable of the Laplace transform, not the saturation parameter). Now our equations of motion can be written

$$\partial_t \begin{bmatrix} \delta\tilde{\rho}_{\rm eg} \\ \delta\tilde{\rho}_{\rm ge} \\ \langle\delta\sigma_z\rangle \end{bmatrix} = \begin{bmatrix} -\frac{\Gamma}{2} + i\Delta & 0 & i\frac{\Omega}{2} \\ 0 & -\frac{\Gamma}{2} - i\Delta & -i\frac{\Omega^*}{2} \\ i\Omega^* & -i\Omega & -\Gamma \end{bmatrix} \begin{bmatrix} \delta\tilde{\rho}_{\rm eg} \\ \delta\tilde{\rho}_{\rm ge} \\ \langle\delta\sigma_z\rangle \end{bmatrix} =: \mathbf{P} \begin{bmatrix} \delta\tilde{\rho}_{\rm eg} \\ \delta\tilde{\rho}_{\rm ge} \\ \langle\delta\sigma_z\rangle \end{bmatrix}. \tag{5.487}$$

Computing the Laplace transform, we find

$$\begin{bmatrix} \mathscr{L}[\delta\tilde{\rho}_{\rm eg}](s) \\ \mathscr{L}[\delta\tilde{\rho}_{\rm ge}](s) \\ \mathscr{L}[\langle\delta\sigma_z\rangle](s) \end{bmatrix} = \frac{1}{s - \mathbf{P}} \begin{bmatrix} \delta\tilde{\rho}_{\rm eg}(0) \\ \delta\tilde{\rho}_{\rm ge}(0) \\ \langle\delta\sigma_z(0)\rangle \end{bmatrix}. \tag{5.488}$$

We only need the Laplace transform at $s = 0$, so it suffices to compute

$$\left.\frac{1}{s-\mathbf{P}}\right|_{s=0} = \frac{1}{-\mathbf{P}} = \frac{1}{4\Gamma[\Gamma^2/4+\Delta^2+|\Omega|^2/2]}\begin{bmatrix} 2|\Omega|^2+2\Gamma(\Gamma+2i\Delta) & 2\Omega^2 & -\Omega(2\Delta-i\Gamma) \\ 2[\Omega^*]^2 & 2|\Omega|^2+2\Gamma(\Gamma-2i\Delta) & -\Omega^*(2\Delta+i\Gamma) \\ -2\Omega^*(2\Delta-i\Gamma) & -2\Omega(2\Delta+i\Gamma) & \Gamma^2+4\Delta^2 \end{bmatrix}. \tag{5.489}$$

Then, since the zero-frequency component of the Laplace transform has the form

$$\begin{aligned}\mathscr{L}[\delta\tilde{\rho}_{\mathrm{eg}}](0) &= \mathscr{L}[\langle\delta\sigma\rangle](0) \\ &= c_1\,\delta\tilde{\rho}_{\mathrm{eg}}(0) + c_2\,\delta\tilde{\rho}_{\mathrm{ge}}(0) + c_3\,\langle\delta\sigma_z(0)\rangle \\ &= [-c_1\,\tilde{\rho}_{\mathrm{eg}}(t\longrightarrow\infty) - c_2\,\tilde{\rho}_{\mathrm{ge}}(t\longrightarrow\infty) - c_3\,\langle\sigma_z(t\longrightarrow\infty)\rangle] + c_1\tilde{\rho}_{\mathrm{eg}}(0) + c_2\tilde{\rho}_{\mathrm{ge}}(0) + c_3\,\langle\sigma_z(0)\rangle\,,\end{aligned} \tag{5.490}$$

we can compare to Eq. (5.475) to identify $G_1 = c_1$, $G_2 = c_2$, $G_3 = c_3$, and

$$G_0 = -G_1\tilde{\rho}_{\mathrm{eg}}(t\longrightarrow\infty) - G_2\tilde{\rho}_{\mathrm{ge}}(t\longrightarrow\infty) - G_3\,\langle\sigma_z(t\longrightarrow\infty)\rangle\,. \tag{5.491}$$

We thus obtain

$$\begin{aligned} G_1 &= \frac{\Gamma + s(\mathbf{r})(\Gamma/2-i\Delta)}{\Gamma[1+s(\mathbf{r})](\Gamma/2-i\Delta)} \\ G_2 &= \frac{\Omega^2}{2\Gamma[(\Gamma/2)^2+\Delta^2][1+s(\mathbf{r})]} \\ G_3 &= \frac{i\Omega}{2\Gamma(\Gamma/2-i\Delta)[1+s(\mathbf{r})]}, \end{aligned} \tag{5.492}$$

and using the steady-state values

$$\begin{aligned} \tilde{\rho}_{\mathrm{eg}}(t\longrightarrow\infty) &= \frac{-i\Omega}{2(\Gamma/2-i\Delta)[1+s(\mathbf{r})]} \\ \rho_{\mathrm{ee}}(t\longrightarrow\infty) &= \frac{s(\mathbf{r})/2}{1+s(\mathbf{r})}, \end{aligned} \tag{5.493}$$

we find

$$G_0 = -\frac{i\Omega}{2[1+s(\mathbf{r})]^2[(\Gamma/2)^2+\Delta^2]}\left[\frac{\Gamma/2-i\Delta}{\Gamma} - \frac{\Gamma}{\Gamma/2-i\Delta} - \frac{i2s\Delta}{\Gamma}\right]. \tag{5.494}$$

Then, putting these integrals into Eq. (5.485), we find our main result

$$\begin{aligned} D_p = {}& \frac{\hbar^2\Gamma}{2}\left(\frac{\nabla s}{2s}\right)^2 \frac{s}{(1+s)^3}\left[1+\left(\frac{\Gamma^2}{(\Gamma/2)^2+\Delta^2}-1\right)s+3s^2+4\frac{(\Gamma/2)^2+\Delta^2}{\Gamma^2}s^3\right] \\ &+\frac{\hbar^2\Gamma}{2}(\nabla\phi)^2\frac{s}{(1+s)^3}\left[1+\left(3-\frac{\Gamma^2}{(\Gamma/2)^2+\Delta^2}\right)s+s^2\right] \\ &+\hbar^2\Delta\left(\frac{\nabla s\cdot\nabla\phi}{s}\right)\frac{s^2}{(1+s)^3}\left[\frac{\Gamma^2}{(\Gamma/2)^2+\Delta^2}+s\right]. \end{aligned}$$

(diffusion coefficient) (5.495)

This expression is somewhat cumbersome, and so we will examine it in the limits of low and high intensity.

5.8.6.4 Interpretation of the Diffusion Rate

In the low-intensity limit, we find that only the first two terms contribute:

$$D_p \approx \frac{\hbar^2\Gamma s}{2}\left[\left(\frac{\nabla s}{2s}\right)^2 + (\nabla\phi)^2\right] + O(s^2).$$

(diffusion coefficient, low-intensity limit) (5.496)

The first responds to the gradient of the field, and we can see explicitly now that the effect here is due to the fluctuations of the atomic dipole interacting with the field gradients. Hence, this effect is referred to as the **stochastic-dipole force**. The other term, which depends on the phase gradient is due to photon absorption; we will interpret this effect more carefully below when we add spontaneous emission. In the high-intensity limit, the stochastic-dipole force dominates so long as $\nabla s \neq 0$, in which case the diffusion rate becomes

$$D_p \approx \frac{\hbar^2|\Omega|^2}{\Gamma}\left(\frac{\nabla s}{2s}\right)^2. \qquad \text{(diffusion coefficient, high-intensity limit)} \quad (5.497)$$

As was the case for the *mean* force, the absorption contribution saturates, whereas the dipole contribution does not. Note that D_p increases like s for very large intensities, while $V_{\rm dip}$ increases only as $\log s$, so that for a nearly conservative and long-lived trap it is not wise to use a very large saturation parameter.

5.8.6.5 Dressed-State Model

In the high-intensity limit, we can understand the diffusion simply in terms of the dressed states.[41] In this limit, the dressed states $|\pm\rangle$ are approximately equal superpositions of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, and vice versa. Each spontaneous-emission event projects the atom into the ground state, and thus into an equal superposition of the dressed states. The atom therefore sees both dressed-state shifts $\pm\hbar\tilde{\Omega}/2 \approx \pm\hbar|\Omega|/2$. We can interpret this as follows: after each spontaneous-emission event, the atom sees a force

$$\mathbf{F} = \mp\frac{\hbar\nabla|\Omega|}{2} = \mp\frac{\hbar|\Omega|}{2}\left(\frac{\nabla s}{2s}\right), \quad (5.498)$$

where the sign is chosen randomly, but with equal probability for the two possibilities. Assuming the atom is moving slowly, even after accumulating momentum, the momentum change associated with a single spontaneous-emission event is

$$\Delta p = \mp\frac{\hbar|\Omega|}{2}\left(\frac{\nabla s}{2s}\right)\xi \quad (5.499)$$

where ξ is the time until the next spontaneous-emission event, which is a random variable ($\xi > 0$) of mean $2/\Gamma$ and exponential probability density

$$f(\xi) = \frac{\Gamma}{2}\exp\left(-\frac{\Gamma}{2}\xi\right). \quad (5.500)$$

To take into account the randomness of the sign, we can write

$$\Delta p = \frac{\hbar|\Omega|}{2}\left(\frac{\nabla s}{2s}\right)\xi' \quad (5.501)$$

where $\xi' \in \mathbb{R}$ has a two-sided exponential probability density

$$f_\pm(\xi') = \frac{\Gamma}{4}\exp\left(-\frac{\Gamma}{2}|\xi'|\right). \quad (5.502)$$

Then the mean-square kick is

$$\left\langle(\Delta p)^2\right\rangle = \frac{\hbar^2|\Omega|^2}{4}\left(\frac{\nabla s}{2s}\right)^2\left\langle\xi'^2\right\rangle = \frac{2\hbar^2|\Omega|^2}{\Gamma^2}\left(\frac{\nabla s}{2s}\right)^2, \quad (5.503)$$

where

$$\left\langle\xi'^2\right\rangle = \frac{\Gamma}{4}\int_{-\infty}^{\infty} d\xi'\,\exp\left(-\frac{\Gamma}{2}|\xi'|\right)\xi'^2 = \frac{8}{\Gamma^2}. \quad (5.504)$$

[41] J. P. Gordon and A. Ashkin, *op. cit.*; see also J. Dalibard and C. Cohen-Tannoudji, *op. cit.*

The diffusion rate is the mean-square step divided by the average step time $\Delta t = 2/\Gamma$, so

$$D_p = \frac{\langle (\Delta p)^2 \rangle}{\Delta t} = \frac{\hbar^2 |\Omega|^2}{\Gamma} \left(\frac{\nabla s}{2s} \right)^2, \tag{5.505}$$

which is precisely what we obtained from the full calculation in this limit.

5.8.6.6 Examples: Plane and Standing Waves

To gain further insight, let's also consider a couple of concrete examples. For a plane wave, $\nabla s = 0$ and $\nabla \phi = \mathbf{k}$, so that only the absorption contribution remains:

$$D_p = \frac{\hbar^2 k^2 \Gamma}{2} \frac{s}{(1+s)^3} \left[1 + \left(3 - \frac{\Gamma^2}{(\Gamma/2)^2 + \Delta^2} \right) s + s^2 \right].$$

(diffusion coefficient, plane wave) (5.506)

For small intensity, this expression becomes

$$D_p \approx \hbar^2 k^2 \Gamma \frac{s}{2} = \hbar^2 k^2 \Gamma \rho_{\mathrm{ee}}(t \longrightarrow \infty).$$

(diffusion coefficient, plane wave, low intensity) (5.507)

As we will see below, this is what we expect for the diffusion due to photon recoils of $\hbar k$ at an average rate $\Gamma \rho_{\mathrm{ee}}(t \longrightarrow \infty)$. However, this is not due to spontaneous emission itself, which we have not yet accounted for, but rather the *absorption* of photons that later result in spontaneous emission events. This conclusion also applies in the high-intensity limit:

$$D_p \approx \frac{\hbar^2 k^2 \Gamma}{2} = \hbar^2 k^2 \Gamma \rho_{\mathrm{ee}}(t \longrightarrow \infty).$$

(diffusion coefficient, plane wave, high intensity) (5.508)

For intermediate intensities, we can see that there are other correction factors in Eq. (5.506). These corrections have been shown to be related to the non-Poissonian character of spontaneous emission.[42] This is related to the antibunching that we already examined; spontaneous emission becomes Poissonian, such that the second-order coherence function $g^{(2)}(\tau) \approx 1$ for all τ, in the limits of low and high intensity.

The other example we will consider is a standing wave of light, where we can take $\nabla \phi = 0$ and $\Omega = \Omega_0 \cos kx$. Note that we are taking $|\Omega| \longrightarrow |\Omega_0| \cos kx$, so we do not have to deal with the sign of the Rabi frequency using the phase ϕ. Thus,

$$\frac{\nabla s}{2s} = \frac{\nabla |\Omega|}{|\Omega|} = -\mathbf{k} \tan kx. \tag{5.509}$$

Only the dipole part contributes, so that the diffusion rate becomes

$$D_p = 2\hbar^2 k^2 \Gamma \frac{\Omega_0^2 \sin^2 kx}{[2(\Delta^2 + \Gamma^2/4) + \Omega_0^2 \cos^2 kx]^3} \times \left[\left(\Delta^2 + \frac{\Gamma^2}{4} \right)^2 + \left(\frac{3}{4}\Gamma^2 - \Delta^2 \right) \Omega_0^2 \cos^2 kx + \frac{3}{4} \Omega_0^4 \cos^4 kx + \frac{\Omega_0^6}{2\Gamma^2} \cos^6 kx \right].$$

(diffusion coefficient, standing wave) (5.510)

For low intensities, this becomes

$$D_p \approx \frac{\hbar^2 k^2 \Gamma s_0}{2} \sin^2 kx,$$

(diffusion coefficient, standing wave, low intensity) (5.511)

[42] Richard J. Cook, "Photon number statistics in resonance fluorescence," *Physical Review A* **23**, 1243 (1981) (doi: 10.1103/PhysRevA.23.1243); Stig Stenholm, "Distribution of photons and atomic momentum in resonance fluorescence," *Physical Review A* **27**, 2513 (1981) (doi: 10.1103/PhysRevA.27.2513).

where

$$s_0 = \frac{\Omega_0^2}{2[(\Gamma/2)^2 + \Delta^2]}. \tag{5.512}$$

We thus see that in this regime, the stochastic-dipole force is maximum where the *gradients* are maximum, which is where the *intensity* is *minimum*. This force is largest precisely where we expect spontaneous-emission noise to be smallest, and vice versa. It turns out that we can interpret the diffusion in this regime as due to stimulated absorption, as we will show below. In the large-intensity limit, the diffusion rate reduces to

$$D_p = \frac{\hbar^2 k^2 \Omega_0^2}{\Gamma} \sin^2 kx, \tag{5.513}$$

(diffusion coefficient, standing wave, high intensity)

which has the same spatial dependence as the small-field case.

We should reiterate here that in writing down these diffusion rates, we have assumed nearly zero atomic velocity and ignored the velocity dependences of the diffusion rates. This is because we have used local values for the internal atomic variables, which is only valid if the atom really is localized, or if we are in the perturbative regime.

5.8.6.7 Spontaneous Emission

In our discussion of the force fluctuations, we have ignored spontaneous emission because we have used a semiclassical field, and we did not use the Bloch equations in the form (5.418) that include the recoil kick in the dissipation terms. Thus, we will need to put it in by hand.

Note that there are two effects on the atomic motion that one might consider: the absorption of the photon and the emission. As we noted above, fluctuations due to absorption are already included in the preceding analysis, but it is worth examining this in a bit more depth. In the perturbative regime, recall that the excited- and ground-state amplitudes are related from Eq. (5.412) by

$$\psi_\mathrm{e} = \frac{\Omega(x)}{2\Delta}\psi_\mathrm{g}. \tag{5.514}$$

Thus, the excited state has an "imprint" of the field profile when compared to the ground state. For concreteness consider the standing-wave case $\Omega(x) = \Omega_0 \cos kx$, so that

$$\psi_\mathrm{e} = \frac{\Omega_0}{2\Delta} \cos kx\, \psi_\mathrm{g}. \tag{5.515}$$

On a spontaneous-emission event, the atomic annihilation operator σ is applied to the atomic state vector (along with the recoil operator for the emission, which we will not consider for the moment). Then the post-emission state is

$$\begin{aligned} \langle x|\sigma|\psi\rangle &= \sigma\left[\psi_\mathrm{e}(x)|\mathrm{e}\rangle + \psi_\mathrm{g}(x)|\mathrm{g}\rangle\right] \\ &\propto \psi_\mathrm{e}(x)|\mathrm{g}\rangle \\ &\propto \cos kx\, \psi_\mathrm{g}(x)|\mathrm{g}\rangle \\ &\propto \left(e^{ikx} + e^{-ikx}\right)\psi_\mathrm{g}(x)|\mathrm{g}\rangle. \end{aligned} \tag{5.516}$$

Thus, the atom is in a superposition of having recoiled by one photon momentum in each direction along the standing wave, due to the indistinguishable possibilities of having absorbed a photon from either traveling wave. Notice also that the disturbance to the wave function due to absorption (multiplication by $\cos kx$) is minimal at the extrema of the cosine function (where the cosine is approximately constant), but there is most disturbance on the gradients of the cosine function: the gradients induced by the absorption represent the added momentum. This explains the $\sin^2 kx$ dependence of the diffusion rate (5.511) in the low-intensity limit, since $\sin^2 kx$ is maximum precisely where the derivatives of $\cos kx$ are most extreme. In the general case, it is somewhat difficult to pin down the absorption effects that contribute to force fluctuations, but in principle we have already accounted for them.

It is easy to account for the diffusion due to spontaneous emission. The photon scattering rate is

$$R_{\rm sc} = \frac{\Gamma}{2}\frac{s(\mathbf{r})}{1+s(\mathbf{r})} = \frac{(\Gamma/2)\Omega^2(\mathbf{r})}{2(\Delta^2+\Gamma^2/4)+\Omega^2(\mathbf{r})}. \tag{5.517}$$

The momentum recoils from the emitted photons, as we saw in the Lorentz atom treatment, result in atomic momentum diffusion at the rate

$$D_p^{(\rm se)} = \hbar^2 k^2 R_{\rm sc} = \frac{\hbar^2 k^2 \Gamma}{2}\frac{s}{1+s} = \frac{\hbar^2 k^2 \Gamma}{2}\frac{\Omega^2(\mathbf{r})}{2(\Delta^2+\Gamma^2/4)+\Omega^2(\mathbf{r})},$$

(diffusion coefficient, spontaneous emission) (5.518)

which is simply the mean-square momentum kick for one photon recoil, $\hbar^2 k^2$, multiplied by the scattering rate. Again, the momentum diffusion coefficient D_p is defined such that $V_p = \langle (p - \langle p \rangle)^2 \rangle$ grows asymptotically as $D_p t$. The form here follows from the fact that photons are emitted into random directions, and thus the resulting photon recoils cause the atom to execute a random walk in momentum space.

The specific example of the plane wave is trivial here, since Ω is just a constant. For the standing wave, the diffusion rate becomes

$$D_p^{(\rm se)} = \frac{\hbar^2 k^2 \Gamma}{2}\frac{\Omega_0^2 \cos^2 kx}{2(\Delta^2+\Gamma^2/4)+\Omega_0^2\cos^2 kx},$$

(diffusion coefficient, spontaneous emission, standing wave) (5.519)

which for small intensities becomes

$$D_p^{(\rm se)} \approx \frac{\hbar^2 k^2 \Gamma}{2} s_0 \cos^2 kx,$$

(diffusion coefficient, spontaneous emission, standing wave, low intensity) (5.520)

which has the same form as Eq. (5.511), but with $\cos^2 kx$ instead of $\sin^2 kx$. When added together, we see that the total diffusion rate (including both absorption and stimulated-emission contributions) becomes independent of position.

If we restrict our attention to a single dimension, then the one-dimensional diffusion rate is

$$D_p^{(\rm se,1)} = \frac{\hbar^2 k^2 \Gamma \zeta^2}{2}\frac{\Omega^2(x)}{2(\Delta^2+\Gamma^2/4)+\Omega^2(x)}, \tag{5.521}$$

where ζ^2 is the mean-square projection of the photon recoil along the direction of the standing wave:

$$\zeta^2 = \int d\Omega \sin^2\theta \cos^2\phi. \tag{5.522}$$

For radiation from a pure linearly oscillating dipole oriented across the axis of interest, $\zeta^2 = 2/5$.

5.8.7 Velocity Dependence

When we take into account the atom's motion, the Rabi frequency Ω becomes time-dependent. Thus, for example,

$$\partial_t \Omega = \mathbf{v}\cdot\nabla\Omega = \Omega\mathbf{v}\cdot\left(\frac{\nabla|\Omega|}{|\Omega|} + i\nabla\phi\right) = \Omega\mathbf{v}\cdot\left(\frac{\nabla s}{2s} + i\nabla\phi\right). \tag{5.523}$$

For the saturation parameter, we can thus also write

$$\partial_t s = \frac{2|\Omega|\partial_t|\Omega|}{2[(\Gamma/2)^2+\Delta^2]} = 2s\frac{\partial_t|\Omega|}{|\Omega|}, \tag{5.524}$$

so that

$$\frac{\partial_t s}{2s} = \frac{\partial_t|\Omega|}{|\Omega|} = \mathbf{v}\cdot\frac{\nabla|\Omega|}{|\Omega|} = \mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right). \tag{5.525}$$

Now we can obtain the atomic density matrix to lowest order in $\mathbf{v}$ as follows. First, differentiate the at-rest steady-state solutions

$$\begin{aligned}\langle\sigma_z(t\longrightarrow\infty)\rangle &= -\frac{1}{1+s}\\ \tilde{\rho}_{\rm eg}(t\longrightarrow\infty) &= -\frac{i\Omega}{2(\Gamma/2-i\Delta)(1+s)}\end{aligned} \tag{5.526}$$

and keep terms to first order in $\mathbf{v}$ to obtain

$$\partial_t\langle\sigma_z\rangle \approx \frac{2s}{(1+s)^2}\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right) \approx -\frac{2s}{1+s}\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right)\langle\sigma_z\rangle \tag{5.527}$$

and

$$\begin{aligned}\partial_t\tilde{\rho}_{\rm eg} &\approx \tilde{\rho}_{\rm eg}\left(\frac{\partial_t\Omega}{\Omega}-\frac{\partial_t s}{1+s}\right)\\ &= \left[\mathbf{v}\cdot\left(\frac{\nabla s}{2s}+i\nabla\phi\right)-\frac{2s}{1+s}\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right)\right]\tilde{\rho}_{\rm eg}\\ &= \left[\left(\frac{1-s}{1+s}\right)\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right)+i\mathbf{v}\cdot\nabla\phi\right]\tilde{\rho}_{\rm eg}.\end{aligned} \tag{5.528}$$

The optical Bloch equations are

$$\begin{aligned}\partial_t\langle\sigma_z\rangle &= i\Omega^*\tilde{\rho}_{\rm eg}-i\Omega\tilde{\rho}_{\rm ge}-\Gamma\langle\sigma_z\rangle-\Gamma\\ \partial_t\tilde{\rho}_{\rm eg} &= \left(-\frac{\Gamma}{2}+i\Delta\right)\tilde{\rho}_{\rm eg}+\frac{i\Omega}{2}\langle\sigma_z\rangle,\end{aligned} \tag{5.529}$$

and equating these relations with the above expressions for the velocity-dependent time derivatives gives

$$\begin{aligned}i\Omega^*\tilde{\rho}_{\rm eg}-i\Omega\tilde{\rho}_{\rm ge}-\Gamma_v\langle\sigma_z\rangle &= \Gamma\\ \gamma_v\tilde{\rho}_{\rm eg}-\frac{i\Omega}{2}\langle\sigma_z\rangle &= 0,\end{aligned} \tag{5.530}$$

where

$$\begin{aligned}\Gamma_v &:= \Gamma-\frac{2s}{1+s}\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right)\\ \gamma_v &:= \frac{\Gamma}{2}-i\Delta+\left(\frac{1-s}{1+s}\right)\mathbf{v}\cdot\left(\frac{\nabla s}{2s}\right)+i\mathbf{v}\cdot\nabla\phi.\end{aligned} \tag{5.531}$$

Eliminating the coherences in the first of Eqs. (5.530) gives the solution

$$\langle\sigma_z\rangle = -\frac{\Gamma}{\Gamma_v+|\Omega|^2\dfrac{{\rm Re}[\gamma_v]}{|\gamma_v|^2}}, \tag{5.532}$$

so that the second of Eqs. (5.530) gives the steady-state coherence

$$\tilde{\rho}_{\rm eg} = \frac{i\Omega}{2\gamma_v}\langle\sigma_z\rangle = -\frac{i\Omega\Gamma/2\gamma_v}{\Gamma_v+|\Omega|^2\dfrac{{\rm Re}[\gamma_v]}{|\gamma_v|^2}} = -\frac{i\Omega\Gamma\gamma_v^*/2}{\Gamma_v|\gamma_v|^2+|\Omega|^2{\rm Re}[\gamma_v]}. \tag{5.533}$$

(velocity-dependent coherence)

The mean force from Eq. (5.441) as usual is then

$$\langle\mathbf{F}\rangle = -\frac{\hbar}{2}\left[\nabla\Omega^*(\mathbf{r})\,\tilde{\rho}_{\rm eg}+\nabla\Omega(\mathbf{r})\,\tilde{\rho}_{\rm ge}\right] = -\frac{\hbar\Omega^*}{2}\left(\nabla\log|\Omega|-i\nabla\phi\right)\tilde{\rho}_{\rm eg}+{\rm c.c.} \tag{5.534}$$

This is again somewhat complicated to interpret in general, so we will work out this expression in a couple of special cases.

For the plane wave, we again have $\nabla s = \nabla|\Omega| = 0$ and $\nabla\phi = \mathbf{k}$. Then $\Gamma_v = \Gamma$ and $\gamma_v = \Gamma/2 - i(\Delta - \mathbf{k}\cdot\mathbf{v})$. Note that this is exactly the same as the static analysis for this case, except for the Doppler-shift replacement $\Delta \longrightarrow \Delta - \mathbf{k}\cdot\mathbf{v}$. Thus,

$$\tilde{\rho}_{\rm eg} = -\frac{i\Omega}{2[\Gamma/2 - i(\Delta - \mathbf{k}\cdot\mathbf{v})](1+s)}, \tag{5.535}$$

and following the steps leading up to Eq. (5.450), the force becomes

$$\langle \mathbf{F}_{\rm rad}\rangle = \frac{\hbar\mathbf{k}\Gamma}{2}\frac{s(\mathbf{v})}{1+s(\mathbf{v})},$$

(velocity-dependent force, plane wave) (5.536)

where

$$s(\mathbf{v}) = \frac{|\Omega|^2}{2[(\Gamma/2)^2 + (\Delta - \mathbf{k}\cdot\mathbf{v})^2]}. \tag{5.537}$$

This force is only valid to first order in $\mathbf{v}$, and so

$$s(\mathbf{v}) = \left(1 + \frac{2\Delta\mathbf{k}\cdot\mathbf{v}}{(\Gamma/2)^2 + \Delta^2}\right)s,$$

(velocity-dependent saturation parameter) (5.538)

where s denotes $s(\mathbf{v} = 0)$, and

$$\frac{1}{1+s(\mathbf{v})} = \left(1 - \frac{s2\Delta\mathbf{k}\cdot\mathbf{v}}{(1+s)[(\Gamma/2)^2 + \Delta^2]}\right)\frac{1}{1+s}, \tag{5.539}$$

so that

$$\begin{aligned}\langle \mathbf{F}_{\rm rad}\rangle &= \frac{\hbar\mathbf{k}\Gamma}{2}\frac{s}{1+s}\left(1 + \frac{2\Delta\mathbf{k}\cdot\mathbf{v}}{(1+s)[(\Gamma/2)^2 + \Delta^2]}\right)\\ &= \hbar\mathbf{k}\Gamma\rho_{\rm ee}(\mathbf{v} = 0, t \longrightarrow \infty)\left(1 + \frac{2\Delta\mathbf{k}\cdot\mathbf{v}}{(1+s)[(\Gamma/2)^2 + \Delta^2]}\right).\end{aligned}$$

(velocity-dependent force, plane wave) (5.540)

The velocity-dependent part of this expression is the Doppler force, which, as we saw from the classical analysis of Section (1.4.2), is a damping, friction-like force for red detunings ($\Delta < 0$), which is what gives rise to optical molasses when multiple beams are present. (Note that from this expression, one beam is sufficient to cool an atom, provided that it is trapped by some other force that cancels the mean radiation pressure.)

For a standing wave in the x-direction, we have $\Omega = \Omega_0 \cos kx$ as usual, so that $\nabla s/2s = -\mathbf{k}\tan kx$ and $\nabla\phi = 0$. Then

$$\begin{aligned}\Gamma_v &= \Gamma + \frac{2s(\mathbf{r})}{1+s(\mathbf{r})}(\mathbf{v}\cdot\mathbf{k})\tan kx\\ \gamma_v &= \frac{\Gamma}{2} - i\Delta - \left(\frac{1-s(\mathbf{r})}{1+s(\mathbf{r})}\right)(\mathbf{v}\cdot\mathbf{k})\tan kx.\end{aligned} \tag{5.541}$$

The algebra here is more complicated, but the velocity-dependent part of the mean force reduces to

$$\langle \mathbf{F}_{\mathbf{v}}\rangle = (\mathbf{k}\cdot\mathbf{v})2\hbar\mathbf{k}\Delta\sin^2 kx\frac{\Gamma^2\Omega_0^2[2(\Delta^2 + \Gamma^2/4) - \Omega_0^2\cos^2 kx] - \Omega_0^6\cos^4 kx}{\Gamma[2(\Delta^2 + \Gamma^2/4) + \Omega_0^2\cos^2 kx]^3}.$$

(velocity-dependent force, standing wave) (5.542)

For small intensities, this expression becomes

$$\langle \mathbf{F}_{\mathbf{v}}\rangle = \hbar\mathbf{k}\Gamma\frac{\Omega_0^2}{2(\Delta^2 + \Gamma^2/4)^2}\Delta(\mathbf{k}\cdot\mathbf{v})\sin^2 kx = \hbar\mathbf{k}\Gamma\bar{\rho}_{\rm ee}(\mathbf{v} = 0, t \longrightarrow \infty)\frac{2\Delta(\mathbf{k}\cdot\mathbf{v})}{(\Gamma/2)^2 + \Delta^2}\sin^2 kx,$$

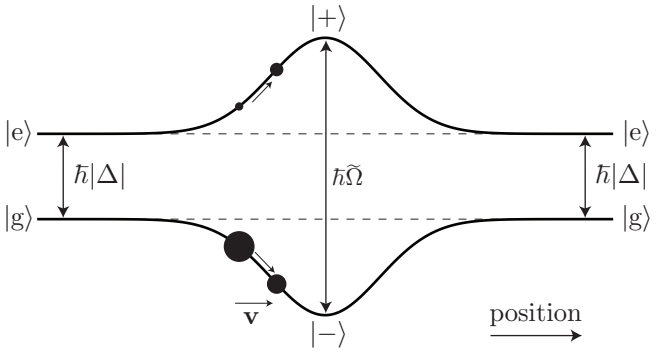
(velocity-dependent force, standing wave, small intensity) (5.543)

where

$$\bar{\rho}_{\rm ee}(\mathbf{v}=0, t\longrightarrow\infty)=\frac{s_0}{2}=\frac{\Omega_0^2}{4[(\Gamma/2)^2+\Delta^2]} \tag{5.544}$$

is the zero-velocity equilibrium population of the atomic excited state, where the intensity is taken to be spatially averaged over the standing wave. Comparison of this expression to Eq. (5.540) makes it clear that, for small intensities, the velocity-dependent force in a standing wave is explained by optical molasses, but where the strength is modulated sinusoidally with period $\lambda/2$.

The more interesting feature here is that the sign of the force changes as the intensity becomes very large. Thus, the velocity-dependent force is damping for red detuning and small intensities, but becomes a heating force for large detunings. The interpretation in the large-intensity regime is that the local steady-state, dressed-level populations lag behind the atomic position.[43]



To visualize this, let's consider an atom moving in a Gaussian beam, with $\Delta<0$. Thus, the atom is primarily in the $|-\rangle$ state, which is more like the ground state than the excited state. Suppose that it is moving to the right. In an instant, when it moves a small distance to the right, the *equilibrium* population for the atom is *less* in $|-\rangle$ and *more* in $|+\rangle$. (The populations are equal in the limit of large intensity.) However, the atom doesn't adjust its populations instantaneously, so when it arrives at the new position, it has ''too much'' population in $|-\rangle$ compared to equilibrium. Thus, the dressed-state shifts don't cancel as much as they would otherwise, and so the force is larger than it would otherwise be. The velocity-dependent ''correction'' is thus in the same direction as the velocity, as we expect for $\Delta<0$. This argument works, of course, when the energy gradient has the opposite sign or when the atom has the opposite velocity.

5.8.8 Doppler Cooling Limit

Now that we have the velocity-dependent force and the diffusion coefficient, we can treat the problem of laser cooling from the quantum-mechanical viewpoint. Recall from Eq. (5.467), the steady-state solution of the linear Fokker–Planck equation with drift coefficient $A(p)=Ap$ and diffusion coefficient $D(p)=D$ is a Gaussian (''thermal'') distribution in momentum of variance

$$V_p(t\longrightarrow\infty)=\frac{D}{2|A|}. \tag{5.545}$$

We will consider the case of two traveling waves, which form a standing wave. The diffusion rate is given by the sum of the diffusion rate from the fluctuation force in Eq. (5.510) and the spontaneous-emission diffusion rate from Eq. (5.518). To avoid a cumbersome expression (and to focus on the cooling regime), we will consider the small-intensity limit, where the diffusion coefficients are given by Eqs. (5.511) and (5.520). The sum is independent of position, and is given by

$$D_p=\frac{\hbar^2k^2\Gamma s_0}{2}, \tag{5.546}$$

[43] J. Dalibard and C. Cohen–Tannoudji, *op. cit.*

where again $s_0 = \Omega_0^2/2[(\Gamma/2)^2 + \Delta^2]$. The linear drift coefficient is given from Eq. (5.543) by

$$A = \frac{\langle \mathbf{F_v} \rangle}{m} = \frac{\hbar \Gamma k^2 s_0}{2m} \frac{\Delta}{(\Gamma/2)^2 + \Delta^2}, \tag{5.547}$$

where we have replaced the $\sin^2 kx$ dependence by the spatially averaged value of 1/2, since we assume the atoms are not so cold that they are stationary on wavelength distance scales. Now remember that D_p represents diffusion in three dimensions, so to have three-dimensional cooling we must have three sets of standing waves. This means that the cooling force is three-dimensional, with the same coefficient A, but the diffusion rate is larger by a factor of three. Thus, the steady-state variance for negative detunings is

$$V_p(t \longrightarrow \infty) = \frac{3D_p}{2|A|} = \frac{3\hbar m}{2} \frac{(\Gamma/2)^2 + \Delta^2}{|\Delta|}. \tag{5.548}$$

We can translate this into a kinetic energy via

$$E_p(t \longrightarrow \infty) = \frac{V_p(t \longrightarrow \infty)}{2m} = \frac{3\hbar}{4} \frac{(\Gamma/2)^2 + \Delta^2}{|\Delta|}. \tag{5.549}$$

We can convert the energy to a temperature by $E_p(t \longrightarrow \infty) = (3/2)k_{\text{B}}T$, giving

$$k_{\text{B}}T = \frac{\hbar\Gamma}{4} \frac{1 + (2\Delta/\Gamma)^2}{2|\Delta|/\Gamma}. \tag{5.550}$$

(Doppler limit to laser cooling)

This is precisely the same result that we obtained from our Lorentz-model treatment of Section 1.4.3.1. Again, all the ingredients are the same: we have included the cooling force (optical molasses) via the velocity-dependent force of Eq. (5.543), the diffusion due to absorption from a random beam in the diffusion rate of Eq. (5.510), and the diffusion rate due to spontaneous emission from Eq. (5.518). In principle, though the more general expressions can also treat more general situations than the weak-field case, and we have now shown more explicitly that we expect (to lowest order in velocity) a thermal steady-state distribution.

5.9 Bloch–Siegert Shift

All of the results that we have developed for the two-level atom have involved the rotating-wave approximation. So, then, what *is* the effect of the neglected term? One well-known effect that is closely related to the dipole shift is the **Bloch–Siegert shift** of the atomic resonance.[44]

Recall that, within the rotating-wave approximation, a monochromatic field of frequency ω induces an ac Stark shift of the ground state, which we can see from Eq. (5.460) is given by

$$\Delta E_{\text{g}} = \frac{\hbar\Omega^2}{4\Delta}, \tag{5.551}$$

to lowest order in Ω^2/Δ^2, where as usual $\Delta = \omega - \omega_0$. The excited-state shift is exactly opposite the ground-state shift:

$$\Delta E_{\text{e}} = -\frac{\hbar\Omega^2}{4\Delta}. \tag{5.552}$$

We would now like to treat the counterrotating term as another monochromatic field, but of frequency $-\omega$. To lowest order, the shift is just an additive shift of the same form, but with $\omega \longrightarrow -\omega$. This treatment is corroborated by the classical treatment of the dipole potential, specifically Eq. (1.76), where the counterrotating field gave an additional dipole shift of the same form, but with the replacement

$$\frac{1}{\Delta} = \frac{1}{\omega - \omega_0} \longrightarrow -\frac{1}{\omega + \omega_0}. \tag{5.553}$$

[44] After F. Bloch and A. Siegert, "Magnetic Resonance for Nonrotating Fields," *Physical Review* **57**, 522 (1940) (doi: 10.1103/PhysRev.57.522).

Thus, we expect a ground-state shift due to the counterrotating term of

$$\Delta E_{\mathrm{g,c}} = -\frac{\hbar\Omega^2}{4(\omega+\omega_0)}, \tag{5.554}$$

while the excited state experiences an equal but opposite shift. The shift of the transition frequency due to the counterrotating field is thus

$$\Delta\omega_{\mathrm{c}} = \frac{\Delta E_{\mathrm{e,c}} - \Delta E_{\mathrm{g,c}}}{\hbar} = \frac{\Omega^2}{2(\omega+\omega_0)}. \tag{5.555}$$

The shift in the atomic resonance is given to lowest order by setting $\omega = \omega_0$ in this expression, so that

$$\Delta\omega_{\mathrm{c}} = \frac{\Omega^2}{4\omega_0}. \tag{5.556}$$
(Bloch–Siegert shift)

This is the lowest-order expression for the Bloch–Siegert shift.[45] The shift is typically quite weak: a relatively large Rabi frequency of $\Omega/2\pi = 1$ GHz is large enough to drive a 2π-pulse in only 1 ns, but at optical frequencies (say for 1 μm light), the Bloch–Siegert shift is only 1 kHz, or a fractional shift of about 10^{-12}. This is well within the power-broadened line shape, and quite difficult to detect.[46] Thus, the first-order result is quite adequate for most optical situations.

An important point regarding the Bloch–Siegert shift is that it is due to a nonresonant interaction. Nonresonance is the justification for treating an atom as a two-state system: the other levels are not resonantly coupled and are therefore ignorable. However, if the detuning is large enough or the desired accuracy of the calculation is high enough that the effects of the counterrotating term are important, then so are the effects of couplings to other levels. That is, *the two-level and rotating-wave approximations are at the same level of accuracy.* For consistency of approximations either both or neither should be made. (Of course, in magnetic resonance it is possible to have an exact two-state system where the Bloch–Siegert shift is significant.)

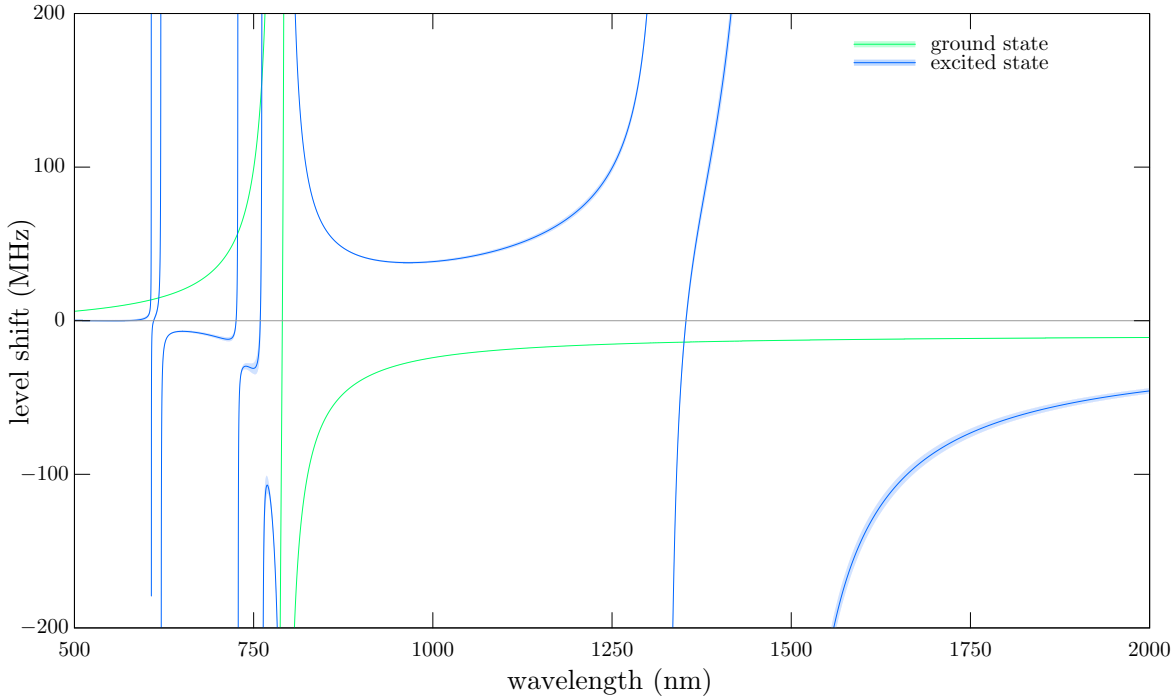
5.9.1 Magic Wavelength

In the two-level atom, the ac Stark shift always results in opposite shifts between the excited and ground states, and thus always leads to a shift of the transition frequency. Of course, the fact that couplings to other states exist can work to our advantage. In particular, when accounting for the counterrotating interactions and couplings to other levels, it may be possible to find situations where the excited- and ground-state shifts are equal, with the *same* sign. This situation happens when the atom is driven at particular wavelengths, called **magic wavelengths**.

To illustrate this, consider the level shifts of ^{87}Rb of the two levels in the D_1 transition at 794 nm due to a monochromatic laser field.

[45] It is possible to obtain higher-order expressions for the Bloch–Siegert shift. See Jon H. Shirley, "Solution of the Schrödinger Equation with a Hamiltonian Periodic in Time," *Physical Review*, **138**, B979 (1965) (doi: 10.1103/PhysRev.138.B979); Stig Stenholm, "Quantum theory of RF resonances: The semiclassical limit," *Journal of Physics B: Atomic and Molecular Physics*, **6**, 1650 (1973) (doi: 10.1088/0022-3700/6/8/042); C. Cohen-Tannoudji, J. Dupont-Roc and C. Fabre, "A quantum calculation of the higher order terms in the Block Siegert shift," *Journal of Physics B: Atomic and Molecular Physics*, **6**, L214 (1973) (doi: 10.1088/0022-3700/6/8/007); P. Hannaford, D. T. Peg, and G. W. Series, "Analytical expressions for the Bloch-Siegert shift," *Journal of Physics B: Atomic and Molecular Physics*, **6**, L222 (1973) (doi: 10.1088/0022-3700/6/8/009).

[46] Note, however, that the shift, including higher order corrections, have been studied experimentally for microwave transitions. See C. Cohen-Tannoudji, J. Dupont-Roc and C. Fabre, "An experimental check of higher order terms in the radiative shift of a coherence resonance," *Journal of Physics B: Atomic and Molecular Physics*, **6**, L218 (1973) (doi: 10.1088/0022-3700/6/8/008).



The plotted energy shifts correspond to those experienced by an atom at the center of a 1 W Gaussian beam of radius $w_0 = 10$ μm. We will defer the details of the calculation until we develop more sophisticated technology for dealing with atomic structure. For now, we will note that we have done the calculation accounting for 24 total atomic transitions, some of which are visible as dispersive resonances in the level-shift curves. Magic wavelengths occur at the locations where the two curves intersect. For example, a laser spot at 1350.39(65) nm can be used as a dipole trap for ^{87}Rb, where the frequency of the D_1 transition is not affected.

Why is this interesting? There are two well-known application. The first is, suppose you have an atom in a dipole trap, and you want to probe a transition with a resonant laser. Well, recall that the stochastic dipole force is due precisely to having different dipole potentials in different states, and then stochastically jumping between them due to spontaneous emission. Of course, the same thing will happen in this situation: when probed, the atom will jump between the ground and excited states, where it will experience different trapping potentials. This can be a large source of heating and decoherence. But trapping at the magic wavelength suppresses this form of heating. This approach is used to great advantage in cavity quantum electrodynamics to obtain long trap lifetimes of a single atom in an ultrahigh-finesse cavity.[47]

The other application is to atomic frequency standards. The current frequency standard is based on a microwave hyperfine transition in ^{133}Cs. But increasing the quality factor $\omega_0/\delta\omega$ of the oscillator standard requires both decreasing the line width $\delta\omega$ as well as pushing the resonance frequency ω_0 into the optical. To allow for long interrogation times, the atoms must be trapped. One promising approach is the single-ion atomic clock, where the oscillating trap fields average to zero to lowest order. Another promising approach is to trap atoms in a magic-wavelength, three-dimensional optical lattice.[48] This allows the atoms to be trapped with little shift of the optical transition. Also, the atoms do not suffer collisions because they are isolated at different lattice sites. So the clock does not suffer from a collisional shift, but retains the advantage of averaging over many atoms.

[47] J. McKeever, J. R. Buck, A. D. Boozer, A. Kuzmich, H.-C. Nägerl, D. M. Stamper-Kurn, and H. J. Kimble, "State-Insensitive Cooling and Trapping of Single Atoms in an Optical Cavity," *Physical Review Letters* **90**, 133602 (2003) (doi: 10.1103/PhysRevLett.90.133602).

[48] Masao Takamoto and Hidetoshi Katori, "Spectroscopy of the $^1S_0$-$^3P_0$ Clock Transition of ^{87}Sr in an Optical Lattice," *Physical Review Letters* **91**, 223001 (2003) (doi: 10.1103/PhysRevLett.91.223001); Carsten Degenhardt, Hardo Stoehr, Uwe Sterr, Fritz Riehle, and Christian Lisdat, "Wavelength-dependent ac Stark shift of the $^1S_0$-$^3P_1$ transition at 657 nm in Ca," *Physical Review A* **70**, 023414 (2004) (doi: 10.1103/PhysRevA.70.023414); Masao Takamoto, Feng-Lei Hong, Ryoichi Higashi, and Hidetoshi Katori, "An optical lattice clock," *Nature* **435**, 321 (2005) (doi: 10.1038/nature03541).

5.10 Exercises

Problem 5.1

In the classical treatment of the dipole potential, we used the expression $-(1/2)\mathbf{d}\cdot\mathbf{E}$ for the dipole energy, where the $1/2$ accounts for the fact that the dipole is induced. However, for the two-level atom, we used the interaction Hamiltonian $H_{\text{AF}} = -\mathbf{d}\cdot\mathbf{E}$, which has no such factor. What gives?

Hint: the interaction Hamiltonian is also valid *classically*, so argue that the above two expressions are consistent in the classical model. In particular, show that the Lorentz model assumed an interaction of the form of H_{AF}. Then, modeling the atom as a classical harmonic oscillator, show that a perturbation of the form H_{AF} leads to an energy shift $-(1/2)\mathbf{d}\cdot\mathbf{E}$ (it suffices to assume a *static* perturbation).

Problem 5.2

(a) Derive the equations of motion for the amplitudes of the quantum state

$$|\psi\rangle = c_{\text{g}}|\text{g}\rangle + c_{\text{e}}|\text{e}\rangle \tag{5.557}$$

under the evolution of the Hamiltonian $H_{\text{A}} + H_{\text{AF}}$.

(b) Then make the transformation to the rotating frame by defining

$$\tilde{c}_{\text{e}} := c_{\text{e}} e^{i\omega t}, \tag{5.558}$$

and rewrite the equations of motion in terms of $\tilde{c}_{\text{e}}$.

(c) Finally, define the rotating-frame quantum state

$$|\tilde{\psi}\rangle = c_{\text{g}}|\text{g}\rangle + \tilde{c}_{\text{e}}|\text{e}\rangle, \tag{5.559}$$

and show that the equations of motion for this state under the rotating-frame Hamiltonian $\tilde{H}_{\text{A}} + \tilde{H}_{\text{AF}}$ are equivalent to your results for part (b).

Problem 5.3

Consider a quantum-mechanical particle in a double-well potential, not necessarily symmetric. For simplicity, we will make a two-state approximation for the particle, restricting our Hilbert space to the lowest two energy levels.

(a) In the uncoupled (i.e., no tunneling through the barrier) limit, we can write the state in the left-hand well as $|\text{L}\rangle$, and the state in the right-hand well is $|\text{R}\rangle$. Using this basis, write down the most general Hamiltonian that describes this system, *including* the tunneling interaction. Introduce new parameters as necessary but explain what they represent and any necessary constraints.

(b) Take the double-well potential to be symmetric. Assume that $|\psi(t=0)\rangle = |\text{L}\rangle$, and let $|\psi(t)\rangle = c_{\text{L}}(t)|\text{L}\rangle + c_{\text{R}}(t)|\text{R}\rangle$. Show that the state will oscillate periodically in time between the two wells. Rewrite the Hamiltonian from part (a) for this case in terms of the period T of oscillation.

(c) For the case in part (b), what are the eigenstates and eigenenergies of the Hamiltonian? How does the structure of the eigenstates and eigenenergies explain the tunneling process? (You need not derive them if you know them, but give precise answers and *explain*.)

Problem 5.4

Show that the eigenstates (dressed states) of the Hamiltonian $\tilde{H}_{\text{A}} + \tilde{H}_{\text{AF}}$ can be written

$$\begin{aligned} |+\rangle &= \sin\theta|\text{g}\rangle + \cos\theta|\text{e}\rangle \\ |-\rangle &= \cos\theta|\text{g}\rangle - \sin\theta|\text{e}\rangle, \end{aligned} \tag{5.560}$$

where

$$\tan 2\theta = -\frac{\Omega}{\Delta} \quad \left(0 \le \theta < \frac{\pi}{2}\right). \tag{5.561}$$

Problem 5.5

Go through the derivation of the Landau–Zener transition probability in Section 5.3.2 *carefully*, filling in the missing steps, to derive the loss probability

$$P_{\text{lost}} = e^{-2\pi|\nu|} = \exp\left(-\frac{\pi\Omega^2}{2|\partial_t\Delta|}\right). \tag{5.562}$$

Problem 5.6

(a) Show that the transition probability for a Ramsey-type interference experiment can be written

$$P_{\text{e}} = 4\frac{\Omega^2}{\tilde{\Omega}^2}\sin^2\left(\frac{\tilde{\Omega}\tau}{2}\right)\left[\cos\left(\frac{\Delta T}{2}\right)\cos\left(\frac{\tilde{\Omega}\tau}{2}\right) - \frac{\Delta}{\tilde{\Omega}}\sin\left(\frac{\Delta T}{2}\right)\sin\left(\frac{\tilde{\Omega}\tau}{2}\right)\right]^2, \tag{5.563}$$

where τ is the interaction of each of the two laser pulses of Rabi frequency Ω and detuning Δ, and T is the "drift" time between pulses.

(b) Make a plot of the excitation vs. the microwave detuning. Use parameters appropriate for the NIST-7 cesium beam clock.

(c) Make another excitation plot, but this time average the fringes over the velocity distribution of the cesium beam. Assume a temperature of 100°C, corresponding to a velocity width of 240 m/s. (I suggest doing the averaging *numerically*, not analytically.)

Problem 5.7

Solve the optical Bloch equations to obtain the population inversion $\langle\sigma_z(t)\rangle$ in the limit of weak excitation, $\Omega \ll \Gamma$, in the homogeneously broadened case ($\gamma_\perp = \Gamma/2$) and for arbitrary detuning Δ. Assume that the atom is initially in the ground state.

Problem 5.8

Find the solution to the optical Bloch equations for $\langle\sigma_z(t)\rangle$ in the strong-drive limit $\Omega \gg \gamma_\perp, \Gamma$, for arbitrary detuning Δ. Keep only lowest-order contributions in Ω^{-2} in your solution. Do *not* assume $\Omega \gg |\Delta|$.

Problem 5.9

(a) Summarize the effects of the nonlinear response of a two-level atom to an externally applied, classical, monochromatic field. (Treat all aspects we covered, but ignore atomic motion.)

(b) In Problem 1.5, you found that adding nonlinearities to the Lorentz model resulted in new frequencies being generated from the original, monochromatic driving field. Explain qualitatively why the nonlinear response of the two-level atom does not similarly generate harmonics of the original driving field. How would the two-level atom need to be generalized to model harmonic generation (assuming that we keep the dipole approximation)?

Problem 5.10

Use the solutions of the optical Bloch equations to derive an expression for the frequency-dependent polarizability $\alpha(\omega)$ for the two-level atom. Show that your results are consistent with the result from the Lorentz atom in the appropriate regime. Discuss the significance of any differences.

Problem 5.11

If the density operator of the two-level atom evolves according to

$$\partial_t\rho = -\frac{i}{\hbar}[H,\rho] + \Gamma\mathcal{D}[\sigma]\rho, \tag{5.564}$$

derive an equation of motion for the purity in terms of the elements ρ_{ee}, ρ_{eg}, ρ_{ge}, and ρ_{gg} of the density matrix.

Problem 5.12

Go through the derivation of the Mollow triplet on resonance, including the details of the quantum regression theorem, filling in the missing steps.

Problem 5.13

Use the quantum regression theorem to derive the inelastic spectrum for a two level atom driven far off resonance, $|\Delta| \gg \Omega, \Gamma$. What are the line widths and (integrated) weights?

Problem 5.14

Derive expressions for the elastic and inelastic emission spectra for a two-level atom with collisional broadening, driven weakly ($\Omega/\Gamma \ll 1$) in the presence of collisional broadening, and show that the inelastic component is a single peak at the atomic resonance frequency.

Problem 5.15

Go through the derivation of the second-order coherence function $g^{(2)}(\tau)$ on resonance, filling in the missing steps.

Problem 5.16

Derive the expression for the Autler–Townes absorption correlation function $g_{\rm a}(\tau)$ in the case of homogeneous broadening but arbitrary pump intensity and detuning, Eq. (5.373) in the notes. To do this, work out the details of the quantum regression theorem for this system, and then solve the resulting system of equations.

Hint: the quantum regression theorem gives the correlation function in terms of a solution of 9 coupled differential equations. However, only two of them are nonzero. Start off by eliminating the unnecessary components.

Problem 5.17

Give an alternate derivation of the general dipole potential

$$V_{\rm dip} = \frac{\hbar\Delta}{2}\log[1+s(\mathbf{r})] \tag{5.565}$$

in the dressed-state picture, and thus interpret the potential as a combination of the Stark shifts of the two dressed levels, plus a contribution from the field due to scattering into the Mollow sidebands. Use the following outline for your derivation.[49]

(a) Show that when the Rabi frequency Ω is complex, the dressed-state energies are

$$E_\pm = -\frac{\hbar\Delta}{2} \pm \frac{\hbar\tilde{\Omega}}{2}, \tag{5.566}$$

where the generalized Rabi frequency is $\tilde{\Omega} = \sqrt{|\Omega|^2+\Delta^2}$, and the dressed states are given by

$$\begin{aligned} |+\rangle &= e^{-i\phi/2}\sin\theta|{\rm g}\rangle + e^{i\phi/2}\cos\theta|{\rm e}\rangle \\ |-\rangle &= e^{-i\phi/2}\cos\theta|{\rm g}\rangle - e^{i\phi/2}\sin\theta|{\rm e}\rangle, \end{aligned} \tag{5.567}$$

where

$$\tan 2\theta = -\frac{|\Omega|}{\Delta} \quad \left(0 \le \theta < \frac{\pi}{2}\right), \tag{5.568}$$

and $\Omega = |\Omega|e^{i\phi}$.

[49] J. Dalibard and C. Cohen-Tannoudji, "Dressed-atom approach to atomic motion in laser light: the dipole force revisited," *Journal of the Optical Society of America B* **2**, 1707 (1985).

Hint: argue that the Hamiltonian has the form

$$\tilde{H} = \hbar \begin{bmatrix} -\Delta & \Omega/2 \\ \Omega^*/2 & 0 \end{bmatrix}, \tag{5.569}$$

and can be diagonalized by a unitary matrix. What is the most general form of a unitary matrix?

(b) Derive the steady-state populations $\rho_{\pm\pm}$ and coherences $\rho_{\pm\mp}$ for the dressed levels in the limit of either large pumping ($|\Omega| \gg \Gamma$) or detuning ($|\Delta| \gg \Gamma$), which is the limit in which the dressed states are well resolved and thus give a valid description of the system.

Hint: first find the density-matrix elements in the bare basis, and make appropriate approximations in this limit. Then use the appropriate matrix transformation to switch to the dressed basis. Your answers should be

$$\begin{aligned} \rho_{++}(t \longrightarrow \infty) &= \frac{1}{2}\left[1 + \frac{\tilde{\Omega}}{\Delta(1+s)}\right] \\ \rho_{+-}(t \longrightarrow \infty) &= 0, \end{aligned} \tag{5.570}$$

where $s = |\Omega|^2/2\Delta^2$.

(c) Starting with the expression from Eq. (5.441),

$$\begin{aligned} \langle \mathbf{F} \rangle &= -\frac{\hbar}{2}\left[\nabla\Omega^*(\mathbf{r})\,\langle\sigma\rangle + \text{c.c.}\right] \\ &= -\frac{\hbar\,\langle\sigma\rangle}{2}\left[(\nabla|\Omega|)e^{-i\phi} - i(\nabla\phi)|\Omega|e^{i\phi}\right] + \text{c.c.}, \end{aligned} \tag{5.571}$$

we drop the radiation-pressure contribution (the term with $\nabla\phi$) to obtain the dipole force

$$\langle \mathbf{F} \rangle_{\text{dip}} = -\frac{\hbar\,\langle\sigma\rangle}{2} e^{-i\phi}\nabla|\Omega| + \text{c.c.} \tag{5.572}$$

Show that the force can be written in terms of the dressed states as

$$\langle \mathbf{F} \rangle_{\text{dip}} = -\frac{\hbar}{2}\nabla\tilde{\Omega}(\rho_{++} - \rho_{--}) - \hbar\tilde{\Omega}\nabla\theta(\rho_{+-} + \rho_{-+}). \tag{5.573}$$

(d) The work to move the atom a distance $d\mathbf{r}$ is

$$\begin{aligned} dW &= -\langle \mathbf{F} \rangle_{\text{dip}} \cdot d\mathbf{r} \\ &= \frac{\hbar}{2}\left(\nabla\tilde{\Omega}\cdot d\mathbf{r}\right)(\rho_{++} - \rho_{--}) + \hbar\tilde{\Omega}\,(\nabla\theta\cdot d\mathbf{r})\,(\rho_{+-} + \rho_{-+}). \end{aligned} \tag{5.574}$$

For a static or slowly moving atom, only the first term contributes to the dipole force; the second term involves the dressed-state coherences, and represents nonadiabatic transitions between the dressed states due to the displacement. Show explicitly based on your answer from (b) that the second term vanishes for adiabatic displacements.

(e) The displacement work can then be written as

$$\begin{aligned} dW &= \frac{\hbar d\tilde{\Omega}}{2}(\rho_{++} - \rho_{--}) \\ &= \rho_{++}dE_+ + \rho_{--}dE_-. \end{aligned} \tag{5.575}$$

Use your results of part (b) to evaluate this expression, and then integrate it with respect to position to obtain the dipole potential (valid in the limit of well-resolved dressed states).

(f) The position-dependent level shift of the dressed states, weighted by the dressed-state population, is

$$U_{\rm A} = \Delta E_+ \rho_{++} + \Delta E_- \rho_{--}, \tag{5.576}$$

where

$$\Delta E_\pm = \pm \frac{\hbar\tilde{\Omega}}{2} \tag{5.577}$$

are the dressed-state energy shifts relative to the mean level energy. Show that $U_{\rm A}$ is *not* the correct dipole potential.

(g) Show that in the limit of either large pumping ($|\Omega| \gg \Gamma$) or detuning ($|\Delta| \gg \Gamma$) the optical Bloch equations give

$$\begin{aligned} \partial_t \rho_{++} &= -\Gamma_{+-}\rho_{++} + \Gamma_{-+}\rho_{--} \\ \partial_t \rho_{--} &= \quad \Gamma_{+-}\rho_{++} - \Gamma_{-+}\rho_{--}, \end{aligned} \tag{5.578}$$

for the (position-dependent) dressed-state populations, where

$$\Gamma_{+-} = \Gamma \cos^4\theta, \qquad \Gamma_{-+} = \Gamma \sin^4\theta \tag{5.579}$$

represent the decay rates between the dressed states. To do this, make the adiabatic, or *secular* approximation, where the populations and coherences evolve on very different time scales and thus can be decoupled by replacing the coherences by their steady-state (mean) values in the population equations. (Why do the populations vary slowly?) Then $\Gamma_{+-}\rho_{++}$ is the rate for $|+\rangle \longrightarrow |-\rangle$ transitions, and thus the rate for producing photons in the $\omega + \tilde{\Omega}$ Mollow side band, and $\Gamma_{-+}\rho_{--}$ is the rate for $|-\rangle \longrightarrow |+\rangle$ transitions, and thus the rate for producing photons in the $\omega - \tilde{\Omega}$ Mollow side band. In equilibrium, the two rates are equal, implying a symmetric Mollow spectrum.

Note that the optical Bloch equations in the dressed-state basis reduce to simple rate equations in the secular approximation. Otherwise, the dressed-state equations are rather more complicated than for the bare states. Thus, the dressed basis is most useful when the secular approximation holds.

(h) Finally, show that

$$dW = dU_{\rm A} + dU_{\rm F}, \tag{5.580}$$

where

$$dU_{\rm F} = \hbar\tilde{\Omega}(\Gamma_{+-}\rho_{++} - \Gamma_{-+}\rho_{--})\, dt. \tag{5.581}$$

Interpret this relation as the energy transferred to the field in time dt due to the displacement of the atom. The dipole potential thus has contributions from both the atom and field energies.

Problem 5.18

Recall the one-dimensional Fokker–Planck equation:

$$\partial_t f(p,t) = -\partial_p A(p) f(p,t) + \frac{1}{2}\partial_p^2 D(p) f(p,t). \tag{5.582}$$

Since the drift coefficient $A(p)$ and diffusion coefficient $D(p)$ depend on momentum, we can't write down a general solution, and the interpretation of these coefficients is a bit tricky. However, computing the equations of motion for the first two moments helps with the interpretation.

(a) Show that

$$\partial_t \langle p \rangle = \langle A(p) \rangle . \tag{5.583}$$

Here, the expectation value refers to an average with respect to the distribution $f(p,t)$. Assume that $f(p,t)$ falls off quickly enough with $|p|$ that any boundary terms are negligible. Thus, "drift" of the distribution mean is caused both by the drift coefficient, averaged over the distribution.

(b) Show that

$$\partial_t V_p = 2\langle pA(p) \rangle - 2\langle p \rangle \langle A(p) \rangle + \langle D(p) \rangle , \tag{5.584}$$

under the same assumptions as in part (a), where $V_p := \langle p^2 \rangle - \langle p \rangle^2$. Thus, the diffusion coefficient causes spreading of the variance, again weighted by the distribution $f(p,t)$. Also, variations in the drift coefficient can cause the distribution to spread or contract, since different accelerations for parts of the distribution at different momenta can tear them apart or squish them together, depending on the relative accelerations.

Problem 5.19

Consider an atom trapped in a potential well of an optical standing wave (one-dimensional optical lattice).

(a) The standing wave is composed of two counterpropagating traveling waves. Suppose that the frequency of *one* of the traveling waves is shifted by a small amount $\delta\omega$. Show mathematically that the new configuration corresponds to a *moving* optical lattice. Intuitively, you can see this because there exists a moving reference frame where the Doppler-shifted frequencies of the two waves are equal, and thus the optical lattice is at rest in this frame. What is the velocity of the standing wave? What is the frequency shift that matches the velocity of an atom moving with momentum $\hbar k$, where k is the wave number of the optical lattice? Compute this last frequency shift for ^{87}Rb, assuming that the optical lattice is tuned close to the 780 nm resonance.

(b) If you add a linear *chirp* (frequency sweep) to the frequency of one of the traveling waves, then you end up with an optical lattice of constant *acceleration*. An atom bound to the optical lattice will accelerate with it, making transitions to higher momentum states. Consider the component of the atom in the $p = 0$ state. Use what you know about Bragg scattering and adiabatic passage to explain qualitatively how the atom in the accelerating optical lattice makes transitions to the states $p = 2\hbar k, 4\hbar k, \ldots$ Also explain how, from the quantum-optics viewpoint (i.e., using an adiabatic passage argument), how the atom is ''lost'' from the lattice and stops accelerating if the acceleration is too large.

(c) Given a particular potential depth V_0, estimate the critical acceleration a_c, above which the atom is lost from the lattice. Give this estimate using both a classical argument, assuming a classical potential of the form $V_0 \cos 2kx$, as well as a quantum-mechanical argument based on adiabatic passage.

Problem 5.20

Derive the expression (5.369)

$$S_a(\omega_p) = \frac{\left[\rho_{gg}(t \longrightarrow \infty) - \rho_{ee}(t \longrightarrow \infty)\right]\left[(\Gamma - i\Delta_p)[\gamma_\perp + i(\Delta - \Delta_p)] + \dfrac{i\Omega^2\Delta_p}{2(\gamma_\perp + i\Delta)}\right]}{(\Gamma - i\Delta_p)[\gamma_\perp + i(\Delta - \Delta_p)][\gamma_\perp - i(\Delta + \Delta_p)] + \Omega^2(\gamma_\perp - i\Delta_p)} + \text{c.c.} \quad (5.585)$$

for the Mollow probe-absorption spectrum, using the following outline.

This problem is a good prototype for how to treat an atom interacting with a *bichromatic field*. By treating one of the fields as a perturbation, the solution is greatly simplified (relatively speaking).

(a) Write down the interaction Hamiltonian for the probe field with the atom in terms of the probe Rabi frequency Ω_p and the probe frequency ω_p. Then write down the same Hamiltonian in the rotating frame of the pump laser field (of Rabi frequency Ω and frequency ω), in terms of the probe-pump detuning $\Delta_p := \omega_p - \omega$. You may assume both Ω and Ω_p to be real.

(b) Write out the equations of motion for the atomic density-matrix elements, obtained from the master equation

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_A + \tilde{H}_{AF} + \tilde{H}_P, \tilde{\rho}\right] + \Gamma\mathcal{D}[\sigma]\tilde{\rho}. \quad (5.586)$$

(c) Now we will treat the probe-field interaction as a perturbation to the steady-state solution of $\tilde{H}_A + \tilde{H}_{AF}$ by making the *ansatz*

$$\tilde{\rho}_{\alpha\beta} = \tilde{\rho}^{(0)}_{\alpha\beta} + \delta\tilde{\rho}^{(0)}_{\alpha\beta} + \delta\tilde{\rho}^{(+)}_{\alpha\beta}e^{-i\Delta_p t} + \delta\tilde{\rho}^{(-)}_{\alpha\beta}e^{i\Delta_p t} \quad (5.587)$$

for the long-time solution to the equations of motion. Here, the $\tilde{\rho}^{(0)}_{\alpha\beta}$ are the steady-state solutions under the evolution of $\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}$; the $\delta\tilde{\rho}^{(0)}_{\alpha\beta}$ are the dc corrections due to the perturbation; and the $\delta\tilde{\rho}^{(\pm)}_{\alpha\beta}$ are (complex) constants giving the amplitudes of the oscillating corrections to the long-time solutions. Note that

$$\begin{aligned} \delta\tilde{\rho}^{(\pm)}_{\alpha\beta} &= \left(\delta\tilde{\rho}^{(\mp)}_{\beta\alpha}\right)^* \\ \delta\rho^{(\alpha)}_{\rm ee} &= -\delta\rho^{(\alpha)}_{\rm gg}, \end{aligned} \tag{5.588}$$

so that only nine of the coefficients are independent. Justify (qualitatively) this form of the *ansatz* as sufficient to describe the corrected solution to lowest order in the probe intensity. Then substitute these equations into the above equations, and use the "steady-state" condition $\partial_t \rho_{\alpha\beta} = 0$ to derive a set of nine(!) coupled, linear equations for the perturbation coefficients (use the above constraints to eliminate the coefficients $\rho^{(\alpha)}_{\rm gg}$).

Check your answers:

$$\begin{aligned} -\frac{i\Omega}{2}\delta\tilde{\rho}^{(0)}_{\rm ge} + \frac{i\Omega}{2}\delta\tilde{\rho}^{(0)}_{\rm eg} - \Gamma\delta\rho^{(0)}_{\rm ee} - \frac{i\Omega_{\rm p}}{2}\delta\tilde{\rho}^{(-)}_{\rm ge} + \frac{i\Omega_{\rm p}}{2}\delta\tilde{\rho}^{(+)}_{\rm eg} &= 0 \\ -\frac{i\Omega}{2}\delta\tilde{\rho}^{(+)}_{\rm ge} - \frac{i\Omega_{\rm p}}{2}\tilde{\rho}^{(0)}_{\rm ge} + \frac{i\Omega}{2}\delta\tilde{\rho}^{(+)}_{\rm eg} - \Gamma\delta\rho^{(+)}_{\rm ee} + i\Delta_{\rm p}\delta\rho^{(+)}_{\rm ee} &= 0 \\ -(\gamma_\perp + i\Delta)\delta\tilde{\rho}^{(0)}_{\rm ge} - \frac{i\Omega}{2}\left(\delta\rho^{(0)}_{\rm ee} - \delta\rho^{(0)}_{\rm gg}\right) - \frac{i\Omega_{\rm p}}{2}\left(\delta\rho^{(+)}_{\rm ee} - \delta\rho^{(+)}_{\rm gg}\right) &= 0 \\ -[\gamma_\perp + i(\Delta - \Delta_{\rm p})]\delta\tilde{\rho}^{(+)}_{\rm ge} - \frac{i\Omega}{2}\left(\delta\rho^{(+)}_{\rm ee} - \delta\rho^{(+)}_{\rm gg}\right) &= 0 \\ -[\gamma_\perp + i(\Delta + \Delta_{\rm p})]\delta\tilde{\rho}^{(-)}_{\rm ge} - \frac{i\Omega}{2}\left(\delta\rho^{(-)}_{\rm ee} - \delta\rho^{(-)}_{\rm gg}\right) - \frac{i\Omega_{\rm p}}{2}\left(\rho^{(0)}_{\rm ee} - \rho^{(0)}_{\rm gg}\right) &= 0. \end{aligned} \tag{5.589}$$

(d) Solve the set of equations to obtain an expression for $\delta\tilde{\rho}^{(+)}_{\rm eg}$. A computer and a symbolic algebra program could save you both time and severe cramping in your writing hand.

(e) Recall the Bloch equation for the atom interacting with the pump field (without the probe)

$$\partial_t \rho_{\rm ee} = \frac{i\Omega}{2}\left(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}\right) - \Gamma\rho_{\rm ee}. \tag{5.590}$$

Identify the function of these terms and argue that in steady state, the rate of photon absorption is

$$R_{\rm abs} = \frac{i\Omega}{2}\left(\tilde{\rho}_{\rm eg} - \tilde{\rho}_{\rm ge}\right). \tag{5.591}$$

Then argue that the rate of photon absorption from the *probe* field is

$$R_{\rm abs,\,probe} = \frac{i\Omega_{\rm p}}{2}\left(\delta\tilde{\rho}^{(+)}_{\rm eg} - \delta\tilde{\rho}^{(-)}_{\rm ge}\right), \tag{5.592}$$

and finally show that $R_{\rm abs,\,probe}$ is equivalent to the above absorption spectrum up to an overall factor.

Problem 5.21

Prove the following form of the quantum regression theorem, for computing a correlation function of the form $\langle A(t+\tau)B(t)\rangle$. In this case, the correlation function may be written

$$\lim_{t\to\infty}\langle A(t+\tau)B(t)\rangle = {\rm Tr}_{\rm S}[A\Lambda(\tau)], \tag{5.593}$$

where $\Lambda(\tau)$ obeys the master equation with initial conditions

$$\Lambda(0) = B\rho(t\to\infty). \tag{5.594}$$

Problem 5.22

Prove the following form of the quantum regression theorem. Suppose that the one-time average of an operator A can be written in the form

$$\langle B(t)\rangle = \sum_j g_j(t)\,\langle B_j(0)\rangle\,, \tag{5.595}$$

where $g_j(t)$ are functions representing the solution in terms of initial conditions $\langle B_j(0)\rangle$ of some set of operators, then the two-time average $\langle A(t)B(t+\tau)C(t)\rangle_{t\to\infty}$ may be written

$$\langle A(t)B(t+\tau)C(t)\rangle_{t\to\infty} \equiv \langle A(0)B(\tau)C(0)\rangle = \sum_j g_j(\tau)\,\langle AB_jC\rangle_{t\to\infty}\,. \tag{5.596}$$

Problem 5.23

Consider the quantum damped harmonic oscillator, with master equation

$$\partial_t\rho = -\frac{i}{\hbar}[H_0 + H_{\rm int},\rho] + \gamma\mathcal{D}[a]\rho, \tag{5.597}$$

with free Hamiltonian

$$H_0 = \hbar\omega_0\left(a^\dagger a + \frac{1}{2}\right) \tag{5.598}$$

and driving Hamiltonian

$$H_{\rm int} = \hbar\mathscr{E}\left(ae^{i\omega t} + a^\dagger e^{-i\omega t}\right). \tag{5.599}$$

(a) Derive the equations of motion for the density matrix elements ρ_{nm} in the energy basis.

(b) Find a transformation to a rotating frame, where the equations of motion for the slow variables $\tilde\rho_{nm}$ have no explicit time dependence.

(c) Write out explicit expressions for the equations of motion for ρ_{11}, $\tilde\rho_{10}$, $\tilde\rho_{01}$, and ρ_{00}, under the assumption that $\tilde\rho_{nm} = 0$ if $n > 1$ or $m > 1$. Compare your results to the optical Bloch equations.

Problem 5.24

In this problem we will construct and analyze a simple model for a laser by using the optical Bloch equations and the results of Problem 1.6 to describe the atom–field interaction. For our purposes here, we will think of a laser as comprised of (1) an **optical cavity** (resonator), which traps light and confines it in some region of space, as in a pair of parallel mirrors (Fabry–Perot cavity); (2) a **gain medium**, consisting of a vapor of two-level atoms uniformly filling the cavity; and (3) some **pump source** that promotes the atoms to the excited state.

(a) Derive the **Maxwell–Bloch equations** for the interaction of a single cavity-field mode with an ensemble of quantum-mechanical two-level atoms of number density N that fill the cavity:

$$\begin{aligned}
\partial_t E_0 &= \frac{N\omega_0 d_{\rm ge}}{\epsilon_0}\langle\sigma_y\rangle - \frac{\kappa}{2}E_0 \\
\partial_t\langle\sigma_y\rangle &= \frac{d_{\rm ge}}{\hbar}E_0\langle\sigma_z\rangle - \gamma_\perp\langle\sigma_y\rangle \\
\partial_t\langle\sigma_z\rangle &= -\frac{d_{\rm ge}}{\hbar}E_0\langle\sigma_y\rangle - \Gamma_R\left[\langle\sigma_z\rangle - \langle\sigma_z\rangle_0\right].
\end{aligned} \tag{5.600}$$

Here, E_0 is the slowly varying field amplitude as defined in Problem 1.6; the field is exactly resonant with the atoms; $\Gamma_R := \Gamma + R$; R is a pumping rate for the atoms to the excited state (by some auxiliary pump mechanism, not the cavity field); $\langle\sigma_z\rangle_0 := -(\Gamma - R)/(\Gamma + R)$ is the steady-state atomic inversion in the absence of coupling to the cavity field; κ is the energy decay rate of the cavity; and $d_{\rm ge} := \langle {\rm g}|d_z|{\rm e}\rangle$ is the matrix element that appears in the Rabi frequency. To do the derivation:

1. Start by writing down the optical Bloch equations on resonance. Assuming the atoms are initially unexcited, argue that $\langle\sigma_x\rangle$ can be removed from the problem. Note that the Rabi frequency here represents the coupling parameter of the atoms to the laser field.
2. Tack on an extra term of the form $-R[\langle\sigma_z\rangle - 1]$ to the $\langle\sigma_z\rangle$ Bloch equation to model excitation due to a pump field. *Justify* why in fact such a term gives the appropriate behavior.
3. Use the results of Problem 1.6 to write down an equation of motion for the cavity-field amplitude E_0, ignoring any spatial dependence of the field amplitude or the polarization amplitude. Also ignore any overlap of the circulating field with itself, as in a ring cavity. Use what you know about the polarization and the atomic dipoles to obtain the first term in the first Maxwell–Bloch equation above. Finally, tack on a damping term to model leakage of the cavity energy through the laser output port.

(b) Let $E_{0\text{ss}}$, $\langle\sigma_y\rangle_{\text{ss}}$, and $\langle\sigma_z\rangle_{\text{ss}}$ denote the steady-state values of E_0, $\langle\sigma_y\rangle$, and $\langle\sigma_z\rangle$, respectively. Now define the variables

$$r := \frac{\langle\sigma_z\rangle_0}{\langle\sigma_z\rangle_{\text{ss}}}; \qquad \sigma := \frac{\kappa}{2\gamma_\perp}; \qquad b := \frac{\Gamma_R}{\gamma_\perp}, \tag{5.601}$$

and define the scaled coordinates

$$\tau := \gamma_\perp t; \qquad x := \sqrt{b(r-1)}\frac{E_0}{E_{0\text{ss}}}; \qquad y := \sqrt{b(r-1)}\frac{\langle\sigma_y\rangle}{\langle\sigma_y\rangle_{\text{ss}}}; \qquad z := \frac{\langle\sigma_z\rangle_0 - \langle\sigma_z\rangle}{\langle\sigma_z\rangle_{\text{ss}}}. \tag{5.602}$$

Show that the Maxwell–Bloch equations may be written in the form[50]

$$\begin{aligned} \partial_\tau x &= -\sigma(x-y) \\ \partial_\tau y &= -y - xz + rx \\ \partial_\tau z &= xy - bz. \end{aligned} \tag{5.603}$$

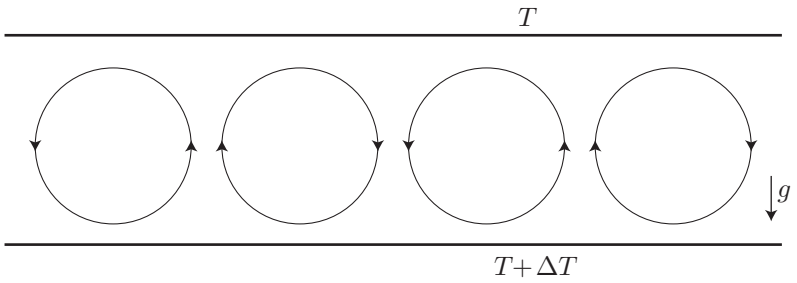
These equations are well-known in the area of nonlinear dynamics as the **Lorenz model**[51] (not Lorentz!), essentially one of the simplest possible nontrivial models for turbulent fluid flow or for the global weather system. Depending on the parameter values, the Lorenz model—and thus the Maxwell–Bloch equations for the laser—display a complex variety of dynamical behaviors, from steady behavior to periodic oscillation to chaos.[52]

The fluid interpretation of the Lorenz model is briefly as follows: the Lorenz model describes a fluid (subject to gravity) confined between two horizontal plates, where the lower plate is maintained a temperature ΔT higher than the upper plate. Then the parameter σ is a sort of scaled fluid viscosity (the *Prandtl number*), r is a scaled temperature difference ΔT (the *Rayleigh number*), and b is a scaled plate separation. Physically, for small temperature differences, the fluid simply supports the temperature gradient by *conduction*, without any movement. For larger temperature differences, the stationary behavior becomes unstable due to the buoyancy of the warm fluid near the bottom plate, and a convection pattern forms: a periodic pattern of ''rolls'' forms, transporting the warm fluid quickly to the upper plate, where it cools and then falls back down.

[50]This correspondence was first shown by H. Haken, ''Analogy Between Higher Instabilities in Fluids and Lasers,'' *Physics Letters A* **53**, 77 (1975).

[51]E. N. Lorenz, ''Deterministic Nonperiodic Flow,'' *Journal of Atmospheric Science* **20**, 130 (1963).

[52]Lorenz–Haken-type instability was observed in an ammonia laser, see C. O. Weiss and J. Brock, ''Evidence for Lorenz-Type Chaos in a Laser,'' *Physical Review Letters* **57**, 2804 (1986) (doi: 10.1103/PhysRevLett.57.2804). For a review, see Eugenio Roldán, G. J. de Valcárcel, R. Vilaseca, R. Corbalán, V. J. Martínez, and R. Gilmore, ''The dynamics of optically pumped molecular lasers. On its relation with the Lorenz–Haken model,'' *Quantum and Semiclassical Optics* **9**, R1 (1997) (doi: 10.1088/1355-5111/9/1/001).



For larger temperature differences, the flow becomes more complicated, but the Lorenz model doesn't capture this: it assumes physical quantities to be periodic in the horizontal direction with a single spatial frequency (i.e., truncating the Fourier series for the velocity, temperature, etc. fields after the first term). The coordinates x, y, and z are respectively the amplitude of the velocity modulation (i.e., the maximum upward velocity), the amplitude of the temperature modulation (i.e., the maximum horizontal temperature difference), and the dc offset of the temperature field from the conductive value. This general fluid problem is called **Rayleigh–Bénard convection**, and the Lorenz model is a greatly simplified description of this system.

You should find stability transitions at $r = 1$ and $r = (\sigma(\sigma + b + 3))/(\sigma - b - 1)$ (the latter transition requiring $\sigma > b + 1$).

(c) To analyze this laser model, begin by finding the **fixed points** or stationary solutions of the Lorenz system. That is, find the vectors (x^*, y^*, z^*) for which $\partial_\tau(x, y, z) = 0$. You should find two types of solutions, corresponding to conductive and convective fluid flow. Identify which is which. In terms of laser output, what is the (correct) interpretation of each fixed point?

(d) Perform a linear stability analysis to find the ranges of the parameter r for which the conductive and convective solutions are stable. You may assume σ, b and r to be positive for the purposes of this problem. Make a plot (sketch) of the fixed-point solutions, plotting x^* vs. r, indicating their stability on your plot.

To do the stability analysis, write each dynamical variable as a small perturbation to a fixed point,

$$x = x^* + \delta x; \qquad y = y^* + \delta x; \qquad z = z^* + \delta x, \tag{5.604}$$

and substitute these into the Lorenz equations, keeping only first order terms in the perturbations. The result is a set of linearized equations for δx, δy, and δz in the neighborhood of the fixed point (x^*, y^*, z^*). Now assume a solution of the form

$$\delta x(t) = \delta x(0)e^{\lambda t}; \qquad \delta y(t) = \delta y(0)e^{\lambda t}; \qquad \delta z(t) = \delta z(0)e^{\lambda t}, \tag{5.605}$$

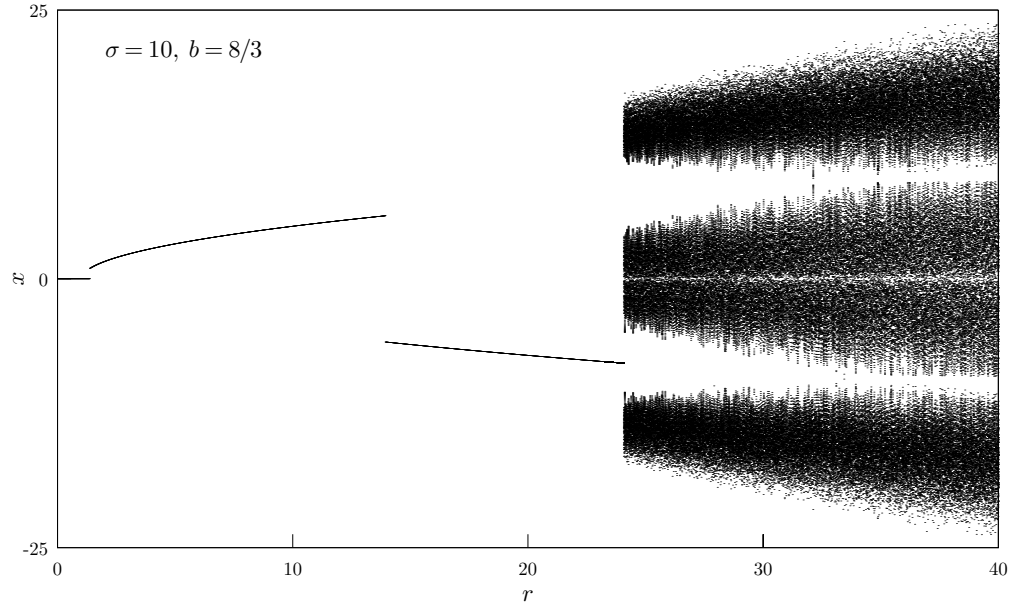
and use the linearized equations to find expressions for λ near each fixed point. If $\mathrm{Re}[\lambda] > 0$, then the solution runs away, and the fixed point is **linearly unstable**; If $\mathrm{Re}[\lambda] \leq 0$, then the solution remains bounded, and the fixed point is **linearly stable**.

Hint: you don't necessarily need to find the values of λ (i.e., the eigenvalues of the evolution matrix) in each case, you just need to determine the signs of the real parts.

(e) A universal feature in laser physics is **threshold behavior**. In a laser, what is a physical mechanism that prevents the unstable conductive solution from occuring? Viewing r as a ''rescaled'' version of the pumping rate R, interpret your solutions for the fixed points and their stability in terms of the steady-state laser output, and explain how these result predict a threshold in the laser output as a function of pump rate. Also, when a laser above threshold begins to oscillate (lase), when starting from a cavity initially in the vacuum state ($x = 0$), it is said to undergo **spontaneous symmetry breaking**. Explain what this term means in the context of your plot from part (d).

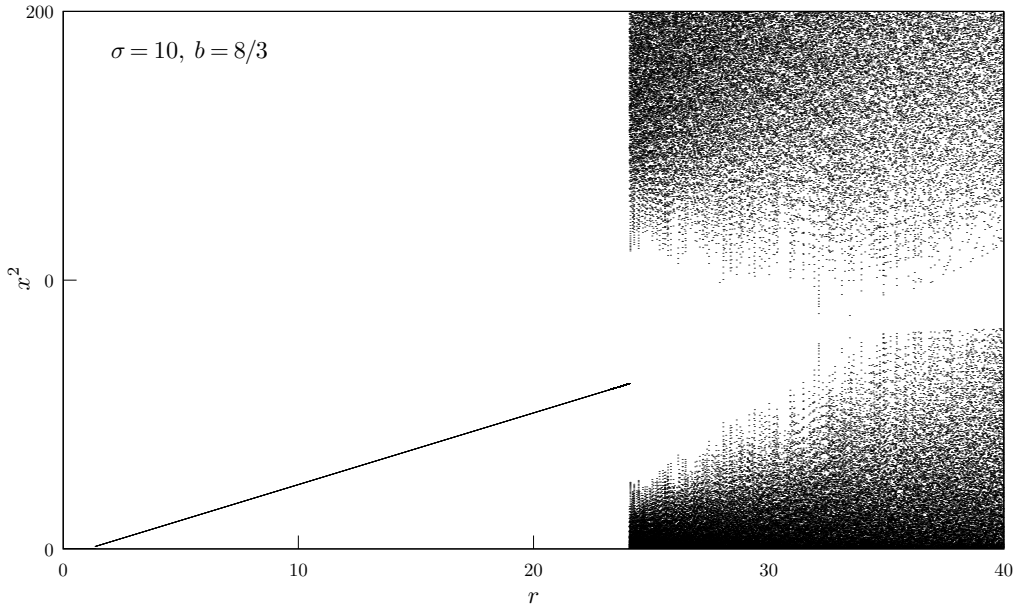
(f) Below is shown a **bifurcation diagram** for the x coordinate of the Lorenz model as a function of r, produced as follows. For each value of r start with some *generic* initial condition (specifically, $x = 0$,

$y = 0.01$, $z = 0.01$, corresponding to a bit of initial atomic excitation) and integrate the Lorenz model forward for a long time to get rid of any transient behavior. Then, integrate it for an even longer time; if the solution has settled to a stationary state, then plot the steady x^* value at coordinates (r, x^*); if it oscillates, plot a point on the graph at coordinates (r, x), each time the slope of the variable x changes sign. Thus, each trajectory will look like a bunch of points on your bifurcation diagram, and gives some idea of what happens at any value of r. For the purposes of this diagram, we have chosen the common values $\sigma = 10$ and $b = 8/3$.

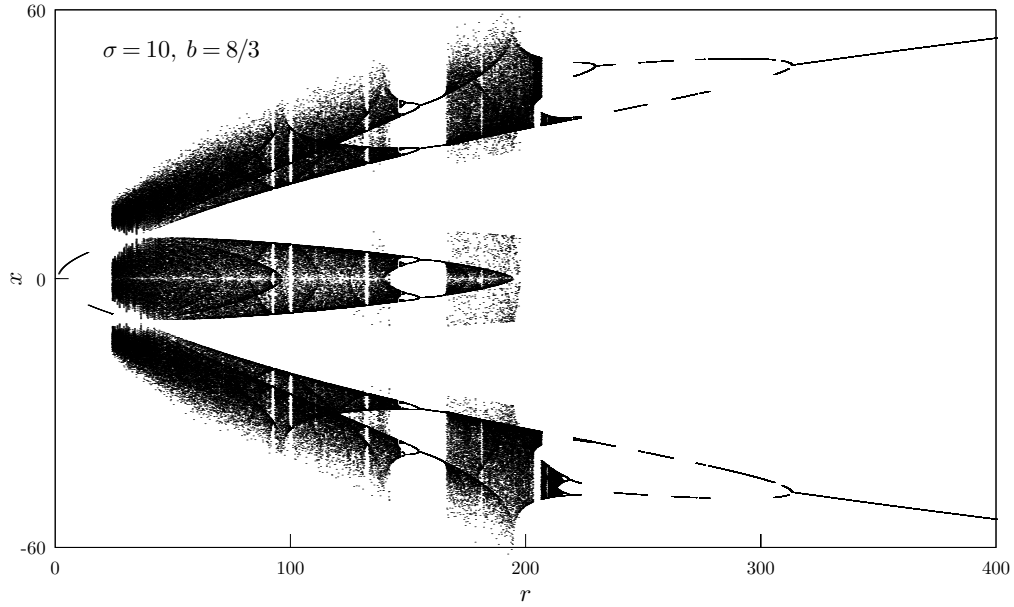


The plot above represents the parameter region that we'll mostly focus on. Explain how the behavior shown here is consistent with your fixed-point stability analysis.

Note that if you were to put a photodetector in front of your laser and measure the power as a function of the pump rate on your oscilloscope, you'd be looking at *intensity*, not the field amplitude, so you'd see something more like this:



Just for kicks, we can also look at a much wider range of r. Note that a large value of r corresponds to a very strong pump, and thus likely puts the laser into a highly nonlinear regime where this model breaks down. We would thus expect higher-order instabilities to make things more complicated. However, even within this model, the bifurcation diagram is quite intricate. You can see all sorts of behavior, such as an inverse period-doubling cascade, stability windows, and period-3 trajectories, which demonstrate that the model must exhibit chaotic behavior[53]



(g) Noting the symmetry of your steady-state sketch in (d), and the apparent reflection symmetry of the bifurcation diagrams, find the underlying symmetry in the Lorenz equations. Give a physical interpretation to the symmetry.

(h) Write a computer program to plot a trajectory of x vs. τ for $r = 0.5$, 10, 20, 30, and 350, with the other parameters and initial conditions as noted in (f). Interpret the behavior you see in each case, and comment on what the dynamical plot tells you about each relevant region in phase space. Note that any noisy behavior you see is likely to be chaos (deterministic ''randomness'' in low-dimensional systems).

(i) In part (g), for your plots at $r = 10$ and 20, you should have seen a phenomenon called **laser spiking** or **relaxation oscillation**, a transient, oscillatory behavior in the laser output when a laser parameter (usually, the pump power) is changed suddenly. It can be explained as follows. When the pump source is suddenly ''switched on,'' the strong pump rapidly excites the gain medium, such that the medium polarization build up to well beyond the threshold value before the cavity field can respond. The intensity, as a result, builds up rapidly, quickly depleting (saturating) the gain medium as the laser oscillates. The cavity field amplitude becomes so large that the gain-medium polarization drops below threshold, and soon afterwards, the field amplitude begins to drop. This process repeats, producing the oscillations.

For the $r = 10$ case, make a plot of $x(\tau)$ and $y(\tau)$ on the same graph, and explain how what you see is consistent with this explanation. Physically, why do these oscillations damp to a steady equilibrium?

(j) 20 years from now, you find yourself in the position of Über-Manager at Acme Laser Corp., overseeing the Nifty New Lasers division. One day, one of your many peons comes up to you in a panic, saying that the new high-power, ultrastable laser prototype that your division is now finally testing after years of engineering work turns out to not be very stable at all. In fact, the output intensity appears to be

[53] Tien-Yien Li and James A. Yorke, ''Period Three Implies Chaos,'' *The American Mathematical Monthly* **82**, 985 (1975) (doi: 10.2307/2318254).

very noisy whenever it is run at full power, even though the power supply is very quiet. Your first hunch, based on years of experience in laser building and tweaking, is that an inappropriate output coupler (output mirror) somehow made it into the design. But to fix things, should you try substituting an output coupler with greater or lesser reflectance than the one in the prototype now? Fortunately, ages ago, you worked out the above theory, which you can use as a semi-quantitative guide to laser stability. By working out how the changing output-coupler reflectance modifies the parameters in the Lorenz model, you make an intelligent guess that should put the laser back into a stable regime. What do you advise your employee, O Sage One? (Note that when you solved this problem years ago in a quantum optics class, you *explained your reasoning.*)

Here are some observations that may help:

1. Changing the mirror reflectance, and hence the value of κ, will affect both the values of r and σ. How do these parameters change with the mirror reflectance? This will not be obvious for r, so you should work out an expression for it in terms of κ and the other physical parameters.
2. Which way should you go in terms of r and σ to get back into a stable region?
3. You may assume that R is not much larger than Γ, as is typical for high-density gain media (for three-level lasers, technically speaking). Also, assume nearly natural broadening, so that $\gamma_\perp \sim \Gamma$. Finally, assume the laser to be in the ''bad cavity'' regime of $\kappa \gg \Gamma, \gamma_\perp$. This doesn't necessarily mean that the cavity is bad (in fact a truly bad cavity would violate the slowly varying amplitude approximation for the electric-field envelope), if the atomic transition is very narrow.

(k) You will get a reasonable answer with the above considerations, but you should keep the following caveat in mind. The Lorenz–Haken instability is unusual in practice, because lasers do not usually operate in the bad-cavity limit—usually, the atomic transition is broadened well beyond the width of the cavity resonance. Usually, though, other instabilities cause transitions to dynamical (though not necessarily chaotic) ''steady'' states.

In the more typical limit of strong broadening (large $\gamma_\perp$), when the atomic coherence can be adiabatically eliminated, show that the laser can then be modeled by the rate equations

$$\begin{aligned} \partial_t I &= \frac{\Gamma}{\gamma_\perp} N c \sigma_0 I \langle \sigma_z \rangle - \kappa I \\ \partial_t \langle \sigma_z \rangle &= -\frac{\Gamma}{\gamma_\perp} \frac{\sigma_0 I}{\hbar \omega_0} I \langle \sigma_z \rangle - \Gamma_R \Big[\langle \sigma_z \rangle - \langle \sigma_z \rangle_0 \Big], \end{aligned} \tag{5.606}$$

which are written only in terms of the cavity intensity I and the atomic inversion. Here $\sigma_0 = \lambda_0^2/2\pi$ is the resonant cross section for homogeneous broadening, with λ_0 the resonant wavelength.

Bonus question: is chaotic instability possible in this regime? Why or why not?

Problem 5.25

Consider again the forced, damped, quantum harmonic oscillator in the form

$$\partial_t \rho = -\frac{i}{\hbar}[H_0 + H_{\rm int}, \rho] + \gamma \mathcal{D}[a]\rho, \tag{5.607}$$

with free-oscillator Hamiltonian

$$H_0 = \hbar\omega_0 \left(a^\dagger a + \frac{1}{2} \right) \tag{5.608}$$

and driving Hamiltonian

$$H_{\rm int} = \hbar \mathscr{E} \left(a e^{i\omega t} + a^\dagger e^{-i\omega t} \right). \tag{5.609}$$

(a) Show that the expectation values evolve under the master equation as

$$\begin{aligned} \partial_t \langle a \rangle &= i\Delta \langle a \rangle - i\mathscr{E} - \frac{\gamma}{2} \langle a \rangle \\ \partial_t \langle a^\dagger a \rangle &= i\mathscr{E} \Big[\langle a \rangle - \langle a^\dagger \rangle \Big] - \gamma \langle a^\dagger a \rangle, \end{aligned} \tag{5.610}$$

once a transformation to a rotating frame has been made. Find the steady-state values for $\langle a \rangle$ and $\langle a^\dagger a \rangle$.

(b) Use the quantum regression theorem to compute the first-order coherence function

$$g^{(1)}(\tau) = \frac{\langle a^\dagger(t)\, a(t+\tau)\rangle_{t\to\infty}}{\langle a^\dagger a\rangle_{t\to\infty}}. \tag{5.611}$$

What is the emission spectrum $S(\omega_s)$ for the damped harmonic oscillator? Comment on how your result compares to the two-level atom modeled by the optical Bloch equations.

Hint: Use the fact that $\langle a(t)\rangle$ may be written in the form

$$\langle a(t)\rangle = g_0(t) + g_1(t)\langle a(0)\rangle + g_2(t)\langle a^\dagger(0)\rangle + g_3(t)\langle a^\dagger(0)a(0)\rangle\,, \tag{5.612}$$

and find the functions $g_\alpha(t)$. Then use the quantum regression theorem to write down an expression for $\langle a^\dagger(0)a(\tau)\rangle$.

Problem 5.26

Consider the quantum damped harmonic oscillator, with master equation

$$\partial_t \rho = -\frac{i}{\hbar}[H,\rho] + \gamma \mathcal{D}[a]\rho, \tag{5.613}$$

with Hamiltonian

$$H = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right). \tag{5.614}$$

The fact that damping occurs here implies that the oscillator also experiences a stochastic (heating) force. This is one way to view the zero-point energy: the damping tends to make the oscillator stop, but in the ground state the damping and stochastic heating exactly balance, producing a steady state of nonzero energy. This is also essentially the content of what is called the **fluctuation–dissipation relation** (see Section 14.3.8.1).

To see this, recall that the momentum diffusion coefficient D_p is be the rate at which the momentum variance increases,

$$\begin{aligned} D_p &= \partial_t V_p \\ &= \partial_t\left(\langle p^2\rangle - \langle p\rangle^2\right) \\ &= \langle pF\rangle + \langle Fp\rangle - 2\langle p\rangle\langle F\rangle\,, \end{aligned} \tag{5.615}$$

and we showed that it is given in terms of the force correlation function as

$$D_p = \int_{-\infty}^{\infty} d\tau \left[\langle F(t)F(t+\tau)\rangle - \langle F(t)\rangle^2\right]. \tag{5.616}$$

(a) Work in the weak-damping ($\gamma \ll \omega$) regime, and use the quantum regression theorem to find the value of the diffusion coefficient for the damped harmonic oscillator.

Hint: Express the force correlation function in terms of the ladder operators, using $a(t) \approx a(0)e^{-i\omega t}$ in the Heisenberg picture for $t \ll 1/\gamma$ to carry out any time derivatives. Then use the solution for $\langle a(t)\rangle$ at arbitrary times and the quantum regression theorem to find the correlation function.

(b) Show that in steady-state, the combined effect of damping and the stochastic force leads to the correct zero-point momentum variance (i.e., the momentum variance of the ground state) in steady state.

Chapter 6

Three-Level Atom Interacting with a Classical Field

Now having developed the theory of two-level atoms quite thoroughly, we will spend a bit more time examining slightly more complicated atoms, those with three levels. Some dramatic effects can take place in such systems due to quantum coherence and interference, much more so than in the two-level atom, due to multiple ''pathways'' between the different levels.

6.1 Stimulated Raman Transitions

One effect that we considered briefly in the context of Bragg scattering in an optical standing wave is the **stimulated Raman effect**. We consider the atomic energy level structure in the **Λ-configuration** shown below, where two ground states $|g_{1,2}\rangle$ are coupled to an excited state $|e\rangle$ by two optical fields. Our goal is to show that under suitable conditions, the atomic population can be driven between the ground states as in an effective, two-level system. In classical, nonlinear optics, you can think of the effect this way: two waves hitting the same atoms are mixed together by the nonlinearity of the atom, leading to an effective polarization wave at the beat (difference) frequency of the two waves, which drives the atomic transition between the ground states. Quantum mechanically, the coherence between the dipole moments associated with each transition causes them to work together, transferring population between the ground states without significantly populating the excited state. Of course, this is only possible with far-detuned excitations, so that spontaneous emission does not ''scramble'' the phases of the dipoles and ruin the quantum coherence. Thus, in this section we will stick to a Schrödinger-equation model, explicitly assuming far-off-resonant excitation to the excited state and thus ignoring spontaneous emission.

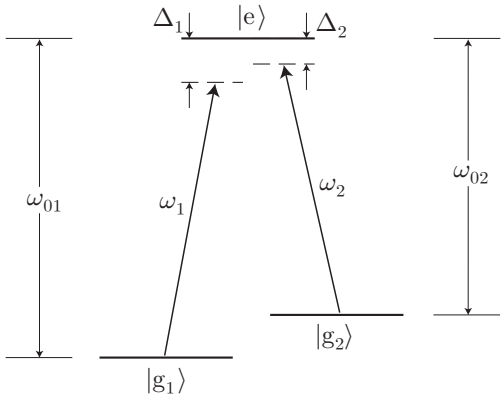
The combined optical field has the form

$$\begin{aligned}\mathbf{E}(\mathbf{r},t) &= \hat{\epsilon}_1 E_{01}\cos(\mathbf{k}_1\cdot\mathbf{r}-\omega_1 t)+\hat{\epsilon}_2 E_{02}\cos(\mathbf{k}_2\cdot\mathbf{r}-\omega_2 t)\\ &= \mathbf{E}^{(+)}(\mathbf{r},t)+\mathbf{E}^{(-)}(\mathbf{r},t),\end{aligned} \tag{6.1}$$

where $\mathbf{E}^{(\pm)}(\mathbf{r},t)$ are the positive and negative rotating components of the field, given by

$$\mathbf{E}^{(\pm)}(\mathbf{r},t)=\frac{1}{2}\left(\hat{\epsilon}_1 E_{01}e^{\pm i\mathbf{k}_1\cdot\mathbf{r}}e^{\mp i\omega_1 t}+\hat{\epsilon}_2 E_{02}e^{\pm i\mathbf{k}_2\cdot\mathbf{r}}e^{\mp i\omega_2 t}\right), \tag{6.2}$$

and $\hat{\epsilon}_{1,2}$ are the unit polarization vectors of the two fields.



The free atomic Hamiltonian can then be written

$$H_{\mathrm{A}} = \frac{p^2}{2m} - \hbar\omega_{01}|\mathrm{g}_1\rangle\langle\mathrm{g}_1| - \hbar\omega_{02}|\mathrm{g}_2\rangle\langle\mathrm{g}_2|, \tag{6.3}$$

where we have taken the excited state to have zero energy. In the dipole and rotating-wave approximations, the atom-field interaction Hamiltonian is

$$H_{\mathrm{AF}} = -\mathbf{d}^{(+)}\cdot\mathbf{E}^{(-)} - \mathbf{d}^{(-)}\cdot\mathbf{E}^{(+)}. \tag{6.4}$$

We have assumed that the ground-state splitting is much smaller than the optical transition frequencies: $\omega_{21} := \omega_{02} - \omega_{01} \ll \omega_{0\alpha}$ (for concreteness, we take $|\mathrm{g}_2\rangle$ to be of higher energy than $|\mathrm{g}_1\rangle$). Additionally, we have decomposed the dipole operator $\mathbf{d}$ into its positive- and negative-rotating components,

$$\begin{aligned}\mathbf{d} &= \mathbf{d}^{(+)} + \mathbf{d}^{(-)} \\ &= \Big[\langle\mathrm{g}_1|\mathbf{d}|\mathrm{e}\rangle\sigma_1 + \langle\mathrm{g}_2|\mathbf{d}|\mathrm{e}\rangle\sigma_2\Big] + \Big[\langle\mathrm{g}_1|\mathbf{d}|\mathrm{e}\rangle\sigma_1^\dagger + \langle\mathrm{g}_2|\mathbf{d}|\mathrm{e}\rangle\sigma_2^\dagger\Big],\end{aligned} \tag{6.5}$$

where $\sigma_\alpha := |\mathrm{g}_\alpha\rangle\langle\mathrm{e}|$ is an annihilation operator, and we assume the dipole matrix elements to be real (in principle, we should include a relative phase for this to be generally true, but this disappears anyway in the transformation to the rotating frame). Substituting (6.5) into (6.4), we find

$$\begin{aligned}H_{\mathrm{AF}} = &-\frac{1}{2}\langle\mathrm{g}_1|\hat{\epsilon}_1\cdot\mathbf{d}|\mathrm{e}\rangle E_{01}\Big(\sigma_1 e^{-i\mathbf{k}_1\cdot\mathbf{r}}e^{i\omega_1 t} + \sigma_1^\dagger e^{i\mathbf{k}_1\cdot\mathbf{r}}e^{-i\omega_1 t}\Big) \\ &-\frac{1}{2}\langle\mathrm{g}_2|\hat{\epsilon}_2\cdot\mathbf{d}|\mathrm{e}\rangle E_{02}\Big(\sigma_2 e^{-i\mathbf{k}_2\cdot\mathbf{r}}e^{i\omega_2 t} + \sigma_2^\dagger e^{i\mathbf{k}_2\cdot\mathbf{r}}e^{-i\omega_2 t}\Big).\end{aligned} \tag{6.6}$$

We will assume the detunings $\Delta_\alpha := \omega_\alpha - \omega_{0\alpha}$ are nearly equal; hence, to make this problem more tractable, we assume that the field $\mathbf{E}_\alpha$ couples only $|\mathrm{g}_\alpha\rangle$ to $|\mathrm{e}\rangle$. After solving this problem, we can treat the cross-couplings as a perturbation to our solutions. If we define the Rabi frequencies

$$\Omega_\alpha := \frac{-\langle\mathrm{g}_\alpha|\hat{\epsilon}_\alpha\cdot\mathbf{d}|\mathrm{e}\rangle E_{0\alpha}}{\hbar}, \tag{6.7}$$

which describe the strength of the coupling from level $|\mathrm{g}_\alpha\rangle$ through field $\mathbf{E}_\alpha$ to the excited level $|\mathrm{e}\rangle$, we arrive at

$$\begin{aligned}H_{\mathrm{AF}} = &\frac{\hbar\Omega_1}{2}\Big(\sigma_1 e^{-i\mathbf{k}_1\cdot\mathbf{r}}e^{i\omega_1 t} + \sigma_1^\dagger e^{i\mathbf{k}_1\cdot\mathbf{r}}e^{-i\omega_1 t}\Big) \\ &+\frac{\hbar\Omega_2}{2}\big(\sigma_2 e^{-i\mathbf{k}_2\cdot\mathbf{r}}e^{i\omega_2 t} + \sigma_2^\dagger e^{i\mathbf{k}_2\cdot\mathbf{r}}e^{-i\omega_2 t}\Big)\end{aligned} \tag{6.8}$$

as a slightly more compact form for the interaction Hamiltonian.

Now, before examining the equations of motion, we transform the *ground* states into the rotating frame of the laser field. This is like our transformation for the two-level atom, as in Section 5.1.5 on p. 154, but

here it is most convenient to use the excited state as the energy reference. Thus, writing the internal part of the state vector as

$$|\psi\rangle = c_1|\mathrm{g}_1\rangle + c_2|\mathrm{g}_2\rangle + c_\mathrm{e}|\mathrm{e}\rangle, \tag{6.9}$$

we can define the rotating-frame state vector by

$$|\tilde{\psi}\rangle = \tilde{c}_1|\mathrm{g}_1\rangle + \tilde{c}_2|\mathrm{g}_2\rangle + c_\mathrm{e}|\mathrm{e}\rangle, \tag{6.10}$$

where the slowly varying ground-state amplitudes are

$$\tilde{c}_\alpha(t) = c_\alpha(t)e^{-i\omega_\alpha t}. \tag{6.11}$$

Since the extra phase factors effectively boost the energies of the $|\mathrm{g}_\alpha\rangle$ states by $\hbar\omega_\alpha$, the dynamics in the rotating frame are generated by the rotating-frame, free-atom Hamiltonian, given by

$$\tilde{H}_\mathrm{A} = \frac{p^2}{2m} + \hbar\Delta_1|\mathrm{g}_1\rangle\langle\mathrm{g}_1| + \hbar\Delta_2|\mathrm{g}_2\rangle\langle\mathrm{g}_2|. \tag{6.12}$$

The interaction Hamiltonian in the rotating frame is

$$\begin{aligned}\tilde{H}_\mathrm{AF} &= -\mathbf{d}^{(+)}\cdot\tilde{\mathbf{E}}^{(-)} - \mathbf{d}^{(-)}\cdot\tilde{\mathbf{E}}^{(+)} \\ &= \frac{\hbar\Omega_1}{2}\Big(\sigma_1 e^{-i\mathbf{k}_1\cdot\mathbf{r}} + \sigma_1^\dagger e^{i\mathbf{k}_1\cdot\mathbf{r}}\Big) + \frac{\hbar\Omega_2}{2}\Big(\sigma_2 e^{-i\mathbf{k}_2\cdot\mathbf{r}} + \sigma_2^\dagger e^{i\mathbf{k}_2\cdot\mathbf{r}}\Big),\end{aligned} \tag{6.13}$$

where the slowly varying field amplitudes are given by $\tilde{E}_{0\alpha}^{(+)} := E_{0\alpha}^{(+)}e^{i\omega_\alpha t}$.

6.1.1 Effective Two-Level Dynamics

Turning to the equations of motion, we will manifestly neglect spontaneous emission, since $\Delta_\alpha \gg \Gamma$, where Γ is the decay rate of $|\mathrm{e}\rangle$, by using a Schrödinger-equation description of the atomic evolution. Then we have

$$i\hbar\partial_t|\psi\rangle = (\tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF})|\psi\rangle, \tag{6.14}$$

where the state vector can be factored into external and internal components as

$$|\psi\rangle = |\psi_{\mathrm{g}_1}\rangle|\mathrm{g}_1\rangle + |\psi_{\mathrm{g}_2}\rangle|\mathrm{g}_2\rangle + |\psi_\mathrm{e}\rangle|\mathrm{e}\rangle. \tag{6.15}$$

Then if $\psi_\alpha(\mathbf{r},t) := \langle\mathbf{r}|\psi_\alpha\rangle$, we obtain the equations of motion

$$\begin{aligned}i\hbar\partial_t\psi_\mathrm{e} &= \frac{p^2}{2m}\psi_\mathrm{e} + \frac{\hbar\Omega_1}{2}e^{i\mathbf{k}_1\cdot\mathbf{r}}\psi_{\mathrm{g}_1} + \frac{\hbar\Omega_2}{2}e^{i\mathbf{k}_2\cdot\mathbf{r}}\psi_{\mathrm{g}_2} - \hbar\Delta\psi_\mathrm{e} \\ i\hbar\partial_t\psi_{\mathrm{g}_1} &= \frac{p^2}{2m}\psi_{\mathrm{g}_1} + \frac{\hbar\Omega_1}{2}e^{-i\mathbf{k}_1\cdot\mathbf{r}}\psi_\mathrm{e} + \hbar(\Delta_1-\Delta)\psi_{\mathrm{g}_1} \\ i\hbar\partial_t\psi_{\mathrm{g}_2} &= \frac{p^2}{2m}\psi_{\mathrm{g}_2} + \frac{\hbar\Omega_2}{2}e^{-i\mathbf{k}_2\cdot\mathbf{r}}\psi_\mathrm{e} + \hbar(\Delta_2-\Delta)\psi_{\mathrm{g}_2},\end{aligned} \tag{6.16}$$

where we have boosted all energies by $-\hbar\Delta$, with $\Delta := (\Delta_1+\Delta_2)/2$ (i.e., setting $\psi_\alpha = [\psi_\alpha e^{i\Delta t}]\,e^{-i\Delta t}$, we identify the bracketed quantity $[\psi_\alpha e^{i\Delta t}]$ as the new state, which amounts to the replacement $\psi_\alpha \longrightarrow \psi_\alpha e^{-i\Delta t}$). Since we assume that $|\Delta_2-\Delta_1| \ll |\Delta|$, it is clear that ψ_e carries the fast time dependence at frequencies of order $|\Delta| \gg \Gamma$. We are interested in motion on timescales slow compared to $1/\Gamma$, and the fast oscillations are damped by coupling to the vacuum on timescales of $1/\Gamma$, so we can adiabatically eliminate ψ_e by making the approximation that it damps to equilibrium instantaneously ($\partial_t\psi_\mathrm{e} = 0$). Also, we use $p^2/2m \ll \hbar|\Delta|$, with the result,

$$\psi_\mathrm{e} = \frac{\Omega_1}{2\Delta}e^{i\mathbf{k}_1\cdot\mathbf{r}}\psi_{\mathrm{g}_1} + \frac{\Omega_2}{2\Delta}e^{i\mathbf{k}_2\cdot\mathbf{r}}\psi_{\mathrm{g}_2}. \tag{6.17}$$

Notice that in deriving this relation, it was important to choose the proper energy shift $-\hbar\Delta$ to minimize the natural rotation of the states that remain after the adiabatic elimination; indeed, if the resonance condition that we will derive is satisfied, the two ground states have no natural oscillatory time dependence. This procedure would be much more clear in a density-matrix treatment (as in Section 5.8.3.1), where the oscillating coherences would be eliminated, but this description is cumbersome due to the number of energy levels in the problem. Using this relation in the remaining equations of motion, we obtain two coupled equations of motion for the ground states,

$$\begin{aligned} i\hbar\partial_t\psi_{g_1} &= \frac{p^2}{2m}\psi_{g_1} + \left[\hbar\Delta_1 + \hbar\omega_{AC1}\right]\psi_{g_1} + \frac{\hbar\Omega_R}{2}e^{i(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{r}}\psi_{g_2} \\ i\hbar\partial_t\psi_{g_2} &= \frac{p^2}{2m}\psi_{g_2} + \left[\hbar\Delta_2 + \hbar\omega_{AC2}\right]\psi_{g_2} + \frac{\hbar\Omega_R}{2}e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}}\psi_{g_1}, \end{aligned} \tag{6.18}$$

where we have removed the energy shift of $-\hbar\Delta$. These equations are formally equivalent to the equations of motion for a two level atom, with Rabi frequency (**Raman** or **two-photon Rabi frequency**)

$$\Omega_R := \frac{\Omega_1\Omega_2}{2\Delta} \tag{6.19}$$
(Raman Rabi frequency)

and Stark shifts

$$\omega_{AC\alpha} := \frac{\Omega_\alpha^2}{4\Delta}. \tag{6.20}$$
(ac Stark shifts)

These equations of motion are just the equations generated by the effective Raman Hamiltonian

$$\begin{aligned} H_R = \frac{p^2}{2m} &+ \hbar(\Delta_1 + \omega_{AC1})|g_1\rangle\langle g_1| + \hbar(\Delta_2 + \omega_{AC2})|g_2\rangle\langle g_2| \\ &+ \frac{\hbar\Omega_R}{2}\left(\sigma_R e^{i(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{r}} + \sigma_R^\dagger e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}}\right), \end{aligned}$$

(effective, two-level Hamiltonian) (6.21)

where the Raman lowering operator is defined as $\sigma_R := |g_1\rangle\langle g_2|$. Noting that the operator $\exp(-i\mathbf{k}\cdot\mathbf{r})$ is a momentum-shift operator, so that $\exp(-i\mathbf{k}\cdot\mathbf{r})|\mathbf{p}\rangle = |\mathbf{p}-\hbar\mathbf{k}\rangle$ (and thus $\exp(-i\mathbf{k}\cdot\mathbf{r})\psi(\mathbf{p}) = \psi(\mathbf{p}+\hbar\mathbf{k})$, where $\psi(\mathbf{p}) := \langle\mathbf{p}|\psi\rangle$), it is clear from the form of the effective Raman Hamiltonian that a transition from $|g_2\rangle$ to $|g_1\rangle$ is accompanied by a kick of up to two photon-recoil momenta, and the reverse transition is accompanied by the opposite kick of up to two photon recoils. We can write out the coupled equations of motion due to the Hamiltonian (6.21) more explicitly as

$$\begin{aligned} i\hbar\partial_t\psi_{g_1}(\mathbf{p}) &= \left[\frac{p^2}{2m} + \hbar\Delta_1 + \hbar\omega_{AC1}\right]\psi_{g_1}(\mathbf{p}) + \frac{\hbar\Omega_R}{2}\psi_{g_2}(\mathbf{p}+2\hbar\delta\mathbf{k}) \\ i\hbar\partial_t\psi_{g_2}(\mathbf{p}+2\hbar\delta\mathbf{k}) &= \left[\frac{(\mathbf{p}+2\hbar\delta\mathbf{k})^2}{2m} + \hbar\Delta_2 + \hbar\omega_{AC2}\right]\psi_{g_2}(\mathbf{p}+2\hbar\delta\mathbf{k}) + \frac{\hbar\Omega_R}{2}\psi_{g_1}(\mathbf{p}), \end{aligned} \tag{6.22}$$

where $2\delta\mathbf{k} := \mathbf{k}_1 - \mathbf{k}_2$. The resonance condition for this transition $|\mathbf{p}\rangle|g_1\rangle \longrightarrow |\mathbf{p}+2\hbar\delta\mathbf{k}\rangle|g_2\rangle$ is

$$\left[\frac{(\mathbf{p}+2\hbar\delta\mathbf{k})^2}{2m\hbar} + \Delta_2 + \omega_{AC2}\right] - \left[\frac{p^2}{2m\hbar} + \Delta_1 + \omega_{AC1}\right] = 0, \tag{6.23}$$

which can be rewritten as

$$\Delta_R := 4\omega_R\left(\frac{p_\parallel + \hbar\delta k}{\hbar\delta k}\right) + (\Delta_2 - \Delta_1) + (\omega_{AC2} - \omega_{AC1}) = 0,$$

(Raman resonance condition) (6.24)

Here, $p_{\|}$ is the component of $\mathbf{p}$ along the direction of $\delta\mathbf{k}$, and we have defined the Raman recoil energy by $\hbar\omega_{\mathrm{R}} := \hbar^2(\delta k)^2/2m$ as the kinetic energy of an atom with momentum $\hbar\delta k/2$. The first term represents the Doppler shift of the two optical fields due to motion at the average of the upper and lower state momenta. Thus, we see that the stimulated Raman problem reduces to an effective two-level system of splitting $\hbar\Delta_{\mathrm{R}}$, coupled by a dc interaction of strength $\hbar\Omega_{\mathrm{R}}/2$.

6.1.1.1 Cross-Couplings

Finally, we account for the effects of the cross-couplings that we previously ignored. The lifetimes of the two ground states are in practice extremely long, so that the line width of the Raman transition is quite narrow, being limited only by the finite interaction time. Since it is assumed that the Raman resonance condition (6.23) is approximately true, the Raman cross-coupling is much further away from resonance than the intended coupling (typically several orders of magnitude), so this extra Raman coupling can be neglected in a secondary rotating-wave approximation. However, the cross-couplings can induce additional ac Stark shifts of the ground levels. So, we simply modify (6.20) to include these extra shifts:

$$\begin{aligned}\omega_{\mathrm{AC1}} &:= \frac{\Omega_1^2}{4\Delta} + \frac{\Omega_{1(2)}^2}{4(\Delta-\omega_{21})}\\ \omega_{\mathrm{AC2}} &:= \frac{\Omega_2^2}{4\Delta} + \frac{\Omega_{2(1)}^2}{4(\Delta+\omega_{21})}.\end{aligned}$$

(ac Stark shifts with cross-couplings) (6.25)

Here, $\Omega_{\alpha(\beta)}$ is the cross-coupling Rabi frequency for field α on transition $|\mathrm{g}_\beta\rangle \longrightarrow |\mathrm{e}\rangle$,

$$\Omega_{\alpha(\beta)} := \frac{-\langle \mathrm{g}_\beta|\hat{\epsilon}_\alpha\cdot\mathbf{d}|\mathrm{e}\rangle E_{0\alpha}}{\hbar}, \tag{6.26}$$

and we have assumed $\omega_{21} \gg |\Delta|$. These additional Stark shifts may not in general be negligible compared to the original Stark shifts.

6.1.2 Spontaneous Emission

We can also obtain an estimate of the spontaneous emission rate, which gives us a measure of how accurate our treatment is (since we have explicitly neglected it), by using (6.17) to write the total excited state population in terms of the density matrix elements:

$$\begin{aligned}R_{\mathrm{sc}} &= \Gamma\rho_{\mathrm{ee}}\\ &= \frac{\Gamma\Omega_1^2}{4\Delta^2}\rho_{\mathrm{g_1g_1}} + \frac{\Gamma\Omega_2^2}{4\Delta^2}\rho_{\mathrm{g_2g_2}} + \frac{\Gamma\Omega_1\Omega_2}{4\Delta^2}e^{i(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{r}}\rho_{\mathrm{g_1g_2}} + \frac{\Gamma\Omega_1\Omega_2}{4\Delta^2}e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}}\rho_{\mathrm{g_2g_1}}.\end{aligned} \tag{6.27}$$

Here, $\rho_{\alpha\alpha}$ is the population in state $|\alpha\rangle$, with $\rho_{\mathrm{g_1g_1}} + \rho_{\mathrm{g_2g_2}} \simeq 1$, and Γ is the total decay rate from the excited state. Note that this result assumes implicitly that $\Delta_1 \approx \Delta_2$. The second two terms represent an enhancement or suppression of spontaneous scattering due to atomic coherences; for example, the state

$$|\psi\rangle = \eta\left[\Omega_2 e^{-i\mathbf{k}_1\cdot\mathbf{r}}|\psi_{\mathrm{g_1}}\rangle - \Omega_1 e^{-i\mathbf{k}_2\cdot\mathbf{r}}|\psi_{\mathrm{g_2}}\rangle\right] \tag{6.28}$$

(where η is the appropriate normalization factor) is dark, since R_{sc} vanishes for this state. However, this state is only dark if the cross-couplings can be ignored. More realistically, the scattering rate can be modeled as an incoherent sum over all the couplings of the form $(\Gamma\Omega^2/4\Delta^2)\rho_{\mathrm{g}_\alpha\mathrm{g}_\alpha}$. This ''dark'' phenomenon is **coherent population trapping**, which we will treat in more detail below.

6.1.3 Multiple Excited States

It turns out we can work out this problem in the case where the ground states are coupled to *multiple* excited states $|\mathrm{e}_n\rangle$, as essentially always happens in real atoms. The idea is the same as above, except now the excited states have energies δ_n with respect to some arbitrary reference in the excited-state manifold

(say, one of the excited-state energies), such that $|\delta_n| \ll \omega_{0\alpha}$ (in our notation here, α enumerates the ground states, and n enumerates the excited states). Generalizing Eq. (6.7), we have Rabi frequencies

$$\Omega_{\alpha n} := \frac{-\langle \mathrm{g}_\alpha|\hat{\epsilon}_\alpha \cdot \mathbf{d}|\mathrm{e}_n\rangle E_{0\alpha}}{\hbar}, \tag{6.29}$$

for each possible transition. Then in the rotating frame the free atomic Hamiltonian is

$$\tilde{H}_{\mathrm{A}} = \frac{p^2}{2m} + \hbar\Delta_1|\mathrm{g}_1\rangle\langle\mathrm{g}_1| + \hbar\Delta_2|\mathrm{g}_2\rangle\langle\mathrm{g}_2| + \sum_n \hbar\delta_n|\mathrm{e}_n\rangle\langle\mathrm{e}_n|, \tag{6.30}$$

and the interaction Hamiltonian in the rotating frame is

$$\tilde{H}_{\mathrm{AF}} = \sum_n \frac{\hbar\Omega_{1n}}{2}\left(\sigma_{1n}e^{-i\mathbf{k_1}\cdot\mathbf{r}} + \sigma_{1n}^\dagger e^{i\mathbf{k_1}\cdot\mathbf{r}}\right) + \sum_n \frac{\hbar\Omega_{2n}}{2}\left(\sigma_{2n}e^{-i\mathbf{k_2}\cdot\mathbf{r}} + \sigma_{2n}^\dagger e^{i\mathbf{k_2}\cdot\mathbf{r}}\right), \tag{6.31}$$

where the lowering operators are now given by $\sigma_{\alpha n} := |\mathrm{g}_\alpha\rangle\langle\mathrm{e}_n|$. This setup is essentially the same as before, except for the summation of the excited states and the dependence of the detunings from the excited state on n. Following the same procedure as above, we find that the effective Raman Hamiltonian (6.21) is still valid, but where the Raman Rabi frequency is given by

$$\Omega_{\mathrm{R}} = \sum_n \frac{\Omega_{1n}\Omega_{2n}}{2(\Delta - \delta_n)},$$

(Raman Rabi frequency, multiple excited states) (6.32)

and the Stark shifts are given by

$$\omega_{\mathrm{AC}\alpha} = \sum_n \frac{\Omega_{\alpha n}^2}{4(\Delta - \delta_n)},$$

(ac Stark shifts, multiple excited states) (6.33)

which can be generalized as above to include cross-couplings and couplings to other levels. Ignoring any interference effects, the spontaneous-emission rate is a sum over terms of the form $\Gamma\Omega_n^2/4\Delta_n^2$ for every transition coupled by the fields.

6.1.4 Velocity Selectivity

From Eq. (6.24), the resonance condition for the stimulated Raman transition is

$$\Delta_{\mathrm{R}} = 4\omega_{\mathrm{R}}\left(\frac{p_\parallel + \hbar\delta k}{\hbar\delta k}\right) + (\Delta_2 - \Delta_1) + (\omega_{\mathrm{AC}2} - \omega_{\mathrm{AC}1}) = 0. \tag{6.34}$$

If we choose to ignore the atomic motion, we can do this by letting $m \longrightarrow \infty$, in which case $\omega_{\mathrm{R}} \longrightarrow 0$, and

$$\Delta_{\mathrm{R}} = (\Delta_2 - \Delta_1) + (\omega_{\mathrm{AC}2} - \omega_{\mathrm{AC}1}) = 0. \tag{6.35}$$

Thus, the transition is resonant if the detunings of the two fields to the excited states are equal, including any ac Stark shifts. However, in general, the resonance condition involves the atomic momentum. Noting that $\omega_{\mathrm{R}} \sim (\hbar\delta k)^2$, we can see that the atomic-velocity contribution is largest when δk is maximum. In particular, for counter propagating beams, $\mathbf{k} = \mathbf{k_2} \approx -\mathbf{k_1}$ (assuming a small ground-state splitting), we have $\delta\mathbf{k} \approx \mathbf{k}$, and the momentum change in the Raman transition is $2\hbar\mathbf{k}$ [this is, in fact, exact if we define $\mathbf{k} := (\mathbf{k}_1 - \mathbf{k}_2)/2$]. Solving for the atomic momentum in the resonance condition for the counterpropagating case, we find the resonant momentum

$$p_\parallel(\mathrm{g}_1) = \hbar k\frac{(\Delta_1 - \Delta_2) + (\omega_{\mathrm{AC}1} - \omega_{\mathrm{AC}2})}{4\omega_{\mathrm{R}}} - \hbar k \tag{6.36}$$

(resonant momentum)

for atoms in the $|g_1\rangle$ state, where in this case the Raman recoil frequency reduces to the usual recoil frequency: $\hbar\omega_R = \hbar^2k^2/2m = \hbar\omega_r$. After making the transition to the $|g_2\rangle$ state, the atoms have the momentum

$$p_\parallel(g_2) = \hbar k\frac{(\Delta_1 - \Delta_2) + (\omega_{AC1} - \omega_{AC2})}{4\omega_R} + \hbar k. \quad (6.37)$$
(resonant momentum)

In the *copropagating* case, when the ground-state splitting is small (say, a microwave transition as in the alkali atoms), the momentum recoil $\hbar\delta\mathbf{k}$ is several orders of magnitude smaller than the optical recoil $\hbar\mathbf{k}_{1,2}$. In this case, the momentum-dependent term makes an essentially negligible contribution to the resonance condition (6.34), because the Raman recoil frequency $\omega_R = \hbar(\delta k)^2/2m$ becomes $\hbar\omega_{21}^2/2mc^2$. For a microwave ground-state splitting $\omega_{21}/2\pi$ of 1 GHz, and a mass of 10^{-25} kg, we find $\omega_R/2\pi = 40$ nHz, compared to a recoil frequency of $\omega_r/2\pi = 4$ kHz if the optical transition is $\omega_{01} \approx \omega_{02} = 5 \times 10^{14}$ Hz. Thus, for typical ''long'' Raman pulses of ms durations, the velocity selectivity can be on the order of the atomic recoil or better for the counterpropagating case, but has essentially no selectivity on the scale of many recoils for the copropagating case.

Thus, in the counterpropagating configuration, the velocity dependence makes stimulated Raman transitions a valuable tool for atomic velocity selection.[1] The idea is fairly simple: to select atoms with a particular velocity, we simply drive a π-pulse between the atomic ground states, tuning the frequency difference between the two laser fields according to the resonance condition to select the desired velocity group. Since the frequency difference must be stable to typically better than the kHz level, the two fields are often generated from the same laser source by acousto-optic or electro-optic modulation, or they are generated by two separate lasers that are actively phase-locked (by detecting the beat note on a photodiode and using a phase-locked loop to feed back to the ''slave'' laser frequency). The line widths of the ground states are typically quite narrow compared to any laser line width, and so the width of the velocity selection is dominated by power-broadening of the two-photon transition. That is, there is an effective range of detunings on the order of Ω_R, so that the width of the selected momentum distribution is $\delta p_\parallel \approx \Omega_R/4\omega_r$ in the counterpropagating case. We will be more quantitative about the distribution below, but for now note that recoil-level resolution requires π-pulses of ms or longer durations. After the Raman pulse, the atoms near the desired momentum are ''tagged,'' by their internal state: if all the atoms start in $|g_1\rangle$, the atoms with the desired momentum end up in $|g_2\rangle$. This may be sufficient for some purposes, or the undesired atoms may be ''blown away'' by a resonant beam. For example, a beam that couples $|g_1\rangle$ to another excited state (that decays only to $|g_1\rangle$ and not to $|g_2\rangle$) can push the atoms away via radiation pressure.

One problem with this technique is extreme sensitivity to stray fields. Magnetic fields cause Zeeman shifts in otherwise degenerate levels of real atoms on the order of 0.1 MHz/G, where the Earth's magnetic field is around 0.7 G. But recoil-level velocity selection requires frequency precisions of kHz or better. Experimentally, stray magnetic fields must therefore be eliminated with compensation coils, high-μ metal shielding, and elimination of ferromagnetic materials in the vicinity of the atoms.

6.1.5 Pulse-Shape Considerations

6.1.5.1 Square Pulse

Since the velocity-selective Raman pulses (in the counterpropagating configuration) are generally used to ''tag'' a subset of an atomic distribution according to their momentum, it is important to consider the impact of the temporal pulse profile on the tagged distribution. The simplest pulse profile is the square profile, where the light is turned on at a constant intensity for some duration. Assuming that the atoms are all initially in the same internal atomic state, the tagging process is described by the solution of the optical Bloch equations for the excited state population of a two-level atom with Rabi frequency Ω_R, Raman detuning Δ_R (given by

[1] Mark Kasevich, David S. Weiss, Erling Riis, Kathryn Moler, Steven Kasapi, and Steven Chu, ''Atomic velocity selection using stimulated Raman transitions,'' *Physical Review Letters* **66**, 2297 (1991) (doi: 10.1103/PhysRevLett.66.2297); Kathryn Moler, David S. Weiss, Mark Kasevich, and Steven Chu, ''Theoretical analysis of velocity-selective Raman transitions,'' *Physical Review A* **45**, 342 (1992) (doi: 10.1103/PhysRevA.45.342).

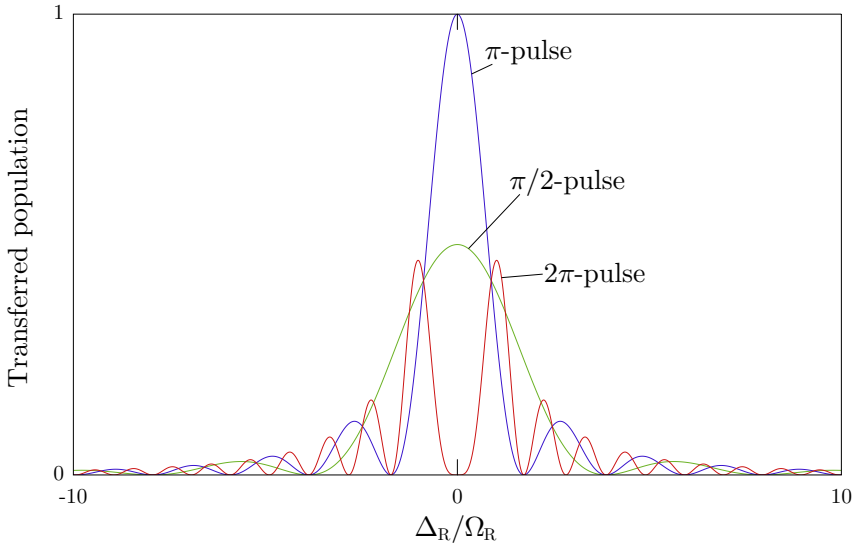
the left-hand side of Eq. (6.24)), and with all initial population in the ground Raman state $|g_1\rangle$:

$$\rho_{g_2g_2}(t) = \frac{\Omega_R^2}{\Omega_R^2 + \Delta_R^2} \sin^2\left(\frac{1}{2}\sqrt{\Omega_R^2 + \Delta_R^2}\, t\right). \tag{6.38}$$

The dynamics here are just the familiar generalized Rabi oscillations from the two-level atom, Eq. (5.60). From Eq. (6.24), we see that a detuning of $\Delta_R = 4\omega_r$ corresponds to a momentum shift of $\hbar k$. This line shape has wings that decay relatively slowly, with a series of locations where the line shape goes to zero. The locations of the zeros for an interaction time of δt is given by

$$\Delta_R = \sqrt{\frac{4n^2\pi^2}{(\delta t)^2} - \Omega_R^2} \tag{6.39}$$

for positive integer n. This relation simplifies for specific interaction times; for example, for a "π-pulse" of duration $\delta t = \pi/\Omega_R$, the locations are at $\Delta_R = \Omega_R\sqrt{4n^2 - 1}$, and for a $\pi/2$-pulse of duration $\delta t = \pi/(2\Omega_R)$, the locations are $\Delta_R = \Omega_R\sqrt{16n^2 - 1}$. These zeros were important in a previous implementation of Raman cooling,[2] where the first zero of the profile (6.38) was placed at zero momentum to form a dark interval where atoms would accumulate. The square-pulse excitation line shape is plotted in below for a $\pi/2$-pulse, a π-pulse, and a 2π-pulse.



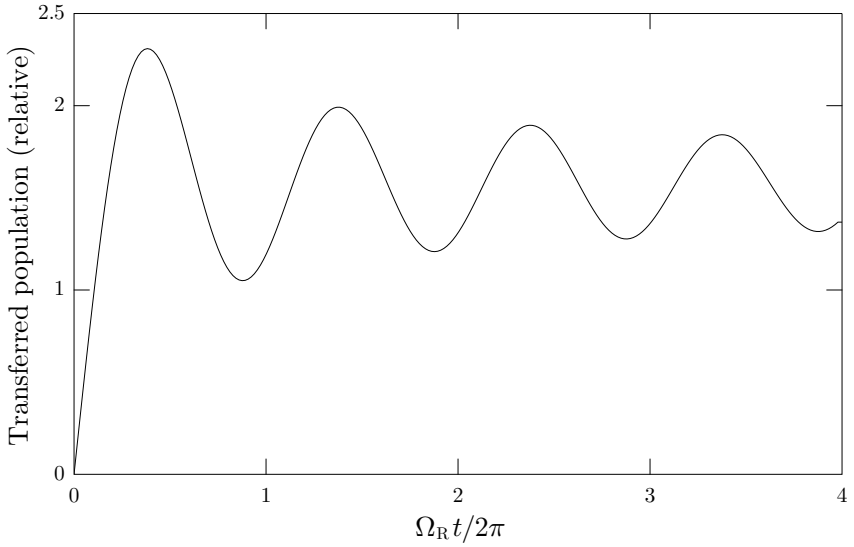
Note that for the important case of the π-pulse, the central population lobe is characterized by a half width at half maximum of $0.799 \cdot \Omega$.

It is also important to note that because one typically excites a range of detunings with a velocity-selective Raman pulse, the transferred population does not undergo simple sinusoidal Rabi oscillations. For a square pulse, the excitation profile (6.38) must be averaged over the atomic velocity distribution. In the limit of a broad velocity distribution, the excited population is proportional to

$$\begin{aligned}\int_{-\infty}^{\infty} \rho_{g_2g_2}(t)\, d\Delta_R &= \frac{\pi\Omega_R}{2} \mathit{Ji}_0(\Omega_R t) \\ &= \frac{\pi\Omega_R^2 t}{2}\left\{J_0(\Omega_R t) + \frac{\pi}{2}[J_1(\Omega_R t)\mathbf{H}_0(\Omega_R t) - J_0(\Omega_R t)\mathbf{H}_1(\Omega_R t)]\right\},\end{aligned} \tag{6.40}$$

where the $J_n(x)$ are ordinary Bessel functions, the $\mathbf{H}_n(x)$ are Struve functions, and $\mathit{Ji}_n(x) := \int_0^x J_n(x')dx'$. The population in this case still oscillates as a function of time, but with some effective damping due to dephasing of the different momenta.

[2] J. Reichel, F. Bardou, M. Ben Dahan, E. Peik, S. Rand, C. Salomon, and C. Cohen-Tannoudji, "Raman Cooling of Cesium below 3 nK: New Approach Inspired by Lévy Flight Statistics," *Physical Review Letters* **75**, 4575 (1995) (doi: 10.1103/PhysRevLett.75.4575); Jakob Reichel, *Refroidissement Raman et vols de Lévy: atomes de césium au nanokelvin*, Thése de Doctorat, École Normale Supérieure (1996).



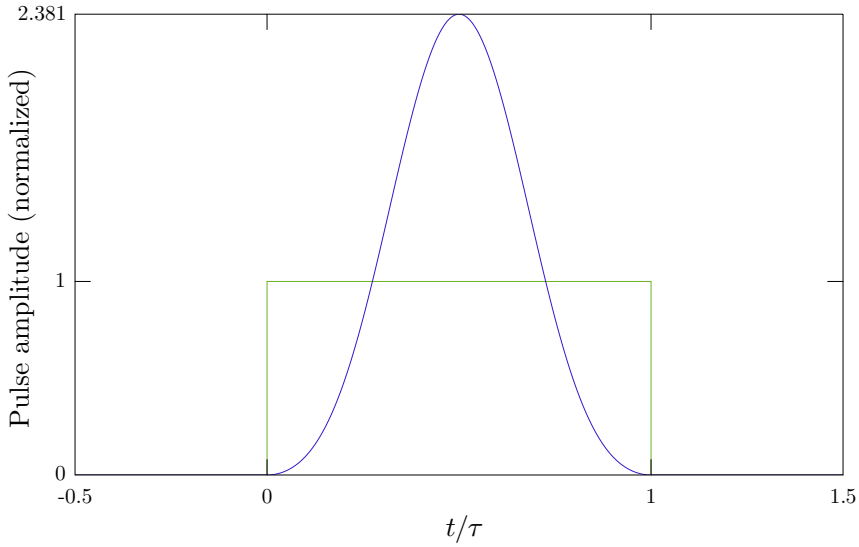
Notice that for short times, the function (6.40) reduces to $(\pi/2)\Omega_R^2 t + O(t^2)$, so that one can associate a nonzero transition rate, proportional to Ω_R^2 (which is in turn proportional to the product of the laser intensities), as long as $\Omega_R t \ll 1$.

6.1.5.2 Blackman Pulse

An alternative pulse profile, the *Blackman pulse profile*, is useful for suppressing the side lobes of the tagged distribution.[3] This profile, when normalized to have unit area, can be written as

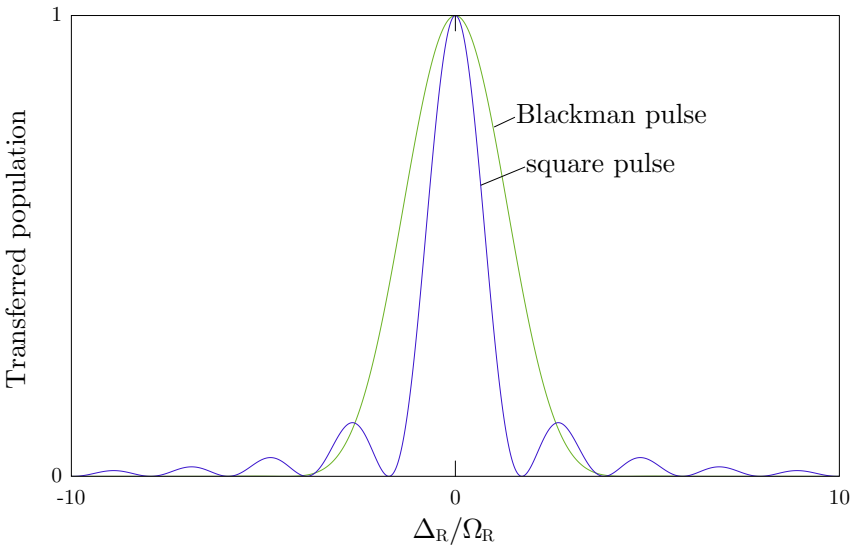
$$f_B(t) = \frac{1}{0.42\tau}[-0.5\cos(2\pi t/\tau) + 0.08\cos(4\pi t/\tau) + 0.42] \tag{6.41}$$

for $t \in [0, \tau]$ (with $f_B(t) = 0$ otherwise), where τ is the duration (support) of the pulse.



The Blackman profile has compact support and also, because it is continuous, has the property that the tails in the Fourier spectrum are suppressed relative to the square pulse. Hence, the Raman excitation spectrum of the Blackman pulse falls off much more sharply than the corresponding square-pulse spectrum.

[3] Mark Kasevich and Steven Chu, "Laser Cooling below a Photon Recoil with Three-Level Atoms," *Physical Review Letters* **69**, 1741 (1992) (doi: 10.1103/PhysRevLett.69.1741); Nir Davidson, Heun Jin Lee, Mark Kasevich, and Steven Chu, "Raman Cooling of Atoms in Two and Three Dimensions," *Physical Review Letters* **72**, 3158 (1994) (doi: 10.1103/PhysRevLett.72.3158).



However, the implementation of Blackman pulses is more complicated if the Raman beams induce an ac Stark shift of the transition, since the Raman frequency must be chirped to match the Stark shift in order to get good frequency resolution.

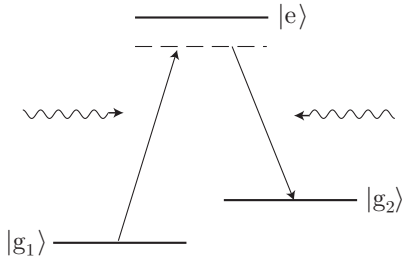
6.1.6 Stimulated Raman Cooling

6.1.6.1 Free Space

The velocity selectivity of stimulated Raman transitions makes them very useful for cooling atoms, and **stimulated Raman cooling** for neutral atoms has been successfully implemented.[4] The method, while difficult to implement in practice, has the advantage of very cold (subrecoil) temperatures without substantial losses of atoms (as in forced evaporation).

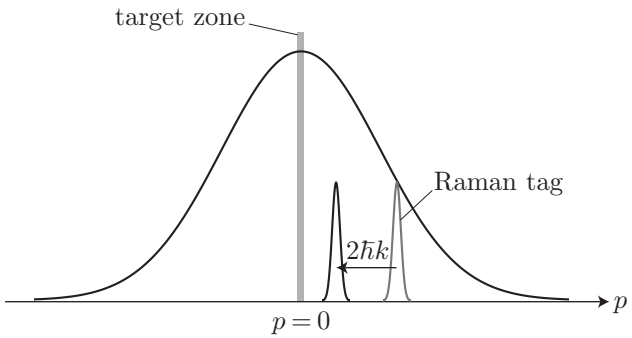
The important preliminary conceptual step is to define a ''target zone'' near $p = 0$, where the atoms will accumulate. Then we proceed with a cycle of steps. For simplicity, we'll consider only one dimension for the moment.

1. Start with all atoms in one state, say $|g_1\rangle$.
2. Tag all atoms *outside* the target zone by transferring them to the $|g_2\rangle$ state, using stimulated Raman pulses in the counterpropagating configuration.



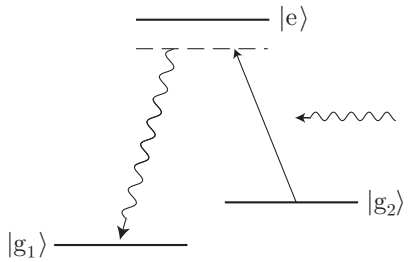
When tagging a particular velocity group, the orientation of the two beams should be such that the recoil of $2\hbar k$ moves the atoms towards the target zone near $p = 0$.

[4]Mark Kasevich and Steven Chu, ''Laser Cooling below a Photon Recoil with Three-Level Atoms,'' *Physical Review Letters* **69**, 1741 (1992) (doi: 10.1103/PhysRevLett.69.1741); J. Reichel, O. Morice, G. M. Tino, and C. Salomon, ''Subrecoil Raman Cooling of Cesium Atoms,'' *Europhysics Letters* **28**, 477 (1994); J. Reichel, F. Bardou, M. Ben Dahan, E. Peik, S. Rand, C. Salomon, and C. Cohen-Tannoudji, ''Raman Cooling of Cesium below 3 nK: New Approach Inspired by Lévy Flight Statistics,'' *Physical Review Letters* **75**, 4575 (1995) (doi: 10.1103/PhysRevLett.75.4575); H. J. Lee, C. S. Adams, M. Kasevich, and S. Chu, ''Raman Cooling of Atoms in an Optical Dipole Trap,'' *Physical Review Letters* **76**, 2658 (1996) (doi: 10.1103/PhysRevLett.76.2658).

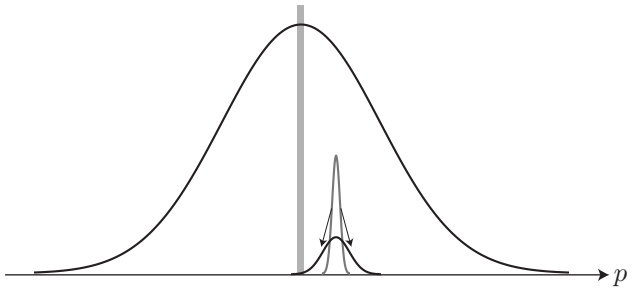


In general, a number of pulses are required to tag all of the necessary atoms. At higher momenta, shorter pulses (tagging wider momentum distributions) may be used, while near the target zone, long, high-resolution pulses are necessary.

3. Now "reset" the tagged atoms by applying light resonant with the $|g_2\rangle \longrightarrow |e\rangle$ transition, so that the atoms eventually decay back to the dark $|g_1\rangle$ state.



This can in general be done by beams from all directions, such as the optical molasses beams that are likely to be present anyway. In this case, the tagged atom distribution from the last step will be broadened in momentum by an amount on the order of $\hbar k$.



This **spontaneous Raman** step provides the dissipation or "exit channel" for entropy necessary for any cooling scheme to work.

4. Repeat the above sequence many times.

Why does this work so effectively? With the above sequence, all the atoms are essentially making a biased random walk towards the target zone. Ideally, once the atoms reach the target zone, they never leave it, because the velocity selectivity of the simulated Raman transitions. Even though the Raman pulses transfer momentum $2\hbar k$ at a time, the spontaneous emission allows the atom to move by fractions of a momentum recoil $\hbar k$, and thus the target zone can be narrower than $\hbar k$, and cooling below the recoil limit has been demonstrated with this method. Of course, the above idealization where the atoms are permanently stuck in the target zone is not quite true: the tails of the tagging distributions as well as off-resonant excitations determine a limited lifetime for atoms in the target zone. So long as this lifetime is much longer than the

time to iterate the above cycle, the cooling still works, and can be understood in terms of Lévy flights,[5] which amounts to diffusive behavior in momentum where every so often, the atoms ''stick'' to the region near $p = 0$ before diffusing again.

In three dimensions, the tagging must take place in all three dimensions on each iteration, so the target region is a small box in the three-dimensional momentum space. The much smaller target region (relative to the initial distribution) implies a much longer cooling time, but the method can still be made to work. Obviously this requires more juggling of laser beams, which makes the method quite challenging to implement. This is especially true considering the sensitivity of Raman transitions to magnetic fields that we discussed above, and subrecoil cooling requires extensive measures against stray fields.

6.1.6.2 Resolved-Sideband Raman Cooling

If atoms are bound in a tightly confining potential, another cooling method becomes possible if the splittings between the vibrational levels becomes much larger than the line width of the relevant atomic transition. In this case, the spectralsidebands are well-resolved, and the cooling method is known as **resolved-sideband Raman cooling**.

The basic idea is as follows. Assuming a nearly harmonic trapping potential of frequency $\omega_{\rm trap}$, we note that the bound atom oscillates mechanically at this frequency. If a monochromatic laser field impinges on the atom in the direction of motion, the atom thus sees a time-varying Doppler shift (phase-modulated wave) of the form

$$E^{(+)}(x,t) = E_0^{(+)} e^{i[kx - \omega t + \delta\phi \sin(\omega_{\rm trap} t)]}. \tag{6.42}$$

The instantaneous frequency is simply given by the time derivative of the phase (up to a minus sign), or $\omega - \delta\phi\,\omega_{\rm trap}\cos(\omega_{\rm trap} t)$, but in view of the decomposition

$$E^{(+)}(x,t) = E_0^{(+)} e^{ikx} e^{-i\omega t} e^{i\delta\phi \sin(\omega_{\rm trap} t)} = E_0^{(+)} e^{ikx} e^{-i\omega t} \sum_{j=-\infty}^{\infty} J_j(\delta\phi) e^{ij\omega_{\rm trap} t}, \tag{6.43}$$

we see that the spectrum is the ''carrier'' at frequency ω plus a sequence of sidebands at frequencies $\omega_j = \omega - j\omega_{\rm trap}$, where j is any nonzero integer. The above decomposition follows from the generating function for the Bessel functions:

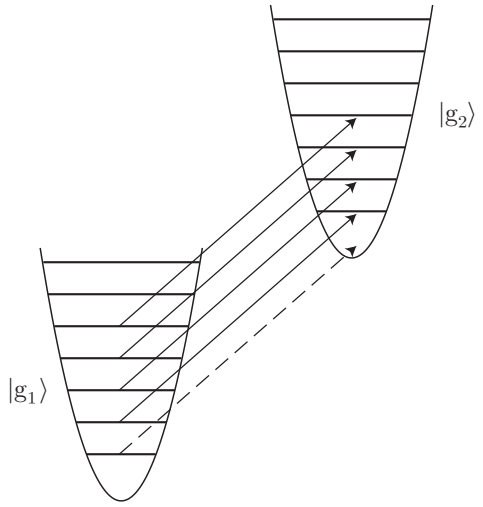
$$\exp\left[\frac{x}{2}\left(t - \frac{1}{t}\right)\right] = \sum_{j=-\infty}^{\infty} J_j(x) t^j. \tag{6.44}$$

The point is that the absorption spectrum of the bound atom consists of the usual atomic resonance ω_0, plus sidebands $\omega_0 + j\omega_{\rm trap}$ spaced at the trap frequency, assuming that the sidebands are well resolved (in the limit where $\omega_{\rm trap}$ is much larger than any decay rates for the ground states). When absorbing on one of the sidebands, energy conservation demands that along with the electronic transition, the vibrational state change by the appropriate number of quanta.

We can write down a recipe similar to that of free-space Raman cooling as follows.

1. Begin with all atoms in the same electronic state, say in $|{\rm g}_1\rangle$.
2. Drive a stimulated Raman transition on the $\omega_0 - \omega_{\rm trap}$ sideband. This implies transitions of the form $|{\rm g}_1, n\rangle \longrightarrow |{\rm g}_2, n-1\rangle$, where the integer labels the vibrational quantum number.

[5] J. Reichel, F. Bardou, M. Ben Dahan, E. Peik, S. Rand, C. Salomon, and C. Cohen-Tannoudji, ''Raman Cooling of Cesium below 3 nK: New Approach Inspired by Lévy Flight Statistics,'' *Physical Review Letters* **75**, 4575 (1995) (doi: 10.1103/PhysRevLett.75.4575).



The vibrational energy is thus reduced by one quantum. Note also that the $|g_1, 0\rangle$ state is dark, because the laser does not resonantly drive it to any other state.

3. Recycle the atoms to $|g_1\rangle$ by resonantly exciting it to the excited state. On average, the vibrational state does not change during the transition, particularly if the vibrational splitting is larger than the transition line width. Thus, on average, the atoms have reduced their vibrational energies by about one quantum.

4. Repeat.

At the end of many iterations, it is possible to find the atom in the ground state with near-unit probability. In three dimensions, all three relevant sidebands must be driven sequentially, assuming nondegenerate trap frequencies, and the beams must not be along a principle axis of the trap. This method has been successfully implemented in ion traps[6] as well as with neutral atoms in optical lattices.[7]

6.1.7 Atom Interferometry

One other application of stimulated Raman transitions is in the realization of atom interferometers, where atoms are split and recombined to effect sensitive physical measurements. The first atom interferometers were realized with thermal atomic beams, with the ''beam splitters'' realized by passing atoms through individual slits in physical aperture masks[8] or through microfabricated (absorptive) diffraction-grating masks.[9] Ultracold-atom interferometers lend themselves naturally to measurements of increased sensitivity due to the high degree of available control and potentially long interaction times.

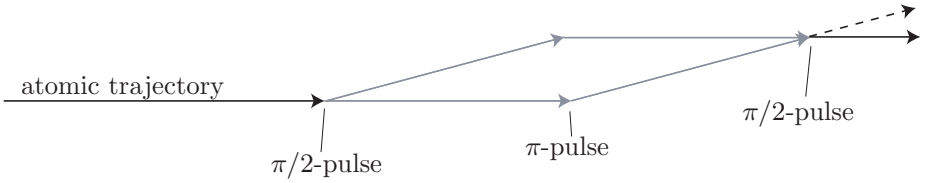
[6]C. Monroe, D. M. Meekhof, B. E. King, S. R. Jefferts, W. M. Itano, D. J. Wineland, and P. Gould, ''Resolved-Sideband Raman Cooling of a Bound Atom to the 3D Zero-Point Energy,'' *Physical Review Letters* **75**, 4011 (1995) (doi: 10.1103/PhysRevLett.75.4011).

[7]S. E. Hamann, D. L. Haycock, G. Klose, P. H. Pax, I. H. Deutsch, and P. S. Jessen, ''Resolved-Sideband Raman Cooling to the Ground State of an Optical Lattice,'' *Physical Review Letters* **80** 4149 (1998) (doi: 10.1103/PhysRevLett.80.4149); Vladan Vuletić , Cheng Chin, Andrew J. Kerman, and Steven Chu, ''Degenerate Raman Sideband Cooling of Trapped Cesium Atoms at Very High Atomic Densities,'' *Physical Review Letters* **81** 5768 (1998) (doi: 10.1103/PhysRevLett.81.5768); Andrew J. Kerman, Vladan Vuletić, Cheng Chin, and Steven Chu, ''Beyond Optical Molasses: 3D Raman Sideband Cooling of Atomic Ceium to High Phase-Space Density,'' *Physical Review Letters* **84** 439 (2000) (doi: 10.1103/PhysRevLett.84.439); Dian-Jiun Han, Steffen Wolf, Steven Oliver, Colin McCormick, Marshall T. DePue, and David S. Weiss, ''3D Raman Sideband Cooling of Cesium Atoms at High Density,'' *Physical Review Letters* **85** 724 (2000) (doi: 10.1103/PhysRevLett.85.724).

[8]O. Carnal and J. Mlynek, ''Young's Double-Slit Experiment with Atoms: A Simple Atom Interferometer,'' *Physical Review Letters* **66**, 2689 (1991) (doi: 10.1103/PhysRevLett.66.2689).

[9]David W. Keith, Christopher R. Ekstrom, Quentin A. Turchette, and David E. Pritchard, ''An Interferometer for Atoms,'' *Physical Review Letters* **66**, 2693 (1991) (doi: 10.1103/PhysRevLett.66.2693).

An atom interferometer based on stimulated Raman transitions might work as follows.[10] As atoms move slowly along (say, in free fall after being launched in an atomic fountain), they are exposed to a set of pulses from stimulated Raman lasers in the counterpropagating configuration.



If the atoms all start in one state $|g_1\rangle$, a $\pi/2$ Raman pulse puts them in a superposition of $|g_1\rangle$ and $|g_2\rangle$. The atoms in $|g_2\rangle$ have also suffered a momentum recoil of $2\hbar k$ in this configuration, and if the Raman lasers are oriented normally to the atoms' path, the atoms in the two states begin to separate transversely. Later, the atoms are exposed to a π Raman pulse, which exchanges the ground states as well as the transverse velocities of the two atomic groups, bringing them back together. When they again overlap, a final $\pi/2$ pulse mixes them and produces interference fringes. Thinking of this interferometer as analogous to the optical Mach–Zehnder interferometer, the $\pi/2$ pulses are analogous to (50/50) beam splitters, while the π pulse is analogous to a set of high reflectors.

Of course, any interaction that induces a relative phase between the two groups of atoms can be sensitively measured with this technique. One of the more successful applications is to the measurement of gravity. In the above figure, we can imagine that gravity points towards the bottom of the page. In this case, the phases accumulated by the two atomic groups during the respective ''tilted segments'' of their journeys should be the same. However, during the ''horizontal segments,'' the two atoms travel along paths with different gravitational potentials, and thus there is a phase shift given by $mg\Delta z\,\Delta t/2\hbar$, where m is the atomic mass, g is the local acceleration of gravity, Δz is the spatial separation of the two interferometer arms, and Δt is the time between the $\pi/2$-pulses. The local gravitational acceleration g has been measured with this system with a resolution of $\delta g/g \sim 10^{-8}$ for ~ 1 s integration times and $\sim 10^{-10}$ for integration times of ~ 1 day,[11] which begins to rival the current method of a falling corner-cube optical interferometer. Further, implementing a simultaneous pair of these measurements at different locations enables measurement of gravity gradients, which are otherwise quite difficult to measure.[12] Further, the sensitivity of these measurements may be greatly enhanced by using Bose–Einstein condensates in place of ordinary cold-atom clouds.[13] On the fundamental side, gravity and gravity gradient measurements enable measurements of the gravitational constant G and tests of general relativity, while on the applied side, interferometers are of technological interest for the detection of underground structures and reservoirs of oil and water, as well as completely passive navigation.

An alternate, but substantially equivalent, method of atom interferometry uses Bragg scattering from optical lattices as atomic beam splitters and mirrors.[14] As we discussed before, Bragg scattering can be viewed as a stimulated Raman process among momentum states, and thus the Bragg scatterings must be set to the equivalents of $\pi/2$- and π-pulses for beam splitters and high reflectors, respectively.

Because the recoil energy enters the resonance condition, a variation on the above interferometer enables the measurement of the fine-structure constant α, which is interesting from a fundamental perspective, since past measurements have had statistically significant discrepancies, and there is some speculation that the fine-structure constant may be time-dependent. The rough idea is that the resonance condition (6.24) for stimulated Raman transitions involves the recoil energy $\hbar\omega_r$ for Raman beams in the counterpropagating

[10] Mark Kasevich and Steven Chu, ''Atomic interferometry using stimulated Raman transitions,'' *Physical Review Letters* **67**, 181 (1991) (doi: 10.1103/PhysRevLett.67.181).

[11] A. Peters, K. Y. Chung, and S. Chu, ''High-precision gravity measurements using atom interferometry,'' *Metrologia* **38**, 25 (2001) (doi: 10.1088/0026-1394/38/1/4).

[12] M. J. Snadden, J. M. McGuirk, P. Bouyer, K. G. Haritos, and M. A. Kasevich, ''Measurement of the Earth's Gravity Gradient with an Atom Interferometer-Based Gravity Gradiometer,'' *Physical Review Letters* **81**, 971 (1998) (doi: 10.1103/PhysRevLett.81.971).

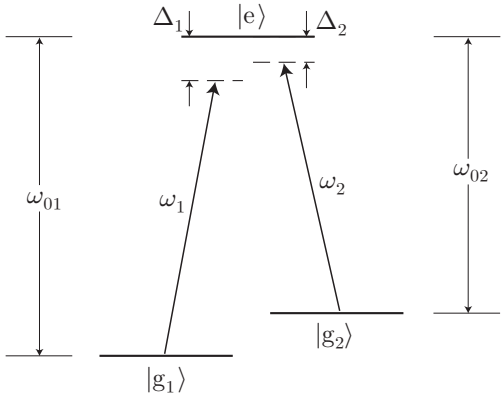
[13] P. Bouyer and M. A. Kasevich, ''Heisenberg-limited spectroscopy with degenerate Bose-Einstein gases,'' *Physical Review A* **56**, R1083 (1997) (doi: 10.1103/PhysRevA.56.R1083).

[14] David M. Giltner, Roger W. McGowan, and Siu Au Lee, ''Atom Interferometer Based on Bragg Scattering from Standing Light Waves,'' *Physical Review Letters* **75**, 2638 (1995) (doi: 10.1103/PhysRevLett.75.2638).

configuration. The recoil energy gives a measurement of h/m, which can be combined with an atomic-mass measurement to give the ratio $\hbar/m_{\mathrm{e}}$ where m_{e} is the electron mass. This ratio can then be converted to a measurement of α.[15]

6.2 Coherent Population Trapping

Another important effect, **coherent population trapping**[16] dramatically shows the influence of quantum interference in Λ atoms. Consider again the Λ atom from our discussion of stimulated Raman transitions of Section 6.1.



Following our previous treatment, but now ignoring center-of-mass motion of the atom, we can write the free-atomic Hamiltonian in the rotating frame as

$$\tilde{H}_{\mathrm{A}} = \hbar\Delta_1|\mathrm{g}_1\rangle\langle\mathrm{g}_1| + \hbar\Delta_2|\mathrm{g}_2\rangle\langle\mathrm{g}_2|. \tag{6.45}$$

Similarly, the atom-field interaction Hamiltonian in the rotating frame is

$$\tilde{H}_{\mathrm{AF}} = \frac{\hbar\Omega_1}{2}\left(\sigma_1 + \sigma_1^\dagger\right) + \frac{\hbar\Omega_2}{2}\left(\sigma_2 + \sigma_2^\dagger\right). \tag{6.46}$$

It turns out that if we make a change of basis for the ground states, one of the new states decouples from the excited state, which of course simplifies things. In particular, motivated by the spontaneous-emission results in the context of stimulated Raman transitions from Section 6.1.2, we can make the transformation

$$\begin{aligned} |\mathrm{g}_+\rangle &= \frac{1}{\sqrt{\Omega_1^2+\Omega_2^2}}\Big(\ \Omega_1|\mathrm{g}_1\rangle + \Omega_2|\mathrm{g}_2\rangle\Big) = \ \ \cos\theta|\mathrm{g}_1\rangle + \sin\theta|\mathrm{g}_2\rangle \\ |\mathrm{g}_-\rangle &= \frac{1}{\sqrt{\Omega_1^2+\Omega_2^2}}\Big(-\Omega_2|\mathrm{g}_1\rangle + \Omega_1|\mathrm{g}_2\rangle\Big) = -\sin\theta|\mathrm{g}_1\rangle + \cos\theta|\mathrm{g}_2\rangle, \end{aligned} \tag{6.47}$$

where the excited state is unchanged, and the rotation angle is defined by

$$\tan\theta = \frac{\Omega_2}{\Omega_1}. \tag{6.48}$$

(decoupling rotation angle)

[15] David S. Weiss, Brenton C. Young, and Steven Chu, "Precision measurement of the photon recoil of an atom using atomic interferometry," *Physical Review Letters* **70**, 2706 (1993) (doi: 10.1103/PhysRevLett.70.2706); Brenton Christopher Young, "A Measurement of the Fine-Structure Constant using Atom Interferometry," Ph.D. dissertation, Stanford University (1997).

[16] E. Arimondo and G. Orriols, "Nonabsorbing atomic coherences by coherent two-photon transitions in a three-level optical pumping," *Lettere al Nuovo Cimento della Societa Italiana di Fisica* **17**, 333 (1976) (doi: 10.1007/BF02746514); H. R. Gray, R. M. Whitley, and C. R. Stroud, Jr., "Coherent trapping of atomic populations," *Optics Letters* **3**, 218 (1978) (doi: 10.1364/OL.3.000218).

Clearly, the new states $|g_+\rangle$ and $|g_-\rangle$ are still normalized and orthogonal. Of course, the opposite basis change is given by reversing the rotation angle:

$$\begin{aligned} |g_1\rangle &= \cos\theta|g_+\rangle - \sin\theta|g_-\rangle \\ |g_2\rangle &= \sin\theta|g_+\rangle + \cos\theta|g_-\rangle. \end{aligned} \tag{6.49}$$

Then, we can put these into Eq. (6.45) to find the atomic Hamiltonian in the new basis,

$$\tilde{H}_{\rm A} = \hbar\Delta_+|g_+\rangle\langle g_+| + \hbar\Delta_-|g_-\rangle\langle g_-| + \hbar\Omega_{\rm g}\Big(|g_+\rangle\langle g_-| + |g_-\rangle\langle g_+|\Big),$$

(transformed free-atom Hamiltonian) (6.50)

where the rotated detunings are

$$\begin{aligned} \Delta_+ &= \cos^2\theta\,\Delta_1 + \sin^2\theta\,\Delta_2 \\ \Delta_- &= \sin^2\theta\,\Delta_1 + \cos^2\theta\,\Delta_2, \end{aligned} \tag{6.51}$$

and the coupling rate between the new states is

$$\Omega_{\rm g} = (\Delta_2 - \Delta_1)\sin\theta\cos\theta = (\Delta_2 - \Delta_1)\frac{\Omega_1\Omega_2}{\Omega_1^2 + \Omega_2^2}. \tag{6.52}$$

Similarly, the transformations (6.49) in Eq. (6.46) give the interaction Hamiltonian in the new basis

$$\tilde{H}_{\rm AF} = \frac{\hbar\Omega_+}{2}\left(\sigma_+ + \sigma_+^\dagger\right) + \frac{\hbar\Omega_-}{2}\left(\sigma_- + \sigma_-^\dagger\right), \tag{6.53}$$

where $\sigma_\pm := |g_\pm\rangle\langle e|$, and the new Rabi frequencies are

$$\begin{aligned} \Omega_+ &= \cos\theta\,\Omega_1 + \sin\theta\,\Omega_2 = \frac{\Omega_1^2 + \Omega_2^2}{\sqrt{\Omega_1^2 + \Omega_2^2}} = \sqrt{\Omega_1^2 + \Omega_2^2} \\ \Omega_- &= \cos\theta\,\Omega_2 - \sin\theta\,\Omega_1 = \frac{\Omega_1\Omega_2 - \Omega_2\Omega_1}{\sqrt{\Omega_1^2 + \Omega_2^2}} = 0. \end{aligned} \tag{6.54}$$

Thus, we may write the interaction Hamiltonian as

$$\tilde{H}_{\rm AF} = \frac{\hbar\sqrt{\Omega_1^2 + \Omega_2^2}}{2}\left(\sigma_+ + \sigma_+^\dagger\right),$$

(transformed interaction Hamiltonian) (6.55)

and thus we see that the coupling between $|g_-\rangle$ and $|e\rangle$ vanishes, while the Rabi frequency for the coupling of $|g_+\rangle$ to $|e\rangle$ is Ω_+.

Furthermore, at Raman resonance ($\Delta_1 = \Delta_2 = \Delta$), the free-atomic Hamiltonian simplifies quite dramatically. This is because $\Omega_{\rm g} = 0$, and $\Delta_+ = \Delta_- = \Delta$:

$$\tilde{H}_{\rm A} = \hbar\Delta\Big(|g_+\rangle\langle g_+| + |g_-\rangle\langle g_-|\Big). \tag{6.56}$$

(free evolution, Raman resonance)

Thus, we see that the free-atomic Hamiltonian becomes diagonal at Raman resonance (while $\tilde{H}_{\rm AF}$ is independent of detuning). We see in this case that under Hamiltonian evolution, $|g_-\rangle$ is *completely uncoupled* from $|g_+\rangle$ and $|e\rangle$.

Now, what about spontaneous emission? The operator form of the master equation reads

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}, \tilde{\rho}\right] + \Gamma_1\mathcal{D}[\sigma_1]\tilde{\rho} + \Gamma_2\mathcal{D}[\sigma_2]\tilde{\rho}, \tag{6.57}$$

where the Γ_α are the decay rates of $|\mathrm{e}\rangle$ to $|\mathrm{g}_\alpha\rangle$, so that the total decay rate of the excited state is $\Gamma = \Gamma_1 + \Gamma_2$. In the new basis, the master equation becomes

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \Gamma_+ \mathcal{D}[\sigma_+]\tilde{\rho} + \Gamma_- \mathcal{D}[\sigma_-]\tilde{\rho} + (\Gamma_2 - \Gamma_1)\sin\theta\cos\theta\left(\sigma_-\tilde{\rho}\sigma_+^\dagger + \sigma_+\tilde{\rho}\sigma_-^\dagger\right), \tag{6.58}$$

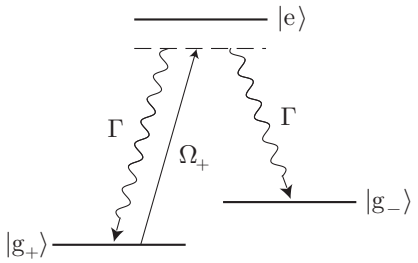
where

$$\begin{aligned} \Gamma_+ &= \cos^2\theta\,\Gamma_1 + \sin^2\theta\,\Gamma_2 \\ \Gamma_- &= \sin^2\theta\,\Gamma_1 + \cos^2\theta\,\Gamma_2. \end{aligned} \tag{6.59}$$

We see that the dissipation terms have a similar form in the new basis, but the last term is a correction term to handle asymmetric decay to $|\mathrm{g}_1\rangle$ and $|\mathrm{g}_2\rangle$. The point here is that $|\mathrm{e}\rangle$ decays as usual to both $|\mathrm{g}_+\rangle$ and $|\mathrm{g}_-\rangle$. Our basic conclusions will thus not be affected by the simplification $\Gamma_1 = \Gamma_2 = \Gamma/2 = \Gamma_+ = \Gamma_-$, so that

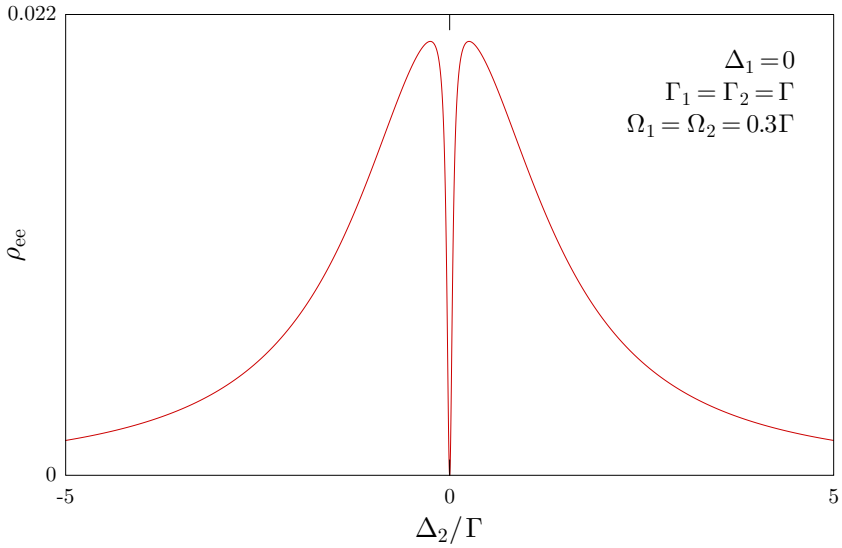
$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \Gamma\mathcal{D}[\sigma_+]\tilde{\rho} + \Gamma\mathcal{D}[\sigma_-]\tilde{\rho}. \tag{6.60}$$

We have thus arrived at a new effective Λ atom, where the excited state decays to both ground states, but only one ground state, $|\mathrm{g}_+\rangle$, is pumped by the external fields to the excited state.



Again, at Raman resonance, there is no coupling between $|\mathrm{g}_+\rangle$ and $|\mathrm{g}_-\rangle$, but away from resonance there is a coupling at rate Ω_g. Thus, at Raman resonance, $|\mathrm{g}_-\rangle$ is a **dark state**, and for $\Omega_+ \neq 0$, all the population will eventually end up in $|\mathrm{g}_-\rangle$. Thus, in steady state, the atom scatters no light. In the original basis, this is because the dipole moments for the two transitions either constructively or destructively interfere. If they destructively interfere, then the atom scatters no light, and effectively decouples from the field. If there is constructive interference, spontaneous emission scrambles the phases of the dipoles until the interference is purely destructive. This effect is **coherent population trapping**, because the population is ''trapped'' in $|\mathrm{g}_-\rangle$ due to quantum interference.

If we look at an absorption spectrum where one frequency, say ω_2, is swept, while the other is held fixed, we should expect to get the usual Lorentzian line shape for absorption, with width $\Gamma_1 + \Gamma_2$. However, we expect to see a dip in the absorption spectrum when $\Delta_1 = \Delta_2$.



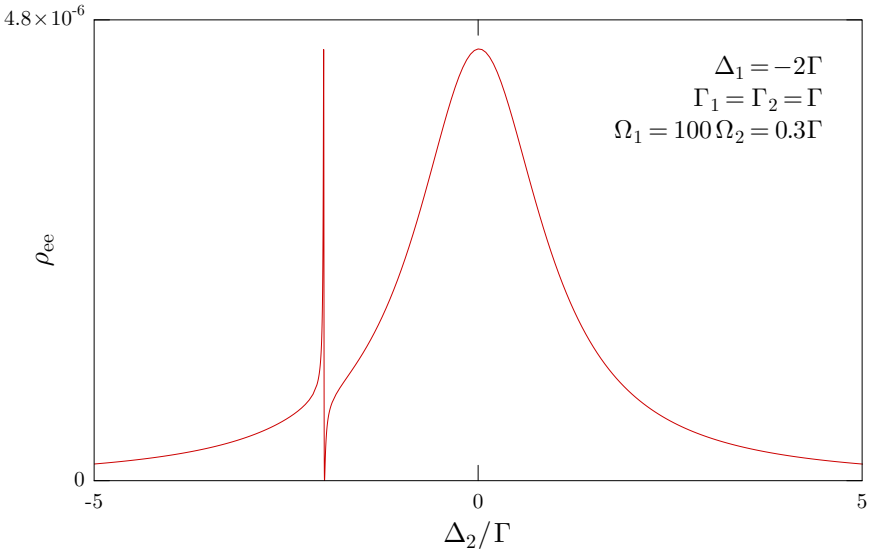
The dip can be quite narrow, and by solving for the steady state of the three-level optical Bloch equations it is possible to show[17] that the line shape is approximately the difference of two Lorentzians, one associated with the natural width of the excited state, and a narrow Lorentzian responsible for the coherent dip, of width (FWHM)

$$\delta\omega = \frac{\Omega_1^2 + \Omega_2^2}{\Gamma_1 + \Gamma_2} \qquad (6.61)$$

(width of coherent dip)

in the weak-driving limit. Since this is a coherent effect, the dip becomes suppressed and wider if any dephasing process reduces the coherence of the ground states.

If the first laser is detuned from resonance, and the second laser is made much weaker and swept in frequency across the first laser, a narrow, dispersive resonance occurs.



This shape has been shown[18] to be a Fano profile[19] (an asymmetric line shape that arises due to interference in ionization spectra).

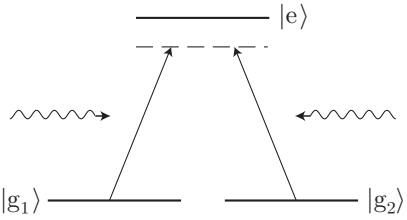
[17] H. R. Gray *et al.*, *op. cit.*

[18] B. Lounis and C. Cohen-Tannoudji, "Coherent population trapping and Fano profiles," *Journal de Physique II (France)* **2**, 579 (1992) (doi: 10.1051/jp2:1992153).

[19] U. Fano, "Effects of Configuration Interaction on Intensities and Phase Shifts," *Physical Review* **124**, 1866 (1961) (doi: 10.1103/PhysRev.124.1866)

6.2.1 VSCPT

If we again account for atomic motion, the dark-state condition ($\Delta_1 = \Delta_2$) must be modified to include the atomic momentum, and is essentially the same as the stimulated Raman resonance condition, Eq. (6.23). Doing so leads to a clever method for cooling atoms to extremely low temperatures, known as **velocity-selective coherent population trapping**, or **VSCPT**.[20] The idea is as follows. Suppose an atom with two degenerate ground states has both levels coupled to the excited state by two counterpropagating lasers of equal optical frequency. (This level structure occurs for an angular-momentum transition of the form $J = 1 \longrightarrow J' = 1$, such as occurs in He atoms.) We assume that each laser only couples one ground state to the excited state (due, e.g., to different polarizations of the two beams).



Then we can see that Doppler laser cooling works as usual, assuming that the common optical detuning is to the red of the atomic resonance: if the atom moves, it scatters photons preferentially from the opposing beam, which is shifted into resonance, thus tending to stop the atom. However, due to the three-level structure, there is a dark state, which for equal intensities is given by

$$|\psi\rangle_{\text{dark}} = \frac{1}{\sqrt{2}}\Big(|\text{g}_1, -\hbar k\rangle + |\text{g}_2, +\hbar k\rangle\Big), \tag{6.62}$$

once we have included the atomic motion. The *frequency width* of this dark state is set mostly by the common Rabi frequency as in Eq. (6.61), and for small intensities the dark-state width can be quite small, corresponding to a momentum width smaller than $\hbar k$. Thus, VSCPT gives rise to a sub-recoil cooling method, with the final momentum distribution consisting of two subrecoil peaks about $\pm\hbar k$. Note also that the final momentum state (the dark state) is, in fact, an entangled state between the atomic internal and external states. In this one-dimensional configuration, this method works best for collimation of atomic beams. However it is possible to extend this method to three dimensions.[21] It is also possible to apply this scheme to more complicated atoms with further degeneracy, so long as the lasers are pulsed to give a Ramsey-fringe-type effect.[22]

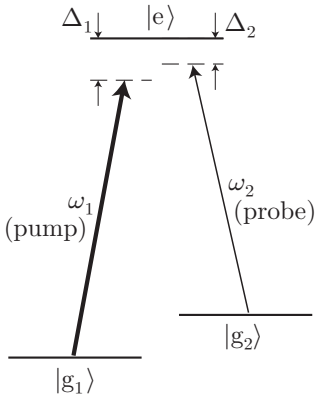
6.2.2 Electromagnetically Induced Transparency

One way to view the phenomenon of coherent population trapping is that the absorption profile for a field is modified by the presence of another field. Thinking of field 1 as a (strong) pump field, and field 2 as a (weak) probe field, the absorption coefficient for the probe can drop to zero due to the presence of the pump.

[20] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, ''Laser Cooling below the One-Photon Recoil Energy by Velocity-Selective Coherent Population Trapping,'' *Physical Review Letters* **61**, 826 (1988) (doi: 10.1103/PhysRevLett.61.826); A. Aspect, Ennio Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, ''Laser cooling below the one-photon recoil energy by velocity-selective coherent population trapping: theoretical analysis,'' *Journal of the Optical Society of America B* **6**, 2112 (1989); M. Widmer, M. R. Doery, M. J. Bellanca, W. F. Buell, T. H. Bergeman, and H. J. Metcalf, ''High-velocity dark states in velocity-selective coherent population trapping,'' *Physical Review A* **53**, 946 (1996) (doi: 10.1103/PhysRevA.53.946).

[21] M. A. Ol'shanii and V. G. Minogin, ''Three-dimensional velocity-selective coherent population trapping of a $(3+3)$-level atom,'' *Optics Communications* **89**, 393 (1992); J. Lawall, S. Kulin, B. Saubamea, N. Bigelow, M. Leduc, and C. Cohen-Tannoudji, ''Three-Dimensional Laser Cooling of Helium Beyond the Single-Photon Recoil Limit,'' *Physical Review Letters* **75**, 4194 (1995) (doi: 10.1103/PhysRevLett.75.4194).

[22] Frank Sander, Thibaut Devolder, Tilman Esslinger, and Theodor W. Hänsch, ''Ramsey-Type Subrecoil Cooling,'' *Physical Review Letters* **78**, 4023 (1997) (doi: 10.1103/PhysRevLett.78.4023).



This phenomenon is known as **electromagnetically induced transparency**, or **EIT**.[23]

Let's analyze this situation in a bit more depth, taking advantage of the assumption of a weak probe field. We will assume essentially the same form of the master equation as in (6.57)

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}, \tilde{\rho}\right] + \Gamma_1\mathcal{D}[\sigma_1]\tilde{\rho} + \Gamma_2\mathcal{D}[\sigma_2]\tilde{\rho} + \gamma_{\rm g}\mathcal{D}[\sigma_{\rm g}]\tilde{\rho}, \tag{6.63}$$

where we have added the final term (with $\sigma_{\rm g} := |{\rm g}_2\rangle\langle{\rm g}_2| - |{\rm g}_1\rangle\langle{\rm g}_1|$) to model coherence-relaxation processes between the ground states such as collisions in atomic vapors or dephasing due to local crystal fields. This master equation implies the equation for the probe-transition coherence

$$\begin{aligned}\partial_t\tilde{\rho}_{{\rm e}{\rm g}_2} &= \left(-\frac{\Gamma_2}{2} + i\Delta_2\right)\tilde{\rho}_{{\rm e}{\rm g}_2} + \frac{i\Omega_2}{2}\left(\rho_{\rm ee} - \rho_{{\rm g}_2{\rm g}_2}\right) - \frac{i\Omega_1}{2}\tilde{\rho}_{{\rm g}_1{\rm g}_2} \\ &\approx \left(-\frac{\Gamma_2}{2} + i\Delta_2\right)\tilde{\rho}_{{\rm e}{\rm g}_2} - \frac{i\Omega_2}{2} - \frac{i\Omega_1}{2}\tilde{\rho}_{{\rm g}_1{\rm g}_2}.\end{aligned} \tag{6.64}$$

Here, to lowest order in Ω_2 we have made the replacements $\rho_{{\rm g}_2{\rm g}_2} \approx 1$ and $\rho_{\rm ee} \approx 0$ to arrive at the second expression, since both are multiplied by Ω_2. We want this coherence in steady state, since it controls the photon-absorption rate of the probe, but it is coupled to $\tilde{\rho}_{{\rm g}_1{\rm g}_2}$. The equation of motion for this ground-state coherence is

$$\begin{aligned}\partial_t\tilde{\rho}_{{\rm g}_1{\rm g}_2} &= i\left(\Delta_2 - \Delta_1\right)\tilde{\rho}_{{\rm g}_1{\rm g}_2} - \gamma_{\rm g}\tilde{\rho}_{{\rm g}_1{\rm g}_2} - \frac{i\Omega_1}{2}\tilde{\rho}_{{\rm e}{\rm g}_2} + \frac{i\Omega_2}{2}\tilde{\rho}_{{\rm g}_1{\rm e}} \\ &\approx i[(\Delta_2 - \Delta_1) + i\gamma_{\rm g}]\tilde{\rho}_{{\rm g}_1{\rm g}_2} - \frac{i\Omega_1}{2}\tilde{\rho}_{{\rm e}{\rm g}_2}.\end{aligned} \tag{6.65}$$

Here, we have dropped the $\tilde{\rho}_{{\rm g}_1{\rm e}}$ term, since it is unpopulated to lowest order in Ω_2, and it already involves a factor of Ω_2. In steady state, $\partial_t\tilde{\rho}_{{\rm g}_1{\rm g}_2} = 0$, and solving the above equation gives

$$\tilde{\rho}_{{\rm g}_1{\rm g}_2} = \frac{\Omega_1\tilde{\rho}_{{\rm e}{\rm g}_2}}{2[(\Delta_2 - \Delta_1) + i\gamma_{\rm g}]}. \tag{6.66}$$

Then setting $\partial_t\tilde{\rho}_{{\rm e}{\rm g}_2} = 0$ to find the steady-state coherence, we find

$$\left(-\frac{\Gamma_2}{2} + i\Delta_2\right)\tilde{\rho}_{{\rm e}{\rm g}_2} - \frac{i\Omega_1}{2}\tilde{\rho}_{{\rm g}_1{\rm g}_2} = \frac{i\Omega_2}{2}. \tag{6.67}$$

Using Eq. (6.66) and solving for $\tilde{\rho}_{{\rm e}{\rm g}_2}$, we find

$$\tilde{\rho}_{{\rm e}{\rm g}_2} = \frac{i(\Omega_2/2)[(\Delta_2 - \Delta_1) + i\gamma_{\rm g}]}{(i\Delta_2 - \Gamma_2/2)[(\Delta_2 - \Delta_1) + i\gamma_{\rm g}] - i(\Omega_1/2)^2}. \tag{6.68}$$

[23]The first proposals for EIT were by Surya P. Tewari and G. S. Agarwal, "Control of Phase Matching and Nonlinear Generation in Dense Media by Resonant Fields," *Physical Review Letters* **56**, 1811 (1986) (doi: 10.1103/PhysRevLett.56.1811); and S. E. Harris, J. E. Field, and A. Imamoğlu, "Nonlinear optical processes using electromagnetically induced transparency," *Physical Review Letters* **64**, 1107 (1990) (doi: 10.1103/PhysRevLett.64.1107). For a good review, see Robert W. Boyd and Daniel J. Gauthier, "'Slow' and 'Fast' Light," in *Progress in Optics*, vol. 43, E. Wolf, ed. (Elsevier, Amsterdam, 2002), p. 497. Our treatment here parallels part of their discussion.

This coherence determines the optical properties of the medium, as far as the probe is concerned. The polarization is given by the dipole moment per unit volume, or

$$P^{(+)} = N\langle \mathrm{g}_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|\mathrm{e}\rangle\tilde{\rho}_{\mathrm{eg}_2} =: \epsilon_0\chi E^{(+)}, \tag{6.69}$$

where χ is the linear susceptibility of the atomic vapor, N is the number density of the atomic vapor, and $E^{(+)}$ is the positive-rotating electric field amplitude for the probe. Thus, we can write the susceptibility as

$$\chi = \frac{-iN|\langle \mathrm{g}_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|\mathrm{e}\rangle|^2}{\epsilon_0\hbar} \frac{[(\Delta_2 - \Delta_1) + i\gamma_\mathrm{g}]}{(i\Delta_2 - \Gamma_2/2)[(\Delta_2 - \Delta_1) + i\gamma_\mathrm{g}] - i(\Omega_1/2)^2}. \tag{6.70}$$

(EIT susceptibility)

Recall that the complex refractive index is the square root of the dielectric constant, so

$$\tilde{n} = \sqrt{1+\chi} \approx 1 + \frac{\chi}{2}, \tag{6.71}$$

since χ is small, assuming a rarefied medium. Also, taking the real part of the refractive index,

$$n = \mathrm{Re}[\tilde{n}] \approx 1 + \frac{\mathrm{Re}[\chi]}{2}. \tag{6.72}$$

The intensity absorption coefficient is related to the imaginary part of the refractive index by comparing the damping part of the plane-wave solution:

$$e^{-\mathrm{Im}[\tilde{n}]k_0 z} = e^{-(a/2)z}, \tag{6.73}$$

so that

$$a = 2k_0\mathrm{Im}[\tilde{n}] \approx k_0\,\mathrm{Im}[\chi]. \tag{6.74}$$

Thus, in this regime, the absorption coefficient for the probe is given by the imaginary part of χ.

We can thus see the induced transparency by looking at the Raman resonance $\Delta_1 = \Delta_2$, and for simplicity we will also consider the resonant case $\Delta_2 = 0$. In this case the susceptibility

$$\chi = \frac{iN|\langle \mathrm{g}_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|\mathrm{e}\rangle|^2}{\epsilon_0\hbar} \frac{\gamma_\mathrm{g}}{(\Gamma_2/2)\gamma_\mathrm{g} + (\Omega_1/2)^2} \tag{6.75}$$

(EIT susceptibility, $\Delta_1 = \Delta_2 = 0$)

becomes purely imaginary (i.e., the phase index becomes unity), and χ drops monotonically to zero with increasing pump intensity (Ω_1). Thus, we see how (on resonance) transparency for the probe is induced by the pump field.

The atomic medium is causal, and since the refractive index represents a causal response of the medium to the applied field, the real and imaginary parts of the complex refractive index obey the Kramers–Kronig relations[24]

$$\begin{aligned} \mathrm{Re}[\tilde{n}(\omega)] &= 1 + \frac{1}{\pi}\fint_{-\infty}^{\infty} \frac{\mathrm{Im}[\tilde{n}(\omega')]}{\omega' - \omega}\,d\omega' \\ \mathrm{Im}[\tilde{n}(\omega)] &= -\frac{1}{\pi}\fint_{-\infty}^{\infty} \frac{\mathrm{Re}[\tilde{n}(\omega')] - 1}{\omega' - \omega}\,d\omega'. \end{aligned} \tag{6.76}$$

The integrals here are **Hilbert transforms** (the cut integration symbols denote Cauchy-principal-value integrals), which are effectively convolutions with the kernel $1/\omega$. Since this kernel changes sign (and is largest in magnitude) near $\omega = 0$, the Hilbert transform is "something like a derivative," as long as you're willing to be not-too-quantitative. Since the coherent dip in EIT can be very narrow, as we saw from our analysis of coherent population trapping, the imaginary part of $\tilde{n}$ has large derivatives, and hence the real

[24] See Section 14.1.4.2, or for an alternate treatment see also Daniel A. Steck, *Classical and Modern Optics*, available online at `http://steck.us/teaching`.

part of $\tilde{n}$ (the phase index) can have large values and steep gradients. For a resonant, arbitrarily strong pump ($\Delta_1 = 0$, large Ω_1), the susceptibility (6.70) becomes

$$\chi = \frac{N|\langle g_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|e\rangle|^2}{\epsilon_0 \hbar} \frac{(i\gamma_g - \Delta_2)}{(\Omega_1/2)^2}, \tag{6.77}$$

so that the phase index $n = \mathrm{Re}[\tilde{n}] \approx \mathrm{Re}[\chi]/2$ becomes

$$n = -\frac{2N|\langle g_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|e\rangle|^2}{\epsilon_0 \hbar} \frac{\Delta_2}{\Omega_1^2}. \tag{6.78}$$

The group index of refraction is given by[25]

$$n_g = n + \omega \frac{dn}{d\omega}, \tag{6.79}$$

and it measures the ratio of the vacuum speed of light to the propagation velocity (group velocity) of an optical pulse. Assuming the second term is the most important, for the EIT medium the group index becomes

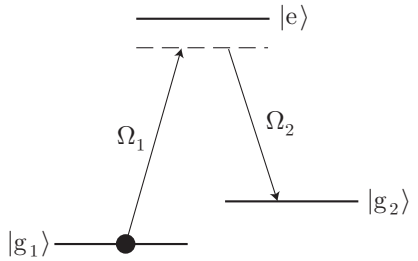
$$n_g \approx \frac{2\omega N|\langle g_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|e\rangle|^2}{\epsilon_0 \hbar \Omega_1^2}.$$

(EIT group index, $\Delta_1 = 0$, large Ω_1, small Δ_2) (6.80)

Putting in some typical numbers for alkali atoms ($\omega/2\pi = 5 \times 10^{14}$ Hz, $N = 10^{11}$ cm^{-3}, $\langle g_2|\hat{\varepsilon}_2 \cdot \mathbf{d}|e\rangle = 10^{-29}$ C·m, and $\Omega/2\pi = 10$ MHz), we find $n_g \approx 2 \times 10^4$! This also occurs where the medium is least absorbing, so long as the pulse spectrum is not too wide. Using this "slow light" technique, optical pulses have been slowed to 17 m/s.[26]

6.2.3 Stimulated Raman Adiabatic Passage

Now suppose in a three-level atom, you want to move all the population from $|g_1\rangle$ to $|g_2\rangle$, using two-photon stimulated Raman transitions.

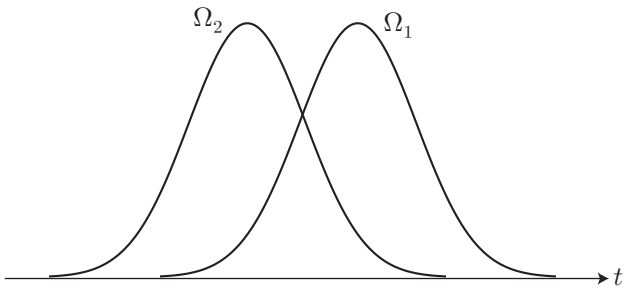


You can do the good old π-pulse, and of course you can chirp the Raman frequency to do adiabatic passage like in the two-level atom. However, because the dark states depend on the relative intensity, there is a different form of adiabatic passage, called **stimulated Raman adiabatic passage**, or **STIRAP**.[27] The idea is that if you have two laser pulses, one for each optical transition, you should do something counterintuitive: you should *first* turn on the laser coupling $|g_2\rangle \longrightarrow |e\rangle$, and then *later* turn on the laser coupling $|g_1\rangle \longrightarrow |e\rangle$.

[25] Daniel A. Steck, *op. cit.*

[26] Lene Vestergaard Hau, S. E. Harris, Zachary Dutton, and Cyrus H. Behroozi, "Light speed reduction to 17 metres per second in an ultracold atomic gas," *Nature* **397** (1999) (doi: 10.1038/17561). See also Michael M. Kash, Vladimir A. Sautenkov, Alexander S. Zibrov, L. Hollberg, George R. Welch, Mikhail D. Lukin, Yuri Rostovtsev, Edward S. Fry, and Marlan O. Scully, "Ultraslow Group Velocity and Enhanced Nonlinear Optical Effects in a Coherently Driven Hot Atomic Gas," *Physical Review Letters* **82**, 5229 (1999) (doi: 10.1103/PhysRevLett.82.5229).

[27] J. Oreg, F. T. Hioe, and J. H. Eberly, "Adiabatic following in multilevel systems," *Physical Review A* **29**, 690 (1984) (doi: 10.1103/PhysRevA.29.690); U. Gaubatz, P. Rudecki, M. Becker, S. Schiemann, M. Külz, and K. Bergmann, "Population switching between vibrational levels in molecular beams," *Chemical Physics Letters* **149**, 463 (1988) (doi: 10.1016/0009-2614(88)80364-6); U. Gaubatz, P. Rudecki, S. Schiemann, and K. Bergmann, "Population transfer between molecular vibrational levels by stimulated Raman scattering with partially overlapping laser fields. A new concept and experimental results," *Journal of Chemical Physics* **92**, 5363 (1990) (doi: 10.1063/1.458514); Martin Weitz, Brenton C. Young, and Steven Chu, "Atomic Interferometer Based on Adiabatic Population Transfer," *Physical Review Letters* **73**, 2563 (1994) (doi: 10.1103/PhysRevLett.73.2563).



The key is the *overlap* of the pulses, and the form of the dark state. When Ω_2 is large and $\Omega_1 = 0$, then clearly the dark state is $|g_1\rangle$. This represents the initial configuration. Similarly, when Ω_1 is large and $\Omega_2 = 0$, the dark state is $|g_2\rangle$, the desired final state. We showed above that there is a dark state $|-\rangle$ for any pair (Ω_1, Ω_2), and thus if we transform the field amplitudes adiabatically, slowly on time scales of $(\Omega_1^2 + \Omega_2^2)^{-1/2}$, then the atom will follow the dark state $|-\rangle$ until it reaches the final state $|g_2\rangle$. Since the atom is always in the dark state, there is no problem with spontaneous emission, even if the lasers are near resonance. Of course, with different pulse shapes it is possible to end in *any* superposition of the two ground states.

6.2.4 Quantum Beats

Until now we have discussed only the three-level atom in the Λ-configuration, but how do atoms in other configurations differ? One of the most significant differences is the possibility of **quantum beats** in resonance fluorescence. The basic idea is fairly simple[28] if we first consider the radiation from the Λ atom. The dipole moment is proportional to the annihilation operators for the two transitions,

$$d^{(+)} \sim \sigma_1 + \sigma_2, \tag{6.81}$$

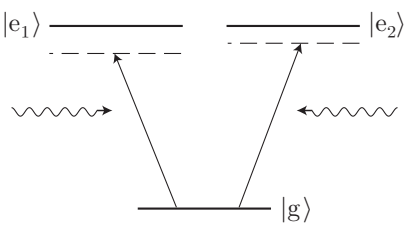
where $\sigma_\alpha := |g_\alpha\rangle\langle e|$. For simplicity we are dropping the dipole matrix elements, which may be different for the two transitions but do not affect our qualitative conclusions. The radiated field intensity thus scales as

$$\begin{aligned} \left\langle E^{(-)}E^{(+)} \right\rangle &\propto \left\langle d^{(-)}d^{(+)} \right\rangle \\ &\propto \left\langle \left(\sigma_1^\dagger + \sigma_2^\dagger\right)(\sigma_1 + \sigma_2) \right\rangle \\ &= 2\rho_{ee}, \end{aligned} \tag{6.82}$$

(radiated intensity, Λ-atom)

where we have used $\sigma_\alpha^\dagger \sigma_\beta = |e\rangle\langle g_\alpha|g_\beta\rangle\langle e| = |e\rangle\langle e|\delta_{\alpha\beta}$. Thus, the total radiation rate is proportional to the excited-state population.

The "vee" atom—where a single ground state $|g\rangle$ is coupled to two excited states $|e_1\rangle$ and $|e_2\rangle$—is more complicated, however.



If the two transitions decay into the same polarization mode, we can also write the dipole operator as

$$d^{(+)} \sim \sigma_1 + \sigma_2, \tag{6.83}$$

[28] P. W. Milonni, "Semiclassical and quantum-electrodynamical approaches in nonrelativistic radiation theory," *Physics Reports* **25**, 1 (1976).

where $\sigma_\alpha := |\mathrm{g}\rangle\langle \mathrm{e}_\alpha|$, and we have again dropped the dipole matrix elements. In this case, we have

$$\sigma_\alpha^\dagger \sigma_\beta = |\mathrm{e}_\alpha\rangle\langle \mathrm{g}|\mathrm{g}\rangle\langle \mathrm{e}_\beta| = |\mathrm{e}_\alpha\rangle\langle \mathrm{e}_\beta|, \tag{6.84}$$

and thus

$$\begin{aligned}\left\langle E^{(-)}E^{(+)}\right\rangle &\propto \left\langle d^{(-)}d^{(+)}\right\rangle \\ &\propto \left\langle \left(\sigma_1^\dagger + \sigma_2^\dagger\right)(\sigma_1+\sigma_2)\right\rangle \\ &= \rho_{\mathrm{e}_1\mathrm{e}_1} + \rho_{\mathrm{e}_2\mathrm{e}_2} + \tilde{\rho}_{\mathrm{e}_1\mathrm{e}_2} + \tilde{\rho}_{\mathrm{e}_2\mathrm{e}_1}.\end{aligned} \tag{6.85}$$

(radiated intensity, V-atom)

We see that the radiated intensity is proportional to the sum of the excited-state populations, which we might expect, but also the last two coherence terms represent *interferences* between the two populations. In the case where $|\mathrm{e}_1\rangle$ and $|\mathrm{e}_2\rangle$ have different energies, these coherences (transiently) rotate at the splitting frequency, thus leading to the quantum beats in the resonance fluorescence. This is the same beat note that we expect from any two radiating oscillators, but it goes to show that spontaneous emission isn't *entirely* coherent—interference effects are manifest in spontaneous emission.

The above argument leading to quantum beats rests on the assumption of decay into the same mode. If the radiation from the two transitions is *distinguishable*, say, if the two transitions radiated orthogonal polarizations, then the decay operators should *not* be added before taking the expectation value, and the quantum beats are not present.

6.2.4.1 Master Equations and Quantum Beats

The above argument addresses an ambiguity that arises when writing down the master equation for the three-level atom. For the Λ atom, we assumed in Eq. (6.57) that the master equation takes the form

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \Gamma_1\mathcal{D}[\sigma_1]\tilde{\rho} + \Gamma_2\mathcal{D}[\sigma_2]\tilde{\rho}. \tag{6.86}$$

(distinguishable radiation)

That is, we use separate dissipation terms for each operator, recalling that the Lindblad superoperator has the form

$$\mathcal{D}[c]\tilde{\rho} := c\tilde{\rho}c^\dagger - \frac{1}{2}\left(c^\dagger c\tilde{\rho} + \tilde{\rho}c^\dagger c\right). \tag{6.87}$$

Of course, we could have used a *single* dissipation term, had we added the operators together, to arrive at the master equation

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \mathcal{D}\left[\sqrt{\Gamma_1}\sigma_1 + \sqrt{\Gamma_2}\sigma_2\right]\tilde{\rho}. \tag{6.88}$$

(indistinguishable radiation)

Which one is correct? It depends on the physical situation. The master equation (6.86) corresponds to the case where the radiation from the two transitions is *distinguishable*, while the master equation (6.88) corresponds to the case where the radiation from the two transitions is *indistinguishable* (and thus we add the "amplitudes," or decay operators, for the two fields together before detection). This interpretation will be more clear when we study master equations in the context of continuous measurement, but the general rule is: we use separate decay terms when the decay processes can be monitored by separate detectors, while if the decay processes *can't* be distinguished, there is no "which-way" information, and we model the resulting potential interference via a single decay term.

To emphasize the difference between Eqs. (6.86) and (6.88), we note that we can rewrite Eq. (6.88) for the Λ atom as

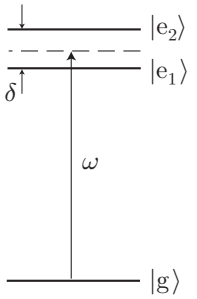
$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\mathrm{A}} + \tilde{H}_{\mathrm{AF}}, \tilde{\rho}\right] + \Gamma_1\mathcal{D}[\sigma_1]\tilde{\rho} + \Gamma_2\mathcal{D}[\sigma_2]\tilde{\rho} + \sqrt{\Gamma_1\Gamma_2}\left(\sigma_1\tilde{\rho}\sigma_2^\dagger + \sigma_2\tilde{\rho}\sigma_1^\dagger\right). \tag{6.89}$$

These last two terms involve the excited-state population and only couple to the ground-state coherences, and represent additional coherence induced between the ground states by spontaneous emission. Normally, the

indistinguishability is not such an important issue for Λ atoms, because even if the radiated polarizations are the same, the ''which-way'' information is provided by the atom itself, since we can in principle interrogate it to see which ground state it is in. However, we have already seen the coherent version of the master equation in the general form of Eq. (6.89) in the context of coherent population trapping, where the extra coherence terms popped up when we switched to the dark/bright-state basis in Eq. (6.58). In that case, the point was that each decay was to a *superposition* of the two ground states in the new basis, which was reflected by the additional coherence terms.

6.2.4.2 Steady-State Quantum Beats

A dramatic manifestation of the above difference between distinguishable and indistinguishable master equations occurs in a variant of the vee atom, resulting in something termed **steady-state quantum beats**.[29] The configuration is the vee atom from above, but in the case where the excited states are nondegenerate with splitting δ but both coupled from the ground state by the same field.



If we define the rotating-frame Hamiltonians, detunings Δ_α, Rabi frequencies Ω_α, and decay rates Γ_α ($\alpha \in \{1, 2\}$) in the usual way, then the Bloch equations for the excited-state populations have the form

$$\begin{aligned} \partial_t \rho_{e_1 e_1} &= \frac{i\Omega_1}{2}\left(\tilde{\rho}_{e_1 g} - \tilde{\rho}_{g e_1}\right) + (\text{dissipations terms}) \\ \partial_t \rho_{e_2 e_2} &= \frac{i\Omega_2}{2}\left(\tilde{\rho}_{e_2 g} - \tilde{\rho}_{g e_2}\right) + (\text{dissipations terms}). \end{aligned} \tag{6.90}$$

We may thus interpret the photon absorption rate to be the rate at which atoms are being excited, and thus

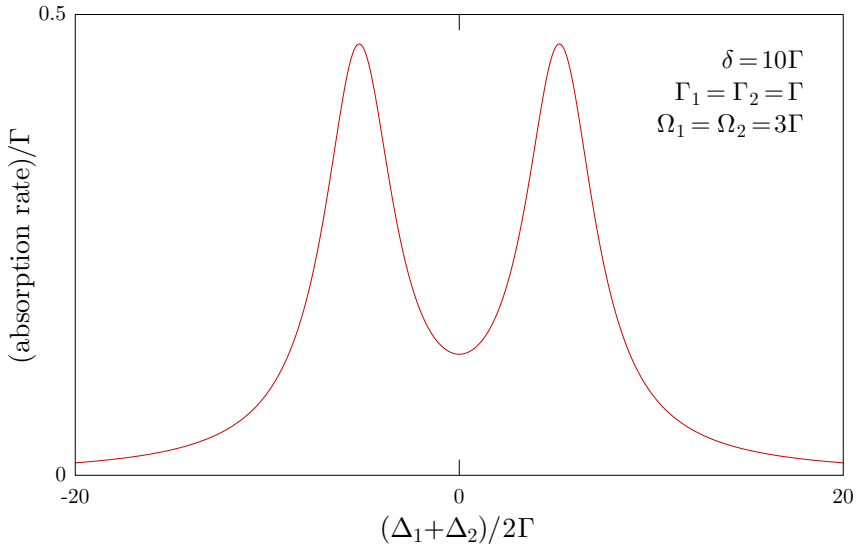
$$R_{\text{abs}} = \frac{i\Omega_1}{2}\left(\tilde{\rho}_{e_1 g} - \tilde{\rho}_{g e_1}\right) + \frac{i\Omega_2}{2}\left(\tilde{\rho}_{e_2 g} - \tilde{\rho}_{g e_2}\right) = \text{Im}\left[\Omega_1 \tilde{\rho}_{g e_1} + \Omega_2 \tilde{\rho}_{g e_2}\right]. \tag{6.91}$$

Again, there are two possible master equations that we can write down to describe this system. In the case where there is *distinguishable* emission from the two excited states (say, different polarizations), we use the master equation with separate decay terms:

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\text{A}} + \tilde{H}_{\text{AF}}, \tilde{\rho}\right] + \Gamma_1 \mathcal{D}[\sigma_1]\tilde{\rho} + \Gamma_2 \mathcal{D}[\sigma_2]\tilde{\rho}. \tag{6.92}$$

A sample absorption spectrum (in steady state) for this system is shown below, plotted as a function of the mean detuning $(\Delta_1 + \Delta_2)/2$, relative to the common decay rate Γ.

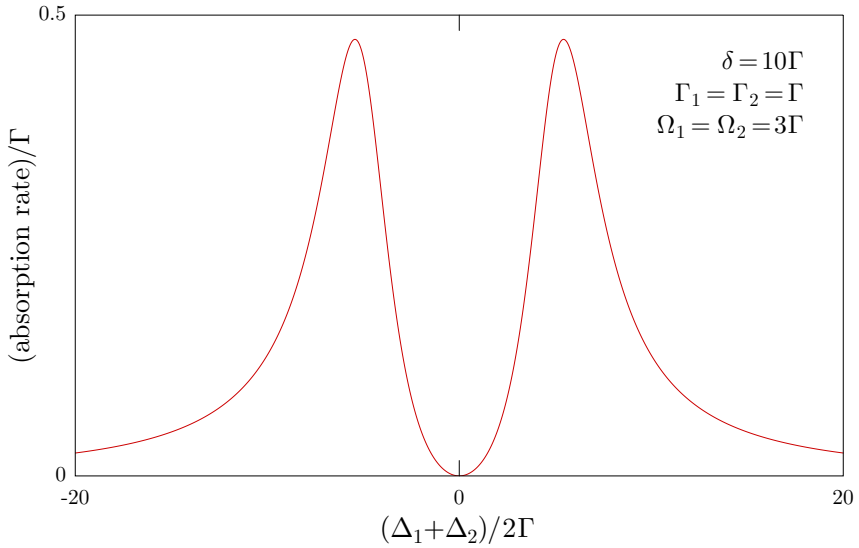
[29] D. A. Cardimona, M. G. Raymer, and C. R. Stroud, Jr., ''Steady-state quantum interference in resonance fluorescence,'' *Journal of Physics B: Atomic and Molecular Physics* **15**, 55 (1982) (doi: 10.1088/0022-3700/15/1/012).



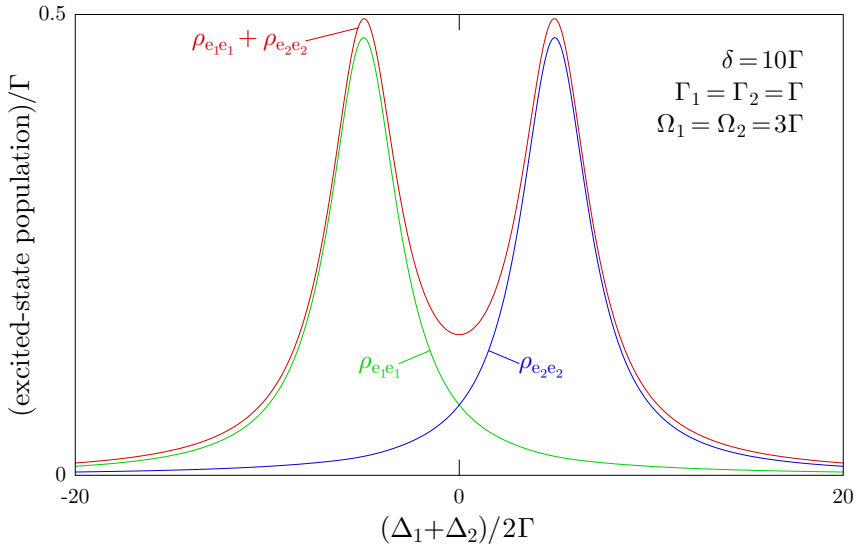
The spectrum consists of two peaks, as one might expect the sum of two Lorentzian peaks if there is no "interaction" between the transitions. For *indistinguishable* emission from the excited states, we again allow for interference of the radiated fields, and we use the master equation

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}, \tilde{\rho}\right] + \mathcal{D}\left[\sqrt{\Gamma_1}\sigma_1 + \sqrt{\Gamma_2}\sigma_2\right]\tilde{\rho}. \tag{6.93}$$

In this case, for the same parameters, something remarkable happens: the absorption *vanishes* at the midpoint between the peaks. This is an effect that persists even at high intensity (note that saturation effects are not negligible for the numerical example here in the plot for $\Omega = 3\Gamma$), and works even for more than two excited states.



This is clearly an interference effect, similar to coherent population trapping, but in a sense more remarkable because there is still population in the excited state, even when the atom is dark, as we can see by examining the total excited-state population.



Of course, the absorption rate must equal the emission rate in steady state, and we have seen that the emission rate is *not* just proportional to the total excited-state population, but rather to

$$\rho_{e_1e_1} + \rho_{e_2e_2} + \tilde{\rho}_{e_1e_2} + \tilde{\rho}_{e_2e_1}. \tag{6.94}$$

The coherences, or interference terms, prevent the atom from decaying even though the excited states are populated. Thus, coherent population trapping is due to interference in the *Hamiltonian* evolution of a Λ atom, while steady-state quantum beating is due to interference in the *dissipative* evolution of a vee atom.

A final amusing thing to note is that steady-state quantum beating gives rise to an alternate interpretation of EIT. In the Λ atom, you can imagine that the pump laser dresses the excited state, splitting it into a doublet (as in the Autler–Townes doublet). The probe beam thus effectively couples to a vee atom, and with the proper detuning, steady-state quantum beating suppresses absorption of the probe.

6.3 Exercises

Problem 6.1

A *phase-modulated* optical wave has the form

$$E^{(+)}(x,t) = E_0^{(+)}(x)e^{-i\omega t+\delta\phi\sin(\omega_{\rm mod}t)}, \tag{6.95}$$

where $\omega_{\rm mod}$ is the modulation frequency. Such a wave could result, for example, by running the wave through an *electro-optic crystal*, whose refractive index is modulated by an applied ac signal with frequency $\omega_{\rm mod}$.

(a) For a wave with time dependence $\exp[-i\phi(t)]$, we can define the *instantaneous frequency* as

$$\omega_{\rm inst} := \frac{d\phi}{dt}. \tag{6.96}$$

Compute the instantaneous frequency of the phase-modulated wave and thus show that the frequency oscillates about ω. That is, phase modulation is in some sense the same as frequency modulation.

(b) Write the phase-modulated wave as a sum of plane waves, with the general form

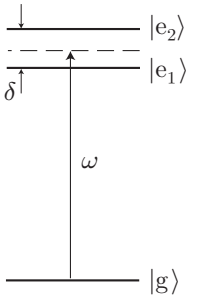
$$\sum_{j=-\infty}^{\infty} c_j e^{i(kx-\omega_j t)}. \tag{6.97}$$

Hint: start by writing down a Bessel series for the function $\exp(iK\sin x)$, using the generating function given in the notes.

(c) From your answer to (b), argue that the intensity spectrum (as viewed through a Fabry–Perot spectrum analyzer) consists of a series of peaks with relative intensity $J_j^2(\delta\phi)$. You may assume the response of the Fabry–Perot analyzer is slow compared to the modulation frequency. This phase-modulation technique is commonly used in the laboratory to shift the frequency of a laser or to generate multiple laser frequencies.

Problem 6.2

Consider the vee atom, where steady-state quantum beating can be observed, where both excited states are coupled by a single, monochromatic, electric field.



(a) Write down expressions for the free atomic Hamiltonian $\tilde{H}_{\rm A}$ and the interaction Hamiltonian $\tilde{H}_{\rm AF}$ in the rotating frame, in terms of the appropriate detunings and Rabi frequencies.

(b) Assuming a master equation of the form

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A}+\tilde{H}_{\rm AF},\tilde{\rho}\right] + \Gamma_1\mathcal{D}[\sigma_1]\rho + \Gamma_2\mathcal{D}[\sigma_2]\tilde{\rho}, \tag{6.98}$$

appropriate for distinguishable decay channels, write out the Bloch equations for the density-matrix elements $\tilde{\rho}_{\alpha\beta}$.

(c) In the case of indistinguishable decay channels, with master equation of the form

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_{\rm A} + \tilde{H}_{\rm AF}, \tilde{\rho}\right] + \mathcal{D}\left[\sqrt{\Gamma_1}\sigma_1 + \sqrt{\Gamma_2}\sigma_2\right]\tilde{\rho}, \tag{6.99}$$

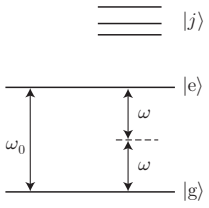
what are the new terms in the Bloch equations compared to what you wrote in part (b)? Give specific interpretations to these extra terms where possible.

Problem 6.3

For the vee atom in Problem 6.2, consider the case of indistinguishable decay channels, with $\Gamma_1 = \Gamma_2 = \Gamma$ and $\Omega_1 = \Omega_2$, with the field tuned exactly halfway between the excited states. Solve for the steady state of the optical Bloch equations for this system to lowest nontrivial order in Ω_1 and Ω_2 (i.e., find the linear response for very weak fields), and thereby prove that steady-state quantum beats occur. The effect does not depend on the smallness of the fields, and the analytic solution can be worked out for arbitrary parameters, but this problem is *much* easier in the perturbative limit.

Problem 6.4

Consider an atomic transition between states $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, of resonance frequency ω_0, driven at nearly *half* the resonance frequency, so that $2\omega \approx \omega_0$. In this case, it is possible to have **two-photon absorption** and nonlinearly drive the transition.



Of course, this happens because of Raman-type transitions involving the other states $|j\rangle$ as intermediate states. However, our stimulated-Raman analysis does *not* apply here, because we cannot make the usual rotating wave approximation, since ω does not resonantly couple $|\mathrm{g}\rangle$ or $|\mathrm{e}\rangle$ to any intermediate level.

Your goal is to work out the theory of two-photon transitions, and thus to show that this system effectively reduces to a two-level system for $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$. To do this, use the following outline.

1. Write down the free atomic Hamiltonian, using the following definitions: the energy of $|\mathrm{g}\rangle$ is zero, and the energy of the $|\mathrm{g}\rangle \longrightarrow |j\rangle$ is ω_j.
2. Write down the atom–field interaction Hamiltonian, using Rabi frequencies $\Omega_{\mathrm{g}j}$ for the $|\mathrm{g}\rangle \longrightarrow |j\rangle$ transitions and $\Omega_{\mathrm{e}j}$ for the $|\mathrm{e}\rangle \longrightarrow |j\rangle$ transitions. For the moment, ignore the direct coupling between $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ (assume, for example, that the transition is dipole-forbidden). *Do not* make any rotating-wave approximations at this stage.
3. Write the state vector as
$$|\psi\rangle = c_{\mathrm{g}}|\mathrm{g}\rangle + c_{\mathrm{e}}|\mathrm{e}\rangle + \sum_j c_j|j\rangle, \tag{6.100}$$
and derive equations of motion for the coefficients.
4. Transform into a rotating frame by changing to the slowly varying coefficient
$$\tilde{c}_{\mathrm{e}} = c_{\mathrm{e}}e^{i2\omega t}, \tag{6.101}$$
which is appropriate for $\omega_0 \approx 2\omega$, and rewrite the coefficient equations in terms of this new variable.
5. Integrate the equation for $\partial_t c_j$ to obtain an approximate expression for $c_j(t)$, assuming that c_{g} and $\tilde{c}_{\mathrm{e}}$ are slowly varying on the time scales of optical oscillations. This is justified since we are interested in the slow dynamics of these variables.

6. Use your approximate result to eliminate c_j from the equations of motion, and write the equations of motion in the form of a two-level system. *Now* you should make appropriate rotating-wave approximations to put the equations in the proper form.

In your answer, give expressions for the Stark shifts of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$, and also for the two-photon Rabi frequency. At the same level of approximation, how do your results change if the transition $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}\rangle$ is also coupled *directly* by the field with Rabi frequency Ω?

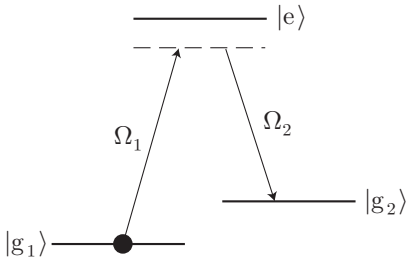
Problem 6.5

Name as many approximations as you can that go into the result

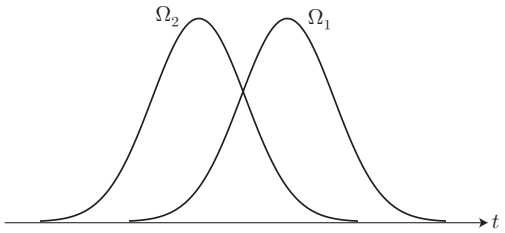
$$\Omega_{\mathrm{R}} = \frac{\Omega_1 \Omega_2}{2\Delta}. \tag{6.102}$$

Problem 6.6

Consider a transition between the two ground states of a Λ atom via the STIRAP procedure.



Estimate the probability that a photon is scattered from the atom during the transition, which happens if the pulse sequence does not drive the atom adiabatically through the transition. To do this, set this problem up as an avoided-crossing problem, and use Landau–Zener theory to estimate the probability that the atom "tunnels" out of the dark state. Model the two laser pulses as Gaussian pulses,



with pulse profile

$$\Omega(t) = \Omega_0 \exp\left[-\frac{t^2}{2\,\delta t^2}\right]. \tag{6.103}$$

Assume that the two laser pulses are identical, both exactly resonant with their respective transitions, and that the peak Rabi frequencies are the same for both transitions. Take the time separation between the two pulse peaks to be τ.

Note: strictly speaking, Landau–Zener theory only applies to an avoided crossing where the bare-state energies are linear functions of time. However, you may approximately apply it to *any* avoided crossing by noting that most of the tunneling across the gap occurs during the times when the gap is narrowest. Thus, you should set up the problem such that the Landau–Zener problem approximates the energy levels of this problem in the vicinity of the avoided crossing.

Problem 6.7

Consider two two-level atoms situated near each other. No optical fields are present other than those radiated by the atoms. Recall that the spatial profile of the atomic radiation field is given by the classical dipole pattern.

(a) Consider the interaction of the two atoms as the usual dipole interaction of one atom with the radiated field of the other. Show that after making a suitable rotating-wave approximation the atom–atom interaction can be written

$$H_{\rm int} = \hbar \left(\Xi^* \sigma_1 \sigma_2^\dagger + \Xi \sigma_1^\dagger \sigma_2 \right), \tag{6.104}$$

where $\sigma_{1,2}$ are the lowering operators for atoms 1 and 2, and the free-atom Hamiltonian is

$$H_0 = \hbar\omega_0 \left(\sigma_1^\dagger \sigma_1 + \sigma_2^\dagger \sigma_2 \right). \tag{6.105}$$

Write down an expression for the coupling rate Ξ, which depends on the separation and orientation of the atoms. Assume the two atomic dipoles have the same orientation.

(b) Argue that in a suitable rotating frame, the interaction Hamiltonian is unchanged but the free-evolution Hamiltonian becomes $\tilde{H}_0 = 0$.

(c) Now consider the *symmetric* initial state

$$|\psi_+\rangle = \frac{1}{\sqrt{2}} \Big[|{\rm e}_1, {\rm g}_2\rangle + |{\rm g}_1, {\rm e}_2\rangle \Big], \tag{6.106}$$

and the *antisymmetric* initial state

$$|\psi_-\rangle = \frac{1}{\sqrt{2}} \Big[|{\rm e}_1, {\rm g}_2\rangle - |{\rm g}_1, {\rm e}_2\rangle \Big], \tag{6.107}$$

both corresponding to a single excitation. Show that in the limit of small atom separations, where $\Xi \in \mathbb{R}$, both states are eigenstates of the rotating-frame Hamiltonian $\tilde{H}_{\rm int}$.

(d) Assume that the atoms evolve according to a master equation of the form

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar} [\tilde{H}_{\rm int}, \tilde{\rho}] + \Gamma \mathcal{D}[\sigma_1 + \sigma_2] \tilde{\rho}. \tag{6.108}$$

That is, we assume the dipole fields radiated by the two atoms to interfere perfectly, which is only true if the atoms have the same orientation, and they are very close together. Show that the decay of the atom pair starting in the *symmetric* state proceeds *more quickly* than for a single, isolated atom in the excited state. (You need only show this to be true at short times.) This effect is called **Dicke superradiance**,[30] and arises physically due to the constructive interference of the two radiated fields.

(e) Show also that the decay of the atom pair starting in the *antisymmetric* state proceeds *more slowly* than for a single, isolated atom in the excited state. This effect is called **subradiance**, and is due to the *destructive* interference of the two radiated fields.

(f) Why is the description ''two atoms playing photon ping-pong'' appropriate to this problem, specifically to the form of $H_{\rm int}$?

[30] R. H. Dicke, ''Coherence in Spontaneous Radiation Processes,'' *Physical Review* **93**, 99 (1954) (doi: 10.1103/PhysRev.93.99).

Chapter 7

Atomic Angular-Momentum Structure

In this chapter, we will review and develop some of the formalism for handling angular momentum, in particular as it applies to the structure of simple (hydrogen-like) atoms. We will use these results to look at fine and hyperfine structure, and in particular how to handle light–matter interactions in the presence of Zeeman-degenerate states (degenerate angular-momentum sublevels).

7.1 Angular Momentum

7.1.1 Operators and Eigenstates

The basics of the quantization of angular momentum is covered well enough in most introductory quantum-mechanics text, so we'll just review them here so we can get on to applications of the theory to atoms and quantum optics. First, we will suppose that we have a set of operators J_x, J_y, and J_z, and we will take them to be defined by the commutation relation

$$[J_\alpha, J_\beta] = i\hbar\epsilon_{\alpha\beta\gamma}J_\gamma, \tag{7.1}$$
(angular-momentum commutator)

where $\epsilon_{\alpha\beta\gamma}$ is the **Levi-Civita symbol** (completely antisymmetric "tensor," though not technically a tensor because it doesn't transform correctly under rotations), having the values $+1$ if $(\alpha\beta\gamma)$ is a cyclic permutation of (xyz), -1 if an odd permutation, and 0 otherwise. These operators will obviously represent angular momenta associated with the three Cartesian axes, and so it will also be useful to define an operator associated with the *total* angular momentum:

$$J^2 = J_x^2 + J_y^2 + J_z^2. \tag{7.2}$$

We assume these operators to correspond to observables, and are thus Hermitian. Out of the set of operators $\{J^2, J_x, J_y, J_z\}$, the above relations (7.1) and (7.2) show that the full set can be expressed in terms of only two (e.g., J_x and J_y). Thus, to completely span the space of angular momentum states, we can choose to have simultaneous eigenstates of any two such operators. This strategy is useful in spherically symmetric systems, where such eigenstates should exist, and so any component J_α is as good as any other. However, in view of the fundamental commutation relation, we can't have simultaneous eigenstates for J_α and J_β if $\alpha \neq \beta$. However, J^2 commutes with J_α:

$$[J_\alpha, J^2] = 0. \tag{7.3}$$
(angular-momentum commutator)

To see this, we can take $J_\alpha = J_x$ without loss of generality, in which case

$$\begin{aligned}
[J_x, J^2] &= [J_x, J_y^2] + [J_x, J_z^2] \\
&= [J_x, J_y]J_y + J_y[J_x, J_y] + [J_x, J_z]J_z + J_z[J_x, J_z] \\
&= (i\hbar J_z)J_y + J_y(i\hbar J_z) + (-i\hbar J_y)J_z + J_z(-i\hbar J_y) \\
&= 0.
\end{aligned} \tag{7.4}$$

Thus, we are free to construct simultaneous eigenstates of J^2 and J_α. We make the arbitrary but conventional choice of taking simultaneous eigenstates of J^2 and J_z. We will thus need two quantum numbers, which we call j and m, and define the eigenvalue λ_j of J^2 to be some function of j,

$$J^2|j\ m\rangle = \lambda_j|j\ m\rangle, \tag{7.5}$$

and the eigenvalue λ_m of J_z will similarly be some function of m,

$$J_z|j\ m\rangle = \lambda_m|j\ m\rangle. \tag{7.6}$$

Our goal in setting up the basic formalism will now be to work out the angular-momentum eigenvalues.

7.1.2 Ladder Operators and Eigenvalues

It is also useful to define two non-Hermitian operators, the **ladder operators**

$$J_\pm := J_x \pm iJ_y, \tag{7.7}$$

which will turn out to be somewhat more convenient than J_x and J_y separately. Given the commutation relation (7.3), we immediately see that the ladder operators commute with J^2:

$$[J^2, J_\pm] = 0. \tag{7.8}$$
(ladder-operator commutator)

The commutators with J_z is not hard to work out,

$$[J_z, J_\pm] = [J_z, J_x] \pm i[J_z, J_y] = i\hbar J_y \pm \hbar J_x, \tag{7.9}$$

or

$$[J_z, J_\pm] = \pm\hbar J_\pm. \tag{7.10}$$
(ladder-operator commutator)

We can also readily compute the commutator of the two ladder operators as

$$[J_+, J_-] = -2i[J_x, J_y] = 2\hbar J_z. \tag{7.11}$$

Now, to put the ladder operators to use, we can consider the action of $J_\pm$ on an eigenstate state $|j\ m\rangle$. In particular, notice that since the J_α commute with J^2, they transform $|j\ m\rangle$ to a state $J_\alpha|j\ m\rangle$ such that

$$J^2(J_\alpha|j\ m\rangle) = J_\alpha J^2|j\ m\rangle = \lambda_j(J_\alpha|j\ m\rangle). \tag{7.12}$$

Thus, $J_\alpha|j\ m\rangle$ is an eigenstate of J^2 with the same eigenvalue as $|j\ m\rangle$, implying that j is unchanged. The same conclusion of course holds for $J_\pm$, and thus, since we will be considering the action of $J_{x,y,z,\pm}$ on the states $|j, m\rangle$ for the rest of this section, we can regard j as a fixed quantity.

We can then use the commutator (7.10) on $|j\ m\rangle$ to write

$$J_zJ_\pm|j\ m\rangle = J_\pm J_z|j\ m\rangle \pm \hbar J_\pm|j\ m\rangle = (\lambda_m \pm \hbar)J_\pm|j\ m\rangle. \tag{7.13}$$

This shows that $J_\pm|j\ m\rangle$ is an eigenstate of J_z with eigenvalue $\lambda_m \pm \hbar$. Now we see the reason for the name "ladder operators," since J_+ acts to raise λ_m by $\hbar$, and J_- lowers it by the same amount. Now since m is an arbitrary label for the states, we may define it such that $\lambda_m = m\hbar$. That is, m represents the projection of angular momentum along the z-axis in multiples of $\hbar$. Then we may write

$$J_z|j\ m\rangle = m\hbar|j\ m\rangle \tag{7.14}$$
(J_z eigenvalues)

for the J_z eigenvalue equation, and for the ladder operators we have thus shown that

$$J_\pm|j\ m\rangle \propto |j\ m \pm 1\rangle. \tag{7.15}$$

To establish the proper normalization, we note that

$$J_\mp J_\pm = J_x^2 + J_y^2 \pm i[J_x, J_y] = J^2 - J_z^2 \mp \hbar J_z = J^2 - J_z(J_z \pm \hbar), \tag{7.16}$$

and thus the norm of the raised/lowered state is

$$\langle j\ m|J_\mp J_\pm|j\ m\rangle = \langle j\ m|[J^2 - J_z(J_z \pm \hbar)]|j\ m\rangle = \lambda_j - m(m \pm 1)\hbar^2. \tag{7.17}$$

Note that the right-hand side becomes negative for sufficiently large m, assuming λ_j to be fixed. However, since $\langle j\ m|J_\mp J_\pm|j\ m\rangle \geq 0$, we can conclude that there is a maximum value of m, say $m_{\max}$, such that

$$J_+|j\ m_{\max}\rangle = 0. \tag{7.18}$$

Then applying Eq. (7.16) to $|j\ m_{\max}\rangle$,

$$J_- J_+|j\ m_{\max}\rangle = [J^2 - J_z(J_z + \hbar)]|j\ m_{\max}\rangle. \tag{7.19}$$

The left-hand side vanishes, so

$$J^2|j\ m_{\max}\rangle = J_z(J_z + \hbar)|j\ m_{\max}\rangle = m_{\max}(m_{\max} + 1)\hbar^2|j\ m_{\max}\rangle. \tag{7.20}$$

Since j is likewise an arbitrary label for the eigenvalue $\lambda_j = m_{\max}(m_{\max}+1)\hbar^2$, we may thus define $j := m_{\max}$, with $j \geq 0$, so that

$$J^2|j\ m\rangle = j(j+1)\hbar^2|j\ m\rangle. \tag{7.21}$$
(J^2 eigenvalue equation)

Repeating this argument, Eq. (7.17) implies a smallest (negative) value of m, say $m_{\min}$, so that

$$J_-|j\ m_{\min}\rangle = 0. \tag{7.22}$$

Again applying Eq. (7.16) to $|j\ m_{\min}\rangle$,

$$J_+ J_-|j\ m_{\min}\rangle = [J^2 - J_z(J_z - \hbar)]|j\ m_{\min}\rangle, \tag{7.23}$$

and since the left-hand side vanishes,

$$J^2|j\ m_{\min}\rangle = J_z(J_z - \hbar)|j\ m_{\min}\rangle = m_{\min}(m_{\min} - 1)\hbar^2|j\ m_{\min}\rangle. \tag{7.24}$$

Thus,

$$j(j+1) = m_{\min}(m_{\min} - 1), \tag{7.25}$$

which is satisfied by $m_{\min} = -j$ (the alternate solution, $m_{\min} = j + 1$, violates the definition $j = m_{\max}$). Thus, m is constrained to be within a bounded range,

$$-j \leq m \leq j. \tag{7.26}$$
(range constraint of m)

Recall that this followed from Eq. (7.17), to avoid a contradiction with the requirement $\langle j\ m|J_\mp J_\pm|j\ m\rangle \geq 0$. In particular, if we start with the state $|j\ {-j}\rangle$ and repeatedly apply J_+, we should eventually end up with (something proportional to) the $|j\ {+j}\rangle$ state. How do we know this? Referring again to Eq. (7.17), which we may rewrite now as

$$\langle j\ m|J_\mp J_\pm|j\ m\rangle = \Big[j(j+1) - m(m \pm 1)\Big]\hbar^2. \tag{7.27}$$

we see that the *only* state that vanishes when hit by J_+ is $|j\ (+j)\rangle$. Thus, the only way to avoid a contradiction (negative state norm) is for $J_+^n|j\ (-j)\rangle \propto |j\ (+j)\rangle$ for some integer n. Further, we may conclude that *every* state $|j\ m\rangle$ may be written as $J_+^n|j\ (-j)\rangle$ (up to a scalar factor) for some integer n; otherwise we would have

a state that, when raised arbitrarily many times by J_+, would not vanish. Thus, we may conclude that m takes on discrete, integer-separated values, according to

$$m \in \{-j, -j+1, \ldots, j-1, j\} \qquad (2j+1 \text{ possible values}), \tag{7.28}$$
(range constraint of m)

which means that there are $2j+1$ possible values for m (i.e., because $m+j$ ranges from 0 to $2j$). Furthermore, $2j+1$ must be an integer, because $2j$ is the number of times J_+ must be applied to $|j\ (-j)\rangle$ to obtain $|j\ (+j)\rangle$. This implies that

$$j \in \mathbb{Z} \quad \text{or} \quad j + \frac{1}{2} \in \mathbb{Z}, \tag{7.29}$$
(integer/half-integer constraint)

with j also nonnegative, as we discussed before. That is, j is either an integer or a half-integer. As we will discuss later, only integer j can correspond to coordinate-space angular momenta; half-integer j are restricted to representing intrinsic particle *spin* angular momenta (which can also have integer j).

Finally, just to tidy up loose ends, we can use Eq. (7.27) to write down

$$\begin{aligned} J_\pm|j\ m\rangle &= \hbar\sqrt{j(j+1) - m(m\pm1)}\ |j\ m\pm1\rangle \\ &= \hbar\sqrt{(j\pm m+1)(j\mp m)}\ |j\ m\pm1\rangle \end{aligned} \tag{7.30}$$
(ladder operator effects)

as the properly normalized action of the ladder operators on the angular-momentum eigenstates.

7.1.3 Addition of Two Angular Momenta: Clebsch–Gordan Coefficients

7.1.3.1 Basis States

Suppose we have two angular momenta $\mathbf{J}_1$ and $\mathbf{J}_2$, and we want to consider their sum $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$. We assume these angular momenta to correspond to independent degrees of freedom, and thus they commute:

$$[J_{1\alpha}, J_{2\beta}] = 0 \qquad \forall_{\alpha,\beta}. \tag{7.31}$$

Treating the two angular momenta as separate entities, we can construct simultaneous eigenstates of J_1^2, J_2^2, J_{1z}, and J_{2z}, since everybody here commutes. We will denote these eigenstates by $|j_1\ m_1; j_2\ m_2\rangle \equiv |j_1\ m_1\rangle|j_2\ m_2\rangle$, so that

$$\begin{aligned} J_1^2|j_1\ m_1; j_2\ m_2\rangle &= j_1(j_1+1)\hbar^2|j_1\ m_1; j_2\ m_2\rangle \\ J_2^2|j_1\ m_1; j_2\ m_2\rangle &= j_2(j_2+1)\hbar^2|j_1\ m_1; j_2\ m_2\rangle \\ J_{1z}|j_1\ m_1; j_2\ m_2\rangle &= m_1\hbar|j_1\ m_1; j_2\ m_2\rangle \\ J_{2z}|j_1\ m_1; j_2\ m_2\rangle &= m_2\hbar|j_1\ m_1; j_2\ m_2\rangle. \end{aligned} \tag{7.32}$$

Now note that the total angular momentum $\mathbf{J}$ has the characteristics of an angular momentum operator, since

$$[J_\alpha, J_\beta] = [J_{1\alpha} + J_{2\alpha}, J_{1\beta} + J_{2\beta}] = [J_{1\alpha}, J_{1\beta}] + [J_{2\alpha}, J_{2\beta}] = i\hbar\epsilon_{\alpha\beta\gamma}J_{1\gamma} + i\hbar\epsilon_{\alpha\beta\gamma}J_{2\gamma} = i\hbar\epsilon_{\alpha\beta\gamma}J_\gamma. \tag{7.33}$$

Thus, we may have simultaneous eigenstates of J^2 and J_z. Also, it is easy to see that J_1^2 and J_2^2 both commute with J^2 and J_z (but J_{1z} and J_{2z} don't commute with J^2), so that we can represent our states in terms of simultaneous eigenstates of J_1^2, J_2^2, J^2, and J_z, which we will label by $|j_1, j_2; j\ m\rangle$, so that

$$\begin{aligned} J_1^2|j_1, j_2; j\ m\rangle &= j_1(j_1+1)\hbar^2|j_1, j_2; j\ m\rangle \\ J_2^2|j_1, j_2; j\ m\rangle &= j_2(j_2+1)\hbar^2|j_1, j_2; j\ m\rangle \\ J^2|j_1, j_2; j\ m\rangle &= j(j+1)\hbar^2|j_1, j_2; j\ m\rangle \\ J_z|j_1, j_2; j\ m\rangle &= m\hbar|j_1, j_2; j\ m\rangle. \end{aligned} \tag{7.34}$$

Sometimes, the state $|j_1, j_2; j\ m\rangle$ is written more succinctly as $|j, m\rangle$ if j_1 and j_2 are clear from the context.

7.1.3.2 Transformation between Bases and Clebsch–Gordan Coefficients

Now we have two distinct bases by which to represent a general state. The basic problem of angular-momentum addition is thus to express any basis state in terms of a superposition of states from the other basis. This is easy to do by using the representations of the identity in each basis:

$$\begin{aligned} |j_1, j_2; j\ m\rangle &= \sum_{j_1' j_2' m_1 m_2} |j_1'\ m_1; j_2'\ m_2\rangle\langle j_1'\ m_1; j_2'\ m_2|j_1, j_2; j\ m\rangle \\ |j_1\ m_1; j_2\ m_2\rangle &= \sum_{j_1' j_2' j m} |j_1', j_2'; j\ m\rangle\langle j_1', j_2'; j\ m|j_1\ m_1; j_2\ m_2\rangle. \end{aligned} \tag{7.35}$$

The inner products on the right-hand sides of the above equations are **Clebsch–Gordan coefficients**. Note that J_1^2 and J_2^2 are Hermitian, and thus

$$\begin{aligned} \langle j_1', j_2'; j\ m|J_1^2|j_1\ m_1; j_2\ m_2\rangle &= j_1'(j_1'+1)\hbar^2\langle j_1', j_2'; j\ m|j_1\ m_1; j_2\ m_2\rangle \\ &= j_1(j_1+1)\hbar^2\langle j_1', j_2'; j\ m|j_1\ m_1; j_2\ m_2\rangle, \end{aligned} \tag{7.36}$$

so that the Clebsch–Gordan coefficient vanishes unless $j_1 = j_1'$ (and similarly $j_2 = j_2'$). Additionally, $J_z = J_{1z} + J_{2z}$, so

$$\begin{aligned} \langle j_1, j_2; j\ m|J_z|j_1\ m_1; j_2\ m_2\rangle &= (m_1+m_2)\hbar\langle j_1, j_2; j\ m|j_1\ m_1; j_2\ m_2\rangle \\ &= m\hbar\langle j_1, j_2; j\ m|j_1\ m_1; j_2\ m_2\rangle, \end{aligned} \tag{7.37}$$

and thus we must have

$$m = m_1 + m_2 \tag{7.38}$$
(angular-momentum conservation)

for the Clebsch–Gordan coefficient to be nonvanishing. Thus, we may rewrite the transformation relations (7.35) as

$$\begin{aligned} |j_1, j_2; j\ m\rangle &= \sum_{\substack{m_1 m_2 \\ (m_1+m_2=m)}} |j_1\ m_1; j_2\ m_2\rangle\langle j_1\ m_1; j_2\ m_2|j_1, j_2; j\ m\rangle \\ |j_1\ m_1; j_2\ m_2\rangle &= \sum_{\substack{jm \\ (m_1+m_2=m)}} |j_1, j_2; j\ m\rangle\langle j_1, j_2; j\ m|j_1\ m_1; j_2\ m_2\rangle, \end{aligned} \tag{7.39}$$

or omitting the redundant labels,

$$\begin{aligned} |j\ m\rangle &= \sum_{\substack{m_1 m_2 \\ (m_1+m_2=m)}} |j_1\ m_1; j_2\ m_2\rangle\langle j_1\ m_1; j_2\ m_2|j\ m\rangle \\ |j_1\ m_1; j_2\ m_2\rangle &= \sum_{\substack{jm \\ (m_1+m_2=m)}} |j\ m\rangle\langle j\ m|j_1\ m_1; j_2\ m_2\rangle, \end{aligned} \tag{7.40}$$
(transformation rules)

The other important constraint is

$$|j_1 - j_2| \le j \le j_1 + j_2 \tag{7.41}$$
(triangular condition)

(recall that $j_1, j_2 \ge 0$). To see this, first note that since $m = m_1 + m_2$, and the maximum value of m is the maximum value of j, but is also given by $j_1 + j_2$. Thus, $j_{\max} = j_1 + j_2$. To find the minimum value of j, note that in the $|j_1\ m_1; j_2\ m_2\rangle$ basis, there are $2j_1 + 1$ states associated with the the $|j_1\ m_1\rangle$ space and $2j_2 + 1$ states associated with the the $|j_2\ m_2\rangle$ space, and thus the composite space is spanned by $(2j_1 + 1)(2j_2 + 1)$

states. In the other basis, we get the correct number of states if $j_{\min} = |j_1 - j_2|$. That is, $j_{\min}$ is the solution to

$$\sum_{j=j_{\min}}^{j_1+j_2} (2j+1) = (2j_1+1)(2j_2+1), \tag{7.42}$$

which we can see because, assuming without loss of generality that $j_1 \geq j_2$,

$$\sum_{j=j_1-j_2}^{j_1+j_2} (2j+1) = \sum_{j=-j_2}^{j_2} [2(j_1+j)+1] = \sum_{j=-j_2}^{j_2} (2j_1+1) + \sum_{j=-j_2}^{j_2} 2j = (2j_1+1)(2j_2+1). \tag{7.43}$$

The cases $j = |j_1 - j_2|$ and $j = j_1 + j_2$ clearly correspond to antialigned and aligned constituent momentum vectors, respectively.

The Clebsch–Gordan coefficients obey orthogonality relations as follows. From the second transformation rule in Eqs. (7.40),

$$\langle j_1\ m_1'; j_2\ m_2' | j_1\ m_1; j_2\ m_2 \rangle = \sum_{jm} \langle j_1\ m_1'; j_2\ m_2' | j\ m \rangle \langle j\ m | j_1\ m_1; j_2\ m_2 \rangle, \tag{7.44}$$

The left-hand side is zero unless $m_1' = m_1$ and $m_2' = m_2$, so

$$\sum_{jm} \langle j_1\ m_1'; j_2\ m_2' | j\ m \rangle \langle j\ m | j_1\ m_1; j_2\ m_2 \rangle = \delta_{m_1 m_1'} \delta_{m_2 m_2'}.$$

(Clebsch–Gordan orthogonality relation) (7.45)

Similarly, the other transformation rule leads to

$$\sum_{m_1 m_2} \langle j\ m | j_1\ m_1; j_2\ m_2 \rangle \langle j_1\ m_1; j_2\ m_2 | j'\ m' \rangle = \delta_{jj'} \delta_{mm'},$$

(Clebsch–Gordan orthogonality relation) (7.46)

which are reasonably obvious applications of different representations of the identity operator.

7.1.3.3 Calculation of Clebsch–Gordan Coefficients

To determine the Clebsch–Gordan coefficients, we make use of the raising and lowering operators

$$\begin{aligned} J_{1\pm} &= J_{1x} \pm i J_{1y} \\ J_{2\pm} &= J_{2x} \pm i J_{2y} \\ J_\pm &:= J_{1\pm} + J_{2\pm}. \end{aligned} \tag{7.47}$$

Then by writing out $\langle j_1\ m_1; j_2\ m_2 | J_\pm | j\ m \rangle = \langle j_1\ m_1; j_2\ m_2 | (J_{1\pm} + J_{2\pm}) | j\ m \rangle$, we find

$$\begin{aligned} \sqrt{j(j+1) - m(m\pm1)}\, \langle j_1\ m_1; j_2\ m_2 | j\ m\pm1 \rangle = {} & \sqrt{j_1(j_1+1) - m_1(m_1\mp1)}\, \langle j_1\ m_1\mp1; j_2\ m_2 | j\ m \rangle \\ & + \sqrt{j_2(j_2+1) - m_2(m_2\mp1)}\, \langle j_1\ m_1; j_2\ m_2\mp1 | j\ m \rangle \end{aligned}$$

(Clebsch–Gordan recursion relation) (7.48)

This recursion relation, in addition to some initial conditions, is sufficient to compute the coefficients. The basic idea is as follows: Setting $m = j$ and taking the upper sign option in Eq. (7.48) gives

$$\begin{aligned} & \sqrt{j_1(j_1+1) - m_1(m_1-1)}\, \langle j_1\ m_1-1; j_2\ m_2 | j\ j \rangle \\ & \qquad + \sqrt{j_2(j_2+1) - m_2(m_2-1)}\, \langle j_1\ m_1; j_2\ m_2-1 | j\ j \rangle = 0, \end{aligned} \tag{7.49}$$

which together with the special cases of Eq. (7.46)

$$\sum_{\substack{m_1 m_2 \\ (m_1+m_2=j)}} |\langle j_1\ m_1; j_2\ m_2 | j\ j \rangle|^2 = 1 \tag{7.50}$$

that pin down the normalization, all coefficients of the form $\langle j_1\ m_1 - 1; j_2\ m_2 | j\ j\rangle$ can be determined up to an arbitrary phase. It is conventional to take the coefficients $\langle j_1\ j_1; j_2\ j - j_1 | j\ j\rangle$ to be real and positive.[1] The recursion relation (7.49) can then generate all the rest of the Clebsch–Gordon coefficients from these "basis cases," and an important consequence of the recursion relation (which only involves real recursion coefficients) and the phase convention is that by convention **all Clebsch–Gordon coefficients are real** (though not necessarily positive).

7.1.3.4 Explicit Formula

The above recursion procedure is fairly cumbersome, although sometimes useful. In a numerical calculation, it is convenient to have explicit formulae to implement any of the coupling coefficients. Fortunately, the Clebsch–Gordan coefficient may be computed according to the rather complicated formula[2]

$$\begin{aligned}
&\langle j_1, m_1; j_2, m_2 | j_3, m_3\rangle \\
&\quad = \delta_{(m_1+m_2),m_3} \sqrt{\frac{(j_1+j_2-j_3)!(j_1+j_3-j_2)!(j_2+j_3-j_1)!}{(j_1+j_2+j_3+1)!}} \\
&\qquad \times \sqrt{(2j_3+1)(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!(j_3+m_3)!(j_3-m_3)!} \\
&\qquad \times \sum_{n=n_{\min}}^{n_{\max}} \frac{(-1)^n}{(j_1-m_1-n)!(j_3-j_2+m_1+n)!(j_2+m_2-n)!(j_3-j_1-m_2+n)!n!(j_1+j_2-j_3-n)!},
\end{aligned}$$

(Clebsch–Gordan coefficient: explicit formula) (7.51)

where the summation limits

$$\begin{aligned}
n_{\min} &= \max\{j_2 - j_3 - m_1, j_1 + m_2 - j_3, 0\} \\
n_{\max} &= \min\{j_1 - m_1, j_2 + m_2, j_1 + j_2 - j_3\}
\end{aligned} \tag{7.52}$$

are chosen such that no factorial arguments are negative. For a nonzero result, we reiterate that we must have $m_1 + m_2 = m_3$, $|j_1 - j_2| \le j_3 \le |j_1 + j_2|$, $j_\alpha \ge 0$, and $|m_\alpha| \le j_\alpha$.

7.1.3.5 Symmetry Relations and Wigner 3-*j* Symbols

Now that we can compute the Clebsch–Gordan coefficients, we can ask, what are the shortcuts to relating them if we just want to permute some symbols? For example, recall that the coupling of two angular momenta according to

$$\mathbf{J}_1 + \mathbf{J}_2 = \mathbf{J}_3 \tag{7.53}$$

is represented by the coefficient $\langle j_1\ m_1; j_2\ m_2 | j_3\ m_3\rangle$. However, $\mathbf{J}_1$ and $\mathbf{J}_2$ are on equal footing in being added together to form $\mathbf{J}_3$, and so we should be able to switch them without a problem, at least up to an overall phase. It turns out that according to the sign convention we have chosen,

$$\langle j_1\ m_1; j_2\ m_2 | j_3\ m_3\rangle = (-1)^{j_1+j_2-j_3} \langle j_2\ m_2; j_1\ m_1 | j_3\ m_3\rangle. \tag{7.54}$$

(symmetry relation)

We can see this by redefining the index n in the explicit formula (7.51) according to $n \longrightarrow (j_1 + j_2 - j_3) - n$, with the limits redefined appropriately to avoid any negative factorials, together with the simultaneous exchanges $j_1 \longleftrightarrow j_2$ and $m_1 \longleftrightarrow m_2$. This transformation leaves the sum invariant, except for the sign $(-1)^{j_1+j_2-j_3}$ (the same exchanges leave the prefactor of the sum invariant as well).

We can go even farther than this. The addition (7.53) is clearly equivalent to the addition

$$\mathbf{J}_3 - \mathbf{J}_1 = \mathbf{J}_2, \tag{7.55}$$

and thus we expect

$$\langle j_1\ -m_1; j_3\ m_3 | j_2\ m_2\rangle \propto \langle j_1\ m_1; j_2\ m_2 | j_3\ m_3\rangle. \tag{7.56}$$

[1] This is known as the **Condon–Shortley phase convention**. See, e.g., D. M. Brink and G. R. Satchler, *Angular Momentum*, 2nd ed. (Oxford, 1968), Section 2.7.2, p. 33.

[2] D. M. Brink and G. R. Satchler, op. cit., p. 34, Eq. (2.34).

In fact, we can see that this is the case by noting that the recursion relation (7.48)

$$\begin{aligned}\sqrt{j_3(j_3+1)-m_3(m_3\pm1)}\,\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\pm1\rangle \\ = \sqrt{j_1(j_1+1)-m_1(m_1\mp1)}\,\langle j_1\ m_1\mp1;j_2\ m_2|j_3\ m_3\rangle \\ + \sqrt{j_2(j_2+1)-m_2(m_2\mp1)}\,\langle j_1\ m_1;j_2\ m_2\mp1|j_3\ m_3\rangle,\end{aligned} \tag{7.57}$$

upon the substitutions $|j_3\,m_3\rangle \longrightarrow |j_2\ m_2\rangle$, $|j_2\,m_2\rangle \longrightarrow |j_1\ -m_1\rangle$, $|j_1\,m_1\rangle \longrightarrow |j_3\ m_3\rangle$, and $\pm \longleftrightarrow \mp$, and multiplying through by $(-1)^{-m_1}$, becomes

$$\begin{aligned}(-1)^{-m_1}\sqrt{j_3(j_3+1)-m_3(m_3\pm1)}\,\langle j_3\ m_3\pm1;j_1\ -m_1|j_2\ m_2\rangle \\ = (-1)^{-(m_1\mp1)}\sqrt{j_1(j_1+1)-(-m_1)[-(m_1\mp1)]}\,\langle j_3\ m_3;j_1\ -(m_1\mp1)|j_2\ m_2\rangle \\ + (-1)^{-m_1}\sqrt{j_2(j_2+1)-m_2(m_2\mp1)}\,\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\mp1\rangle.\end{aligned} \tag{7.58}$$

This recursion relation has the same form as the original, and indicates that $(-1)^{-m_1}\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle$ obeys the same recursion relation as $\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle$. Since the recursion relation determines the m-dependence of the Clebsch–Gordan coefficients, we conclude that these two coefficients are proportional,

$$\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle \propto (-1)^{-m_1}\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle, \tag{7.59}$$

with the remaining proportionality constant to be determined depends only on the j's. To get the j-dependent amplitude, note that from Eq. (7.46) we may write

$$\begin{aligned}\sum_{m_1m_2}|\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle|^2 = 1 \\ \sum_{m_1m_3}|\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle|^2 = 1.\end{aligned} \tag{7.60}$$

but since $m_1+m_2=m_3$ holds in either case, the sums simplify to

$$\begin{aligned}\sum_{m_2}|\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle|^2 = 1 \\ \sum_{m_3}|\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle|^2 = 1.\end{aligned} \tag{7.61}$$

Noting that the coefficients are equivalent in each case and that we have already taken care of the m-dependence, we count $2j_2+1$ terms in the first sum and $2j_3+1$ in the second. Thus, for the sums to be equivalent, we require

$$\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle \propto (-1)^{-m_1}\sqrt{\frac{2j_3+1}{2j_2+1}}\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle \tag{7.62}$$

where the remaining proportionality constant is a j-dependent phase. This we establish by noting the convention we already mentioned that $\langle j_1\ j_1;j_2\ (j_3-j_1)|j_3\ j_3\rangle$ is always positive. In this case, we need a factor of $(-1)^{j_1}$ to cancel the factor of $(-1)^{-m_1}$ for this case, and so we can finally write the symmetry relation

$$\langle j_1\ m_1;j_2\ m_2|j_3\ m_3\rangle = (-1)^{j_1-m_1}\sqrt{\frac{2j_3+1}{2j_2+1}}\langle j_3\ m_3;j_1\ -m_1|j_2\ m_2\rangle. \tag{7.63}$$

(Clebsch–Gordan symmetry rule)

We can then take $\langle j_1\ m_1; j_2\ m_2 | j_3\ m_3\rangle$, apply Eq. (7.54), apply Eq. (7.63), and apply Eq. (7.54) again to find

$$\begin{aligned}\langle j_1\ m_1; j_2\ m_2|j_3\ m_3\rangle &= (-1)^{j_1+j_2-j_3}\langle j_2\ m_2; j_1\ m_1|j_3\ m_3\rangle \\ &= (-1)^{j_1+j_2-j_3}(-1)^{j_2-m_2}\sqrt{\frac{2j_3+1}{2j_1+1}}\langle j_3\ m_3; j_2\ -m_2|j_1\ m_1\rangle \\ &= (-1)^{j_1+j_2-j_3}(-1)^{j_2-m_2}(-1)^{j_3+j_2-j_1}\sqrt{\frac{2j_3+1}{2j_1+1}}\langle j_2\ -m_2; j_3\ m_3|j_1\ m_1\rangle \\ &= (-1)^{j_2+m_2}(-1)^{2(j_2-m_2)}\sqrt{\frac{2j_3+1}{2j_1+1}}\langle j_2\ -m_2; j_3\ m_3|j_1\ m_1\rangle,\end{aligned} \tag{7.64}$$

and noting that $j_2 - m_2$ is always an integer, so $(-1)^{2(j_2-m_2)} = 1$, we find the alternate symmetry relation

$$\langle j_1\ m_1; j_2\ m_2|j_3\ m_3\rangle = (-1)^{j_2+m_2}\sqrt{\frac{2j_3+1}{2j_1+1}}\langle j_2\ -m_2; j_3\ m_3|j_1\ m_1\rangle. \tag{7.65}$$

(Clebsch–Gordan symmetry rule)

Noting that this rule amounts to a cyclic permutation of the angular momenta while flipping the orientation of one, we can apply this rule three times to find

$$\begin{aligned}\langle j_1\ m_1; j_2\ m_2|j_3\ m_3\rangle &= (-1)^{j_2+m_2}\sqrt{\frac{2j_3+1}{2j_1+1}}\langle j_2\ -m_2; j_3\ m_3|j_1\ m_1\rangle \\ &= (-1)^{j_2+j_3+m_2+m_3}\sqrt{\frac{2j_3+1}{2j_2+1}}\langle j_3\ -m_3; j_1\ m_1|j_2\ -m_2\rangle \\ &= (-1)^{j_1+j_2+j_3+m_1+m_2+m_3}\langle j_1\ -m_1; j_2\ -m_2|j_3\ -m_3\rangle.\end{aligned} \tag{7.66}$$

Noting that $j_3 - m_3$ and $m_1 + m_2 - j_3$ are both integers, we can rewrite this as the final symmetry rule

$$\langle j_1\ m_1; j_2\ m_2|j_3\ m_3\rangle = (-1)^{j_1+j_2-j_3}\langle j_1\ -m_1; j_2\ -m_2|j_3\ -m_3\rangle. \tag{7.67}$$

(Clebsch–Gordan symmetry rule)

A nice way to summarize the symmetry relations here is to define the **Wigner 3-*j* symbol** in terms of the Clebsch–Gordan coefficient as

$$\begin{pmatrix} j_1 & j_2 & j_3 \\ m_1 & m_2 & m_3 \end{pmatrix} := \frac{(-1)^{j_1-j_2-m_3}}{\sqrt{2j_3+1}}\langle j_1, m_1; j_2, m_2|j_3, -m_3\rangle. \tag{7.68}$$

(Wigner 3-j symbol)

Then the symmetries are as follows: The symbol on the left-hand side of Eq. (7.68) is invariant under *even* permutations of the columns, but *odd* permutations are accompanied by a factor $(-1)^{j_1+j_2+j_3}$. The simultaneous, triple replacement $m_{1,2,3} \longrightarrow -m_{1,2,3}$ is similarly accompanied by the same factor. Finally, the symbol is only nonvanishing if $m_1 + m_2 + m_3 = 0$, and if j_1, j_2, and j_3 obey the usual triangle condition.

7.1.4 Addition of Three Angular Momenta: Racah Coefficients and Wigner 6-*j* Symbols

Suppose now that we want to couple *three* angular momenta, $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3$. We can use the formalism we have just developed for adding together two angular momenta, and simply iterate it. Unfortunately, the result of doing this turns out not to be unique: it depends on which two angular momenta are coupled first. For example, suppose we first add $\mathbf{J}_{12} = \mathbf{J}_1 + \mathbf{J}_2$, using Eq. (7.40):

$$|j_{12}\ m_{12}\rangle = \sum_{\substack{m_1 m_2 \\ (m_1+m_2=m_{12})}} |j_1\ m_1; j_2\ m_2\rangle\langle j_1\ m_1; j_2\ m_2|j_{12}\ m_{12}\rangle. \tag{7.69}$$

Now we can add $\mathbf{J} = \mathbf{J}_{12} + \mathbf{J}_3$ to obtain

$$|j_{12}, j_3; j\ m\rangle = \sum_{\substack{m_{12} m_3 \\ (m_{12}+m_3=m)}} |j_{12}\ m_{12}; j_3\ m_3\rangle \langle j_{12}\ m_{12}; j_3\ m_3 | j\ m\rangle. \tag{7.70}$$

Combining these two relations, we find the composite state in terms of the three original angular momenta as

$$|j_{12}, j_3; j\ m\rangle = \sum_{\substack{m_1 m_2 m_{12} m_3 \\ (m_1+m_2=m_{12}) \\ (m_{12}+m_3=m)}} |j_1\ m_1; j_2\ m_2; j_3\ m_3\rangle \langle j_{12}\ m_{12}; j_3\ m_3 | j\ m\rangle \langle j_1\ m_1; j_2\ m_2 | j_{12}\ m_{12}\rangle. \tag{7.71}$$

On the other hand, suppose we instead first added $\mathbf{J}_{23} = \mathbf{J}_2 + \mathbf{J}_3$, and then $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_{23}$, Then we instead obtain

$$\begin{aligned} |j_1, j_{23}; j\ m\rangle &= \sum_{\substack{m_1 m_{23} \\ (m_1+m_{23}=m)}} |j_1\ m_1; j_{23}\ m_{23}\rangle \langle j_1\ m_1; j_{23}\ m_{23} | j\ m\rangle \\ &= \sum_{\substack{m_1 m_2 m_3 m_{23} \\ (m_2+m_3=m_{23}) \\ (m_1+m_{23}=m)}} |j_1\ m_1; j_2\ m_2; j_3\ m_3\rangle \langle j_1\ m_1; j_{23}\ m_{23} | j\ m\rangle \langle j_2\ m_2; j_3\ m_3 | j_{23}\ m_{23}\rangle. \end{aligned} \tag{7.72}$$

This expression is clearly different from Eqs. (7.70) and (7.71), but equally valid. Since we have proceeded via the established procedure of adding two angular momenta, where the composite states form a complete basis for the uncoupled states, we know that both $|j_{12}; j_3; j\ m\rangle$ and $|j_1; j_{23}; j\ m\rangle$ form alternate, complete bases for the original space spanned by $|j_1\ m_1; j_2\ m_2; j_3\ m_3\rangle$. Thus, there exists a unitary transformation between the bases, which we can write in the same way as for Clebsch–Gordan coefficients as

$$|j_{12}, j_3; j\ m\rangle = \sum_{j_{23}} |j_1, j_{23}; j\ m\rangle \langle j_1, j_{23}; j\ m | j_{12}, j_3; j\ m\rangle. \tag{7.73}$$

(No sum over m is required, due to the orthogonality of the composite basis vectors in either addition scheme.) The coefficient is in fact independent of m. To see this, set $m = j$,

$$|j_{12}, j_3; j\ j\rangle = \sum_{j_{23}} |j_1, j_{23}; j\ j\rangle \langle j_1, j_{23}; j\ j | j_{12}, j_3; j\ j\rangle, \tag{7.74}$$

Now apply the lowering operator J_- from Eq. (7.30)

$$|j_{12}, j_3; j\ j-1\rangle = \sum_{j_{23}} |j_1, j_{23}; j\ j-1\rangle \langle j_1, j_{23}; j\ j | j_{12}, j_3; j\ j\rangle, \tag{7.75}$$

since the normalization coefficient will be the same on either side of the equation. We can continue to apply the lowering operator to obtain the relation between states of any m, which are transformed with the same coefficient as for $m = j$; thus, the coefficient is m-independent. Finally, we may write this ''recoupling equation'' as

$$\begin{aligned} |j_{12}, j_3; j\ m\rangle &= \sum_{j_{23}} |j_1, j_{23}; j\ m\rangle \sqrt{(2j_{12}+1)(2j_{23}+1)}\, W(j_1\ j_2\ j\ j_3; j_{12}\ j_{23}) \\ &= \sum_{j_{23}} |j_1, j_{23}; j\ m\rangle (-1)^{j_1+j_2+j_3+j} \sqrt{(2j_{12}+1)(2j_{23}+1)} \begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_3 & j & j_{23} \end{Bmatrix}, \end{aligned}$$

(recoupling relation) (7.76)

where the m-independent **Racah W-coefficient**[3] is

$$W(j_1\ j_2\ j\ j_3; j_{12}\ j_{23}) := \frac{\langle j_1, j_{23}; j\ m | j_{12}, j_3; j\ m\rangle}{\sqrt{(2j_{12}+1)(2j_{23}+1)}} \tag{7.77}$$

(Racah W-coefficient)

for any m, and the m-independent **Wigner 6-j symbol**[4] is

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_{12} \\ j_3 & j & j_{23} \end{array} \right\} := \frac{(-1)^{j_1+j_2+j_3+j}\langle j_1, j_{23}; j\ m | j_{12}, j_3; j\ m\rangle}{\sqrt{(2j_{12}+1)(2j_{23}+1)}}. \tag{7.78}$$

(Wigner 6-j symbol)

Evidently, the two symbols are related by

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{array} \right\} = (-1)^{j_1+j_2+l_1+l_2} W(j_1\ j_2\ l_2\ l_1; j_3\ l_3). \tag{7.79}$$

(relation between Racah and Wigner 6-j symbols)

The two symbols are equivalent up to a sign, with the Wigner 6-j symbol being somewhat more symmetric in terms of permutation relations, as we discuss below. Both symbols are commonly used, though we will generally stick with the 6-j symbol.

Note that from the definition in Eq. (7.73), the (real) inner product $\langle j_1; j_{23}; j\ m | j_{12}; j_3; j\ m\rangle$ is a unitary and thus orthogonal matrix, with rows and columns labeled by j_{12} and j_{23}. In particular, this means that

$$\sqrt{(2j'+1)(2j''+1)} \left\{ \begin{array}{ccc} j_1 & j_2 & j' \\ j_3 & j_4 & j'' \end{array} \right\} \tag{7.80}$$

represents an orthogonal matrix with indices j' and j''. Then matrix multiplication with its transpose leads to the identity matrix,

$$\sum_j \sqrt{(2j+1)(2j'+1)} \left\{ \begin{array}{ccc} j_1 & j_2 & j \\ j_3 & j_4 & j' \end{array} \right\} \sqrt{(2j+1)(2j''+1)} \left\{ \begin{array}{ccc} j_1 & j_2 & j \\ j_3 & j_4 & j'' \end{array} \right\} = \delta_{j'j''}. \tag{7.81}$$

Since the result is nonzero only when $j' = j''$, we may instead write

$$\sum_j (2j+1)(2j''+1) \left\{ \begin{array}{ccc} j_1 & j_2 & j \\ j_3 & j_4 & j' \end{array} \right\} \left\{ \begin{array}{ccc} j_1 & j_2 & j \\ j_3 & j_4 & j'' \end{array} \right\} = \delta_{j'j''} \tag{7.82}$$

(orthogonality relation)

as an orthogonality relation in terms of 6-j symbols.

7.1.4.1 Explicit Forms

To obtain a more useful expression for the 6-j symbol, we can first invert Eq. (7.72) to obtain

$$|j_1\ m_1; j_2\ m_2; j_3\ m_3\rangle = \sum_{j_{23} m_{23} j m} |j_1; j_{23}; j\ m\rangle\ \langle j_1\ m_1; j_{23}\ m_{23} | j\ m\rangle \langle j_2\ m_2; j_3\ m_3 | j_{23}\ m_{23}\rangle. \tag{7.83}$$

Putting this into Eq. (7.71),

$$\begin{aligned} |j_{12}; j_3; j\ m\rangle = \sum_{\substack{m_1 m_2 m_3 m_{12} \\ j_{23} m_{23} j' m'}} & |j_1; j_{23}; j'\ m'\rangle\ \langle j_{12}\ m_{12}; j_3\ m_3 | j'\ m'\rangle \langle j_1\ m_1; j_2\ m_2 | j_{12}\ m_{12}\rangle \\ & \times \langle j_1\ m_1; j_{23}\ m_{23} | j'\ m'\rangle \langle j_2\ m_2; j_3\ m_3 | j_{23}\ m_{23}\rangle, \end{aligned} \tag{7.84}$$

[3] Giulio Racah, "Theory of Complex Spectra. II," *Physical Review* **62** 438 (1942) (doi: 10.1103/PhysRev.62.438); D. M. Brink and G. R. Satchler, *Angular Momentum*, 2nd ed. (Oxford, 1968), Section 3.2, p. 40.

[4] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, 1957), Section 3.3, p. 40.

and then projecting with $\langle j_1; j_{23}; j\ m|$, we find

$$\langle j_1; j_{23}; j\ m|j_{12}; j_3; j\ m\rangle = \sum_{\substack{m_1 m_2 m_3 \\ m_{12} m_{23}}} \langle j_{12}\ m_{12}; j_3\ m_3|j\ m\rangle \langle j_1\ m_1; j_2\ m_2|j_{12}\ m_{12}\rangle \times \langle j_1\ m_1; j_{23}\ m_{23}|j\ m\rangle \langle j_2\ m_2; j_3\ m_3|j_{23}\ m_{23}\rangle. \tag{7.85}$$

Thus, we have the explicit form

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_{12} \\ j_3 & j & j_{23} \end{array} \right\} = \frac{(-1)^{j_1+j_2+j_3+j}}{\sqrt{(2j_{12}+1)(2j_{23}+1)}} \sum_{\substack{m_1 m_2 m_3 \\ m_{12} m_{23}}} \langle j_{12}\ m_{12}; j_3\ m_3|j\ m\rangle \langle j_1\ m_1; j_2\ m_2|j_{12}\ m_{12}\rangle \times \langle j_1\ m_1; j_{23}\ m_{23}|j\ m\rangle \langle j_2\ m_2; j_3\ m_3|j_{23}\ m_{23}\rangle.$$

(Wigner 6-j symbol, connection to Clebsch–Gordan coefficients) (7.86)

A somewhat simpler explicit formula due to Racah[5] comes from putting in the explicit formula (7.51) for the Clebsch–Gordan coefficients into the above expression, with the result

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{array} \right\} = \Delta(j_1, j_2, j_3)\, \Delta(j_1, l_2, l_3)\, \Delta(l_1, j_2, l_3)\, \Delta(l_1, l_2, j_3) \times \sum_{n=n_{\min}}^{n_{\max}} \frac{(-1)^n (n+1)!}{(n-J)!(n-k_1)!(n-k_2)!(n-k_3)!(m_1-n)!(m_2-n)!(m_3-n)!},$$

(Wigner 6-j symbol: explicit formula) (7.87)

where we have used the shorthand symbols

$$\begin{aligned} J &= j_1 + j_2 + j_3 \\ k_1 &= j_1 + l_2 + l_3 \\ k_2 &= l_1 + j_2 + l_3 \\ k_3 &= l_1 + l_2 + j_3 \\ m_1 &= j_1 + j_2 + l_1 + l_2 \\ m_2 &= j_2 + j_3 + l_2 + l_3 \\ m_3 &= j_3 + j_1 + l_3 + l_1 \\ n_{\min} &= \max\{j, k_1, k_2, k_3\} \\ n_{\max} &= \min\{m_1, m_2, m_3\} \\ \Delta(a,b,c) &= \sqrt{\frac{(a+b-c)!(b+c-a)!(c+a-b)!}{(a+b+c+1)!}}. \end{aligned} \tag{7.88}$$

The 6-j symbol must satisfy the triangular constraints for four sets of ordered triples,

$$\begin{aligned} &(j_1, j_2, j_3): && |j_1 - j_2| \le j_3 \le j_1 + j_2, && |j_3 - j_1| \le j_2 \le j_3 + j_1, && |j_2 - j_3| \le j_1 \le j_2 + j_3, \\ &(j_1, l_2, l_3): && |j_1 - l_2| \le l_3 \le j_1 + l_2, && |l_3 - j_1| \le l_2 \le l_3 + j_1, && |l_2 - l_3| \le j_1 \le l_2 + l_3, \\ &(l_1, j_2, l_3): && |l_1 - j_2| \le l_3 \le l_1 + j_2, && |l_3 - l_1| \le j_2 \le l_3 + l_1, && |j_2 - l_3| \le l_1 \le j_2 + l_3, \\ &(l_1, l_2, j_3): && |l_1 - l_2| \le j_3 \le l_1 + l_2, && |j_3 - l_1| \le l_2 \le j_3 + l_1, && |l_2 - j_3| \le l_1 \le l_2 + j_3, \end{aligned} \tag{7.89}$$

which follow from the constraint (7.41) for the Clebsch–Gordan coefficients [where we may exchange $j_1 \longleftrightarrow j_{23}$ and $j_{12} \longleftrightarrow j_3$ in view of the definition (7.85)], the expression (7.86) for the 6-j symbol, and the permutation symmetries in the following section. (The permutation symmetries lead to yet more triangle inequalities.) The 6-j symbols must also obviously have $j_\alpha \ge 0$, $l_\alpha \ge 0$ and also $2j_\alpha \in \mathbb{Z}$ and $2l_\alpha \in \mathbb{Z}$

[5] A. R. Edmonds, *op. cit.*, p. 99; Giulio Racah, *op. cit.*

to represent angular-momentum eigenvalues (either integer or half-integer spin). For the 6-j symbol to be nonzero, the elements of the above triples must also add up to an integer,

$$\begin{aligned} j_1 + j_2 + j_3 &\in \mathbb{Z} \\ j_1 + l_2 + l_3 &\in \mathbb{Z} \\ l_1 + j_2 + l_3 &\in \mathbb{Z} \\ l_1 + l_2 + j_3 &\in \mathbb{Z}, \end{aligned} \tag{7.90}$$

due again to the expression (7.86) for the Clebsch–Gordan coefficients, which effectively add any two elements of each triple to obtain the third: all integer momenta are okay, but two half-integer momenta must add to produce an integer momentum. Of course, once the 6-j symbols are computed this way, the Racah coefficient can then be found using Eq. (7.79).

7.1.4.2 Symmetry Relations

As we mentioned above, the 6-j symbols are simpler than the Racah coefficients under permutations of the elements. We will simply summarize the symmetry relations, which follow from the above formulae, in particular by recasting Eq. (7.86) in a very symmetric form as a product of four 3-j symbols. The 6-j symbols are invariant under any exchange of columns, such as

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{array} \right\} = \left\{ \begin{array}{ccc} j_2 & j_1 & j_3 \\ l_2 & l_1 & l_3 \end{array} \right\} = \left\{ \begin{array}{ccc} j_3 & j_2 & j_1 \\ l_3 & l_2 & l_1 \end{array} \right\} = \left\{ \begin{array}{ccc} j_1 & j_3 & j_2 \\ l_1 & l_3 & l_2 \end{array} \right\}, \tag{7.91}$$

and so on. The 6-j symbols are also invariant under the following interchanges,

$$\left\{ \begin{array}{ccc} j_1 & j_2 & j_3 \\ l_1 & l_2 & l_3 \end{array} \right\} = \left\{ \begin{array}{ccc} l_1 & l_2 & j_3 \\ j_1 & j_2 & l_3 \end{array} \right\} = \left\{ \begin{array}{ccc} l_1 & j_2 & l_3 \\ j_1 & l_2 & j_3 \end{array} \right\} = \left\{ \begin{array}{ccc} j_1 & l_2 & l_3 \\ l_1 & j_2 & j_3 \end{array} \right\}, \tag{7.92}$$

where the upper and lower values are interchanged in any two columns.

The Wigner 6-j symbols will be useful in decomposing the reduced matrix elements for the dipole operator that we will derive. We will thus consider the 6-j symbols again below.

7.1.4.3 Addition of Four Angular Momenta: Wigner 9-j Symbols

Now let's consider the coupling of *four* angular momenta,[6] $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3 + \mathbf{J}_4$. Obviously, from our discussion of adding three angular momenta, there will be no unique way to add these together. For example, suppose that we add $\mathbf{J}_{12} = \mathbf{J}_1 + \mathbf{J}_2$ and $\mathbf{J}_{34} = \mathbf{J}_3 + \mathbf{J}_4$, and then finally $\mathbf{J} = \mathbf{J}_{12} + \mathbf{J}_{34}$. We can denote an eigenstate coupled in this fashion as

$$|j_{12}, j_{34}; j\ m\rangle, \tag{7.93}$$

where we keep only the last addition of uncoupled momenta to make it obvious how we arrived at the result (i.e., the dependence on j_1, j_2, j_3, and j_4 is implied). We could also couple the angular momenta in alternate pairs, adding $\mathbf{J}_{13} = \mathbf{J}_1 + \mathbf{J}_3$ and $\mathbf{J}_{24} = \mathbf{J}_2 + \mathbf{J}_4$, and then finally $\mathbf{J} = \mathbf{J}_{13} + \mathbf{J}_{24}$, where we denote the eigenstate

$$|j_{13}, j_{24}; j\ m\rangle. \tag{7.94}$$

Again, there is an m-independent, orthogonal transformation between these two bases, which we use to define the **Wigner 9-j symbol**:[7]

$$\langle j_{12}, j_{34}; j | j_{13}, j_{24}; j \rangle =: \sqrt{(2j_{12}+1)(2j_{34}+1)(2j_{13}+1)(2j_{24}+1)} \left\{ \begin{array}{ccc} j_1 & j_2 & j_{12} \\ j_3 & j_4 & j_{34} \\ j_{13} & j_{24} & j \end{array} \right\}. \tag{7.95}$$

(Wigner 9-j symbol)

[6] "Four, *four* angular momentum vectors, ah ah ah…" –Count von Count

[7] Edmonds, *op. cit.*, Section 6.4, p. 100.

Note that we have dropped the m quantum number, since the result is m-independent anyway. To obtain an expression for the 9-j symbol, we can perform the recoupling of the angular momenta in multiple steps as

$$\begin{aligned}\langle j_{12}, j_{34}; j | j_{13}, j_{24}; j\rangle &= \sum_{j_{234}} \langle j_{12}, j_{34}; j | j_1, j_{2(34)}; j\rangle \, \langle j_1, j_{2(34)}; j | j_{13}, j_{24}; j\rangle \\ &= \sum_{j_{234}} \langle j_{12}, j_{34}; j | j_1, j_{2(34)}; j\rangle \, \langle j_2, j_{34}; j_{2(34)} | j_3, j_{42}; j_{3(42)}\rangle \, \langle j_1, j_{3(42)}; j | j_{13}, j_{24}; j\rangle \\ &= \sum_{j_{234}} \langle j_1, j_{2(34)}; j | j_{12}, j_{34}; j\rangle \, (-1)^{j_2+j_{34}-j_{234}} \langle j_3, j_{42}; j_{3(42)} | j_{34}, j_2; j_{(34)2}\rangle \\ &\qquad \times (-1)^{j_2+j_4-j_{42}} \langle j_1, j_{3(42)}; j | j_{13}, j_{42}; j\rangle\end{aligned} \tag{7.96}$$

where we have used parentheses in the subscripts in cases where the order of coupling is ambiguous—though note that the *numerical* values are the same for the same set of subscripts, as in $j_{24} = j_{42}$ or $j_{234} = j_{2(34)} = j_{(34)2} = j_{3(42)}$, after any summations are carried out. In the second step, we used

$$\begin{aligned}\langle j_1, j_{2(34)}; j | &= \langle j_1, j_{2(34)}; j | j_1, j_{3(42)}; j\rangle \, \langle j_1, j_{3(42)}; j | \\ &= \langle j_2, j_{34}; j_{2(34)} | j_3, j_{42}; j_{3(42)}\rangle \, \langle j_1, j_{3(42)}; j |,\end{aligned} \tag{7.97}$$

while in the last step of Eq. (7.96), we used the symmetry rule (7.54) to change the order of two couplings. These three coefficients each represent the coupling of *three* angular momenta, and thus we can use the definition (7.78) of the 6-j symbol three times to obtain

$$\begin{aligned}\langle j_{12}, j_{34}; j | j_{13}, j_{24}; j\rangle &= \sum_{j_{234}} (-1)^{-(j_1+j_2+j_{34}+j)} \sqrt{(2j_{12}+1)(2j_{234}+1)} \begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_{34} & j & j_{234} \end{Bmatrix} \\ &\qquad \times (-1)^{j_2+j_{34}-j_{234}} (-1)^{-(j_2+j_3+j_4+j_{234})} \sqrt{(2j_{24}+1)(2j_{34}+1)} \begin{Bmatrix} j_3 & j_4 & j_{34} \\ j_2 & j_{234} & j_{24} \end{Bmatrix} \\ &\qquad \times (-1)^{j_2+j_4-j_{24}} (-1)^{-(j_1+j_3+j_{24}+j)} \sqrt{(2j_{13}+1)(2j_{234}+1)} \begin{Bmatrix} j_1 & j_3 & j_{13} \\ j_{24} & j & j_{234} \end{Bmatrix} \\ &= \sum_{j_{234}} (-1)^{2j_{234}} (2j_{234}+1) \sqrt{(2j_{12}+1)(2j_{13}+1)(2j_{24}+1)(2j_{34}+1)} \\ &\qquad \times \begin{Bmatrix} j_1 & j_2 & j_{12} \\ j_{34} & j & j_{234} \end{Bmatrix} \begin{Bmatrix} j_3 & j_4 & j_{34} \\ j_2 & j_{234} & j_{24} \end{Bmatrix} \begin{Bmatrix} j_{13} & j_{24} & j \\ j_{234} & j_1 & j_3 \end{Bmatrix},\end{aligned} \tag{7.98}$$

where in the last step we used $j_1 + j + j_{234} \in \mathbb{Z}$ and $j_3 + j_{234} + j_{24} \in \mathbb{Z}$, according to the constraints of the 6-j symbols, to simplify the expression for the sign. Comparing to the definition (7.95) of the 9-j symbol, we find the explicit formula

$$\begin{Bmatrix} j_1 & j_2 & j_3 \\ k_1 & k_2 & k_3 \\ \ell_1 & \ell_2 & \ell_3 \end{Bmatrix} = \sum_s (-1)^{2s} (2s+1) \begin{Bmatrix} j_1 & j_2 & j_3 \\ k_3 & \ell_3 & s \end{Bmatrix} \begin{Bmatrix} k_1 & k_2 & k_3 \\ j_2 & s & \ell_2 \end{Bmatrix} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ s & j_1 & k_1 \end{Bmatrix},$$

(Wigner 9-j symbol in terms of 6-j symbols) (7.99)

in terms of a sum over products of 6-j symbols, after changing to a more symmetric notation.

We will be able to accomplish what we want in terms of angular-momentum structure without having to resort to the 9-j symbol. However, it *will* help to consider the coupling of four angular momenta in a

slightly different way. First, we can couple the four angular momenta in two stages as

$$\begin{aligned}\langle j_{(12)3}\ j_4; j | j_{23}\ j_{14}; j\rangle &= \langle j_{(12)3}\ j_4; j | j_{1(23)}, j_4; j\rangle\ \langle j_{1(23)}, j_4; j | j_{23}\ j_{14}; j\rangle \\ &= \langle j_{1(23)}, j_4; j | j_{(12)3}\ j_4; j\rangle\ (-1)^{j_1+j_{23}-j_{123}} \langle j_{23}\ j_{14}; j | j_{(23)1}, j_4; j\rangle \\ &= (-1)^{-j_1-j_2-j_3-j_4-2j_{123}-j} \sqrt{(2j_{12}+1)(2j_{23}+1)(2j_{123}+1)(2j_{14}+1)} \\ &\qquad \times \left\{ \begin{array}{ccc} j_1 & j_2 & j_{12} \\ j_3 & j_{123} & j_{23} \end{array} \right\} \left\{ \begin{array}{ccc} j_{23} & j_1 & j_{123} \\ j_4 & j & j_{14} \end{array} \right\},\end{aligned} \tag{7.100}$$

while we can make the same coupling in three stages, as we did for the 9-j symbol:

$$\begin{aligned}\langle j_{(12)3}\ j_4; j | j_{23}\ j_{14}; j\rangle &= \sum_{j_{124}} \langle j_{(12)3}\ j_4; j | j_{(12)4}, j_3; j\rangle\ \langle j_{(12)4}, j_3; j | j_{(14)2}, j_3; j\rangle\ \langle j_{(14)2}, j_3; j | j_{23}\ j_{14}; j\rangle \\ &= \sum_{j_{124}} (-1)^{j_{123}+j_4-j} (-1)^{j_{12}+j_4-j_{124}} \langle j_4\ j_{(12)3}; j | j_{4(12)}, j_3; j\rangle \\ &\qquad \times (-1)^{j_{12}+j_4-j_{124}} (-1)^{j_1+j_4-j_{14}} \langle j_{4(12)}, j_3; j | j_{(41)2}, j_3; j\rangle \\ &\qquad \times (-1)^{j_{14}+j_{23}-j} \langle j_{14}\ j_{23}; j | j_{(14)2}, j_3; j\rangle \\ &= \sum_{j_{124}} (-1)^{j_1-2j+2j_{12}+j_{23}+j_{123}-2j_{124}} \\ &\qquad \times (-1)^{-(j_4+j_{12}+j_3+j)} \sqrt{(2j_{123}+1)(2j_{124}+1)} \left\{ \begin{array}{ccc} j_4 & j_{12} & j_{124} \\ j_3 & j & j_{123} \end{array} \right\} \\ &\qquad \times (-1)^{-(j_4+j_1+j_2+j_{124})} \sqrt{(2j_{12}+1)(2j_{14}+1)} \left\{ \begin{array}{ccc} j_4 & j_1 & j_{14} \\ j_2 & j_{124} & j_{12} \end{array} \right\} \\ &\qquad \times (-1)^{-(j_{14}+j_2+j_3+j)} \sqrt{(2j_{23}+1)(2j_{124}+1)} \left\{ \begin{array}{ccc} j_{14} & j_2 & j_{124} \\ j_3 & j & j_{23} \end{array} \right\}.\end{aligned} \tag{7.101}$$

Equating these two expressions and permuting some of the 6-j symbol elements, we find the **Biedenharn–Elliott sum rule** [8]

$$\begin{aligned}\left\{ \begin{array}{ccc} j_1 & j_2 & j_{12} \\ j_3 & j_{123} & j_{23} \end{array} \right\} \left\{ \begin{array}{ccc} j_{23} & j_1 & j_{123} \\ j_4 & j & j_{14} \end{array} \right\} &= \sum_{j_{124}} (-1)^{j_1+j_2+j_3+j_4+j_{12}+j_{23}+j_{14}+j_{123}+j_{124}+j} (2j_{124}+1) \\ &\qquad \times \left\{ \begin{array}{ccc} j_3 & j_2 & j_{23} \\ j_{14} & j & j_{124} \end{array} \right\} \left\{ \begin{array}{ccc} j_2 & j_1 & j_{12} \\ j_4 & j_{124} & j_{14} \end{array} \right\} \left\{ \begin{array}{ccc} j_3 & j_{12} & j_{123} \\ j_4 & j & j_{124} \end{array} \right\}\end{aligned}$$

(Biedenharn–Elliott sum rule) (7.102)

after using the usual tricks to simplify the sign factor.

7.2 Static Angular-Momentum Structure of Atoms

In the standard textbook version of the nonrelativistic, quantum-mechanical hydrogen-like atom,[9] an electron of reduced mass

$$m = \frac{m_{\rm e} m_{\rm n}}{m_{\rm e} + m_{\rm n}} \approx m_{\rm e}, \tag{7.103}$$

where $m_{\rm e}$ is the electron mass, and $m_{\rm n}$ is the nuclear mass, moves in the central potential

$$V(r) = -\frac{Ze^2}{4\pi\epsilon_0 r}, \tag{7.104}$$

[8] L. C. Biedenharn, "An Identity Satisfied by the Racah Coefficients," *Journal of Mathematics and Physics* **31**, 287 (1953); J. P. Elliott, "Theoretical Studies in Nuclear Structure. V. The Matrix Elements of Non-Central Forces with an Application to the 2p-Shell," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **218**, 345 (1953); A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, 1957), p. 97, Eq. (6.2.12).

[9] See, e.g., John L. Powell and Bernd Crasemann, *Quantum Mechanics* (Addison–Wesley, 1961), Section 7-7, p. 220.

where e is the fundamental charge, and the nuclear charge is Ze. The standard result for the energies is

$$E_n = -\left(\frac{mc^2}{2}(Z\alpha)^2\right)\frac{1}{n^2}, \tag{7.105}$$

where n is the radial (principle) quantum number,

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} \tag{7.106}$$

is the **fine-structure constant**, and the coefficient of $1/n^2$ has the approximate value $-Z^2(13.6\text{ eV})$. This energy expression says that, at this crude level of approximation, the hydrogen-like-atom energies do not depend on any angular-momentum quantum numbers.

In what follows, we will use the standard notation of n, l, and m as the usual quantum numbers referring to the single-electron state $|n\ l\ m\rangle$. However, to be a bit more general, for multielectron atoms we will refer to the total quantities using capital letters. That is, L is the quantum number for the *total* electron orbital angular momentum, S is the quantum number for the total electron spin, and so on.

7.2.1 Fine Structure

At the next level of approximation, we find that angular momentum *does* contribute some energy shifts, splitting some of the degenerate lines in the above simplistic treatment. Because of the relatively small splittings, at least in lighter atoms—for example 0.58 nm for the common yellow 589 nm line of sodium (seen in sodium lamps everywhere)—this splitting goes by the name of **fine structure**.[10] We can treat this effect to lowest order as follows. Because the electron orbits the nucleus, it moves through the nuclear Coulomb field and thus ''sees'' in its rest frame an effective magnetic field

$$\mathbf{B} = -\frac{\mathbf{v}}{c^2}\times\mathbf{E} \tag{7.107}$$

via the Lorentz transformation for electromagnetic fields.[11] The Coulomb force on the electron is

$$\mathbf{F} = -e\mathbf{E} = -\nabla V(\mathbf{r}) = -\frac{\mathbf{r}}{r}\partial_r V(r), \tag{7.108}$$

and so with the orbital angular momentum

$$\mathbf{L} = \mathbf{r}\times\mathbf{p} = m\mathbf{r}\times\mathbf{v}, \tag{7.109}$$

the effective magnetic field becomes

$$\mathbf{B} = -\frac{\partial_r V(r)}{c^2 er}\mathbf{v}\times\mathbf{r} = \frac{\partial_r V(r)}{mc^2 er}\mathbf{L}. \tag{7.110}$$

The electron's magnetic moment due to its intrinsic spin is

$$\boldsymbol{\mu}_S = -\mu_{\text{B}} g_S \frac{\mathbf{S}}{\hbar}, \tag{7.111}$$

where $\mu_{\text{B}} := e\hbar/2m_{\text{e}}$ is the **Bohr magneton** ($\mu_{\text{B}} = 9.274\ 009\ 15(23)\times 10^{-24}$ J/T $= h\cdot 1.399\ 624\ 604(35)$ MHz/G[12]), g_S is the electron g-factor ($g_S \approx 2$ in Dirac theory, but due to quantum effects is slightly larger,

[10] Fine structure was first described by A. Sommerfeld, ''Zur Quantentheorie der Spektrallinien,'' *Annalen der Physik* **356**, 1 (1916) (doi: 10.1002/andp.19163561702).

[11] See David J. Griffiths, *Introduction to Electrodynamics*, 2nd ed. (Prentice-Hall, 1989), Eq. (10.120), p. 497.

[12] 2006 CODATA recommended value; see P. J. Mohr, B. N. Taylor, and D. B. Newell, ''The 2006 CODATA Recommended Values of the Fundamental Physical Constants, Web Version 5.1,'' available at `http://physics.nist.gov/constants` (National Institute of Standards and Technology, Gaithersburg, MD 20899, 31 December 2007).

$g_S = 2.002\ 319\ 304\ 3622(15)$[13]), and $\mathbf{S}$ is the electron spin operator. The interaction energy of the magnetic field (7.110) with the magnetic moment (7.111) gives the fine-structure Hamiltonian

$$H_{\rm fs} = -\boldsymbol{\mu}_S \cdot \mathbf{B} = \frac{\mu_{\rm B} g_S \partial_r V(r)}{mc^2 \hbar e r}\, \mathbf{L}\cdot\mathbf{S} = \left(\frac{Ze^2}{4\pi\epsilon_0}\right)\frac{g_S}{2m^2c^2r^3}\,\mathbf{L}\cdot\mathbf{S}. \tag{7.112}$$

The problem with this expression is that we still need to transform back into the lab frame, which is a noninertial transformation. This correction is **Thomas precession**, and the correction amounts to adding the Thomas-precession Hamiltonian[14]

$$H_{\rm T} = -H_{\rm fs}(g_S \longrightarrow 1), \tag{7.113}$$

so that the real fine-structure Hamiltonian is

$$H_{\rm fs} + H_{\rm T} = \frac{\mu_{\rm B}(g_S - 1)\partial_r V(r)}{mc^2 \hbar e r}\, \mathbf{L}\cdot\mathbf{S} = \left(\frac{Ze^2}{4\pi\epsilon_0}\right)\frac{(g_S-1)}{2m^2c^2r^3}\,\mathbf{L}\cdot\mathbf{S}. \tag{7.114}$$

The coupling is thus proportional to $\mathbf{L}\cdot\mathbf{S}$.

The uncoupled states are of the form $|n\ L\ m_L\ m_S\rangle$, where L are represent eigenvalues of the $\mathbf{L^2}$ operator, m_L represent eigenvalues of the L_z operator, and m_S represent eigenvalues of the S_z operator. (We suppress dependence on the S quantum number, since it is always 1/2.) Under the $\mathbf{L}\cdot\mathbf{S}$ coupling, these are no longer good quantum numbers. We can thus introduce the composite quantum number

$$\mathbf{J} = \mathbf{L} + \mathbf{S}, \tag{7.115}$$
(fine-structure angular momentum)

where from the triangularity condition (7.41) we have the new quantum number in the range

$$|L - S| \le J \le L + S \tag{7.116}$$

The magnitude of $\mathbf{J}$ is

$$\mathbf{J}^2 = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L}\cdot\mathbf{S}, \tag{7.117}$$

or solving for the dot product,

$$\mathbf{L}\cdot\mathbf{S} = \frac{1}{2}\left(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2\right). \tag{7.118}$$

Thus, under the interaction we may still have eigenstates of L, S, J and m_J. In particular, the fine-structure shift due to this interaction is given to by simply taking the expectation value of the interaction Hamltonian in the coupled basis:

$$\begin{aligned}\Delta E_{\rm fs} &= \langle n; L, S; J\ m_J | H_{\rm fs} | n/L, S; J\ m_J\rangle \\ &= \frac{\mu_{\rm B}\hbar(g_S-1)\langle n|(1/r)\partial_r V(r)|n\rangle}{2mc^2 e}\left[J(J+1) - L(L+1) - S(S+1)\right].\end{aligned} \tag{7.119}$$

The fine-structure shift then depends on J, breaking the degeneracy of different L levels. Actually, in writing down this expression, we are ignoring a relativistic correction of similar order, which also depends on L.[15] However, the form of the perturbation is the important issue; when dealing with heavier alkali atoms it is difficult to obtain a quantitatively accurate expression anyway. Thus the main point is the introduction of the composite angular momentum $\mathbf{J}$ to label the fine-structure states, where the energy levels can be taken to be experimentally known, generally to high accuracy.

[13] 2006 CODATA recommended value.
[14] John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 11.8, p. 548.
[15] David J. Griffiths, *Introduction to Quantum Mechanics* (Prentice-Hall, 1995), Section 6.3.1, p. 236.

7.2.1.1 Spectroscopic Notation for Simple Atoms

Now that we have introduced the basics of atomic structure, we briefly describe the common spectroscopic labels, of the form

$$n\,^{2S+1}L_J. \tag{7.120}$$

The n is the principal quantum number of the active electron. The $2S+1$ gives the multiplicity of the electron spin, or the number of possible electron-spin states. For a single-electron (hydrogen-like) atom, $2S+1$ is always 2, since $S = 1/2$. The L quantum number is represented by a letter: S for $L = 0$, P for $L = 1$, D for $L = 2$, F for $L = 3$, G for $L = 4$, H for $L = 5$, and so on. The first four letters stand for descriptors for lines in alkali spectra (sharp, principal, diffuse, fundamental), and the rest are alphabetic continuations. Since the scheme is based on abbreviations, the letters should, in the author's opinion, be set in roman, not italics, as is commonly the case in the literature. Finally, the subscript indicates the J quantum number. For example, the principle $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}\rangle$ laser-cooling transition for cesium is the D_2 line, which is written

$$6\,^2\mathrm{S}_{1/2} \longrightarrow 6\,^2\mathrm{P}_{3/2}. \tag{7.121}$$

The principal quantum number is 6 for both states, being the lowest available for the valence electron. Again for a single active electron in alkali atoms, $2S+1 = 2$. The ground and excited L quantum numbers are 0 and 1, respectively. Finally, $J = 1/2$ for the ground state, while $J = 3/2$ for the excited state (since $L = 1$ for the excited state, either $J = 1/2$ or $J = 3/2$, corresponding to $|L - S$ and $L + S$, respectively).

As a slightly more complicated example, we consider strontium, which has two valence electrons, and a slightly more general notation. Here, for example, the narrow, second-stage laser-cooling transition is the ''intercombination line''

$$5\,\mathrm{s}^2\,{}^1\mathrm{S}_0 \longrightarrow 5\,\mathrm{p}\,{}^3\mathrm{P}_1. \tag{7.122}$$

(The $^1\mathrm{S}_0$ here would be read as ''singlet S zero,'' and the $^3\mathrm{P}_1$ would be read as ''triplet P one.'') Note that the configurations for the two electrons are also given in lower case. In the ground state, both electrons are in the s ($l = 0$) orbital, while in the excited state, one is promoted to the p ($l = 1$) orbital (the 5s for the unexcited electron is implied). In these two levels the respective total orbital quantum number is given also by S ($L = 0$) and P ($L = 1$). The ground-state $2S+1$ value reflects $S = 0$, while the excited-state value reflects the other possible $S = 1$ value, where $\mathbf{S}$ now represents the total electron spin $\mathbf{S}_1 + \mathbf{S}_2$. Hence, the ''intercombination line'' name, since the transition flips an atomic spin, which is electric-dipole-forbidden (a magnetic field must flip the spin). Finally $J = 0$ for the ground state, since $S = L = 0$, but for the excited state, where $S = 1$ and $L = 1$, J could be 0, 1, or 2.

7.2.2 Hyperfine Structure

The **hyperfine structure** of an atom arises from the interaction between the total atomic angular momentum $\mathbf{J}$ and the nuclear angular momentum $\mathbf{I}$. We will develop this a bit more carefully, as in atomic physics and quantum optics a single laser could interact almost resonantly with hyperfine-split states, which is not as often the case with fine structure. The basic idea is essentially the same as for the fine-structure case. The nuclear magnetic moment is

$$\boldsymbol{\mu}_I = -\mu_\mathrm{B} g_I \frac{\mathbf{I}}{\hbar}, \tag{7.123}$$

where $\mathbf{I}$ is the nuclear spin operator and g_I is the nuclear spin g-factor.[16] Again, the electron is effectively a current loop, and generates a magnetic field of the form

$$\mathbf{B} = -b\mathbf{J}, \tag{7.124}$$

where b is some positive constant, since $\mathbf{B}$ and $\mathbf{J}$ should be antiparallel for an electron where the charge is negative. The interaction is then given by (ignoring hyperfine couplings between different J, and thus

[16] Experimentally measured values for g_I for the alkali atoms are given by E. Arimondo, M. Inguscio, and P. Violino, ''Experimental determinations of the hyperfine structure in the alkali atoms,'' *Reviews of Modern Physics* **49**, 31 (1977) (doi: 10.1103/RevModPhys.49.31).

assuming J is still a good quantum number)

$$H_{\rm hfs} = -\boldsymbol{\mu}_I \cdot \mathbf{B} = -\frac{\mu_{\rm B} g_I b}{\hbar}\, \mathbf{I}\cdot\mathbf{J} =: A_{\rm hfs}\, \frac{\mathbf{I}\cdot\mathbf{J}}{\hbar^2}, \tag{7.125}$$

where, since we have considered the interaction of the nuclear and electron magnetic dipoles, $A_{\rm hfs}$ is called the **magnetic dipole hyperfine constant** and has the dimension of energy. For all abundant alkali atoms, $A_{\rm hfs} > 0$, since $g_I < 0$ (with the exception of ^{40}K, where the sign is opposite). The form of this interaction ($\mathbf{I}\cdot\mathbf{J}$) is very similar to the fine-structure ($\mathbf{L}\cdot\mathbf{S}$) interaction.

As in the fine-structure case, we can add the angular momenta to obtain the total atomic angular momentum

$$\mathbf{F} = \mathbf{J} + \mathbf{I}. \tag{7.126}$$
(hyperfine-structure angular momentum)

Under this interaction, we can use the new **hyperfine** quantum number F to label the new eigenstates; as in the fine-structure case, we square Eq. (7.126) to obtain the operator equation

$$\mathbf{F}^2 = \mathbf{J}^2 + \mathbf{I}^2 + 2\mathbf{I}\cdot\mathbf{J}, \tag{7.127}$$

and thus when the operator $\mathbf{I}\cdot\mathbf{J}$ acts on a hyperfine state $|JIF\rangle$ we find that it is an eigenstate,

$$(\mathbf{I}\cdot\mathbf{J})|JIF\rangle = \frac{\hbar^2 K}{2}|JIF\rangle, \tag{7.128}$$

where the eigenvalue is written in terms of the combination

$$K = F(F+1) - I(I+1) - J(J+1) \tag{7.129}$$

of angular-momentum quantum numbers. Thus, the energy shift due to this interaction is simply

$$\Delta E_{\rm hfs} = \frac{1}{2} A_{\rm hfs} K. \tag{7.130}$$

Here $A_{\rm hfs}$ can be computed, though the calculation can be complex, or simply experimentally measured. The hyperfine shift is much smaller than the fine-structure shift. This is because of the weak nuclear magnetic moment: while the electron moment was of the order of $\mu_{\rm B} = e\hbar/2m_{\rm e}$, the nuclear moment is of the order of $\mu_{\rm N} = e\hbar/2m_{\rm p}$, where $m_{\rm p}$ is the proton mass. The nuclear moment is thus smaller by a factor on the order of $m_{\rm e}/m_{\rm p} \approx 1/1836$, and so the hyperfine interaction should be smaller than the fine-structure interaction by a factor on the same order.

The higher-order corrections to this simple theory become quite involved, and so we summarize the main points here.[17] In general, the interaction between the nuclear and electron angular momenta can be expanded in a multipole series,

$$H_{\rm hfs} = \sum_k \mathbf{T}^{(k)}_{\rm e} \cdot \mathbf{T}^{(k)}_{\rm n}, \tag{7.131}$$

where $\mathbf{T}^{(k)}_{\rm e}$ and $\mathbf{T}^{(k)}_{\rm n}$ are spherical tensor operators of rank k (defined below in Section 7.3.3), that respectively operate on only the electronic and nuclear Hilbert spaces. The $k = 0$ monopole term has already been included in the fine-structure calculation. We have treated the $k = 1$ magnetic-dipole term above. The $k = 2$ and $k = 3$ terms correspond respectively to the electric-quadrupole and magnetic-octupole terms. Because we will compute expectation values with respect to $|J\,I\,F\rangle$ states, due to parity considerations either the electric or magnetic interaction will alternatingly vanish at each multipole order (with the electric dipole operator coupling only states of opposite parity, for example). The interaction between the electron and

[17] Charles Schwartz, "Theory of Hyperfine Structure," *Physical Review* **97**, 380 (1955) (doi: 10.1103/PhysRev.97.380). The electric hexadecapole term is given here in addition to the ones we have shown. See also Lloyd Armstrong, Jr., *Theory of the Hyperfine Structure of Free Atoms* (Wiley-Interscience, New York, 1971), Eqs. (IV-25), (IV-28), and (IV-31).

nuclear angular momenta is given by evaluating these operators, with the result up to the magnetic-octupole contribution reading

$$H_{\rm hfs} = A_{\rm hfs}\frac{\mathbf{I}\cdot\mathbf{J}}{\hbar^2} + B_{\rm hfs}\frac{\frac{3}{\hbar^2}(\mathbf{I}\cdot\mathbf{J})^2 + \frac{3}{2\hbar}(\mathbf{I}\cdot\mathbf{J}) - I(I+1)J(J+1)}{2I(2I-1)J(2J-1)} + C_{\rm hfs}\frac{\frac{10}{\hbar^3}(\mathbf{I}\cdot\mathbf{J})^3 + \frac{20}{\hbar^2}(\mathbf{I}\cdot\mathbf{J})^2 + \frac{2}{\hbar}(\mathbf{I}\cdot\mathbf{J})[I(I+1)+J(J+1)+3-3I(I+1)J(J+1)] - 5I(I+1)J(J+1)}{I(I-1)(2I-1)J(J-1)(2J-1)}. \tag{7.132}$$

Again, the three terms on the right-hand side respectively represent magnetic-dipole (applicable for $I, J > 0$), electric-quadrupole (applicable for $I, J > 1/2$), and magnetic-octupole (applicable for $I, J > 1$) couplings. Thus, $A_{\rm hfs}$ is the magnetic-dipole hyperfine constant, $B_{\rm hfs}$ is the electric-quadrupole hyperfine constant, and $C_{\rm hfs}$ is the magnetic-octupole hyperfine constant. We can see which terms are applicable to a given transition from the general product rule (7.267) that we prove later for commuting tensor operators,

$$\langle J, I; F\ m_F|\mathbf{T}^{(k)}_{\rm e}\cdot\mathbf{T}^{(k)}_{\rm n}|J, I; F\ m_F\rangle = (-1)^{J+I+F}\sqrt{(2J+1)(2I+1)}\left\{\begin{array}{ccc} F & I & J \\ k & J & I\end{array}\right\}\langle J||\mathbf{T}^{(k)}_{\rm e}||J\rangle\,\langle I||\mathbf{T}^{(k)}_{\rm n}||I\rangle, \tag{7.133}$$

where to satisfy the triangle inequalities for the 6-j symbol, we must have $J \geq k/2$ and $I \geq k/2$.

The eigenenergies under the hyperfine interaction may then be written in terms of the shift

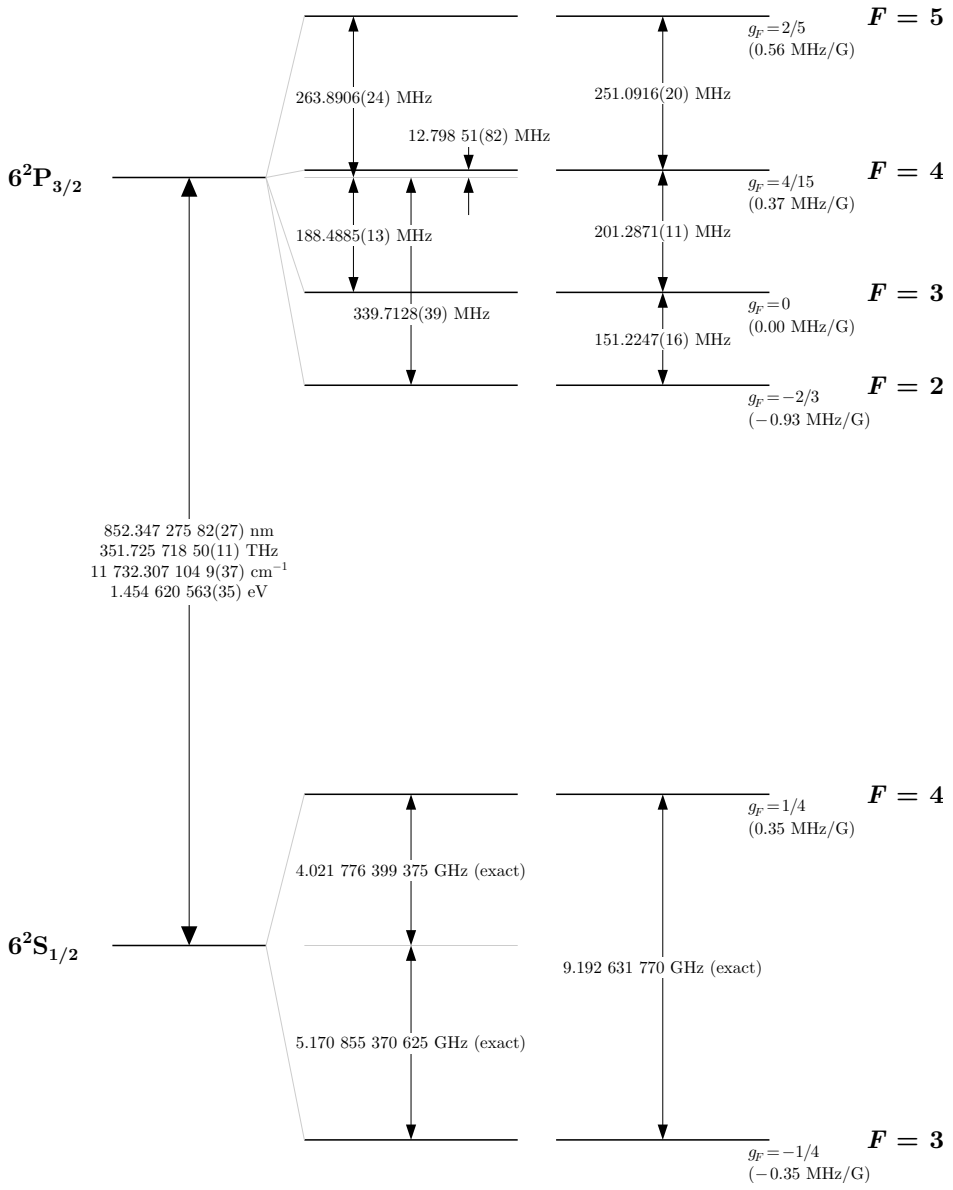
$$\Delta E_{\rm hfs} = \frac{1}{2}A_{\rm hfs}K + B_{\rm hfs}\frac{\frac{3}{2}K(K+1) - 2I(I+1)J(J+1)}{4I(2I-1)J(2J-1)} + C_{\rm hfs}\frac{5K^2(K/4+1) + K[I(I+1)+J(J+1)+3-3I(I+1)J(J+1)] - 5I(I+1)J(J+1)}{I(I-1)(2I-1)J(J-1)(2J-1)}. \tag{7.134}$$

Generally the effect of the last octupole term is quite difficult to observe, but it has been observed in the hyperfine structure of cesium,[18] where for example the values $A_{\rm hfs} = 50.288\ 27(23)$ MHz, $B_{\rm hfs} = -0.4934(17)$ MHz, and $C_{\rm hfs} = 0.56(7)$ kHz were reported for the $6^2{\rm P}_{3/2}$ (D_2 excited) level manifold. Thus, the octupole interaction contributes in cesium to the hyperfine splittings only at the kHz level, a very difficult level of accuracy to achieve in observing optical transitions.

To illustrate the hyperfine structure, the hyperfine structure of the cesium D_2 (laser-cooling) transition is shown here.[19] Note that the ground-state hyperfine splitting of ^{133}Cs is particularly significant, as it defines our measure of time: the second is defined such that the ground-state hyperfine splitting of an isolated ^{133}Cs atom is exactly 9.192 631 770 GHz.

[18] Vladislav Gerginov, Andrei Derevianko, and Carol E. Tanner, ''Observation of the Nuclear Magnetic Octupole Moment of ^{133}Cs,'' *Physical Review Letters* **91**, 072501 (2003) (doi: 10.1103/PhysRevLett.91.072501).

[19] For sources of the measured values, along with more compiled data and a more terse description of hyperfine structure, see Daniel A. Steck, ''Cesium D_2 Line Data,'' available online at `http://steck.us/alkalidata`.



7.3 Rotations and Irreducible Tensor Operators

7.3.1 Rotation Operator

What is the operator that induces a rotation in quantum mechanics? Rather than deduce it directly, we will ''cheat'' and simply quantize the classical version of a rotation. Consider the classical angle (generalized coordinate) ζ, with conjugate (angular) momentum J. The Hamiltonian $H = J$ leads to Hamilton's equations $\partial_t \zeta = 1$ and $\partial_t J = 0$. That is, J is a constant of the motion, and ζ evolves linearly in time as $\zeta(t) = t$. Thus, time evolution according to this Hamiltonian is equivalent to a rotation through an angle t. Quantum mechanically, since this Hamiltonian is time-independent, the time-evolution operator is

$$U(t,0) = \exp\left(-\frac{iHt}{\hbar}\right) = \exp\left(-\frac{iJ\zeta}{\hbar}\right) =: R(\zeta), \tag{7.135}$$

then we have the rotation operator (for a two-dimensional system) for a rotation through angle ζ. Generalizing the Hamiltonian to $H = J_x + J_y + J_z$, a rotation by angle $\boldsymbol{\zeta}$ (that is, a rotation about the $\boldsymbol{\zeta}$-axis by

angle ζ) is induced by the unitary rotation operator

$$R(\boldsymbol{\zeta}) := \exp\left(-\frac{i\mathbf{J}\cdot\boldsymbol{\zeta}}{\hbar}\right). \tag{7.136}$$
(rotation operator)

Note that a rotation of an angular-momentum state $|j\ m\rangle$ about the z-axis (the quantization axis) is particularly simple, as

$$R(\zeta\hat{z})|j\ m\rangle = \exp\left(-\frac{iJ_z\zeta}{\hbar}\right)|j\ m\rangle = e^{-im\zeta}|j\ m\rangle. \tag{7.137}$$
(rotation operator)

However, a rotation about any other axis is more complicated, as the result will in general be a superposition of angular-momentum states. Being a rotation, the j quantum number must be left unchanged (following directly from $[J^2, \mathbf{J}] = 0$, since this implies $[J^2, R(\boldsymbol{\zeta})] = 0$), but the superposition will involve states of other m values. This is a good way to see that the angular orientation of a state is encoded in the m quantum number; the j quantum number by itself doesn't tell you about orientation.

7.3.1.1 Rotation Matrix

To formalize the transformation of $|j\ m\rangle$ into a superposition of states $|j\ m'\rangle$ by a rotation we can write out an explicit rotation matrix in the basis of angular-momentum states. Wigner's convention is to write such a matrix as

$$R(\boldsymbol{\zeta})|j\ m\rangle = \sum_{m'=-j}^{j} |j\ m'\rangle\, d^{(j)}_{m'm}(\boldsymbol{\zeta}), \tag{7.138}$$
(action of rotation matrix)

where

$$d^{(j)}_{m'm}(\boldsymbol{\zeta}) := \langle j\ m'|R(\boldsymbol{\zeta})|j\ m\rangle = \langle j\ m'|e^{-i\mathbf{J}\cdot\boldsymbol{\zeta}/\hbar}|j\ m\rangle. \tag{7.139}$$
(rotation matrix)

Note the "backwards" convention for the matrix indices for the matrix-vector product in Eq. (7.138). The point is that there is a $(2j+1)\times(2j+1)$ rotation matrix $\mathbf{d}^{(j)}(\boldsymbol{\zeta})$ associated with the rotation operator $R(\boldsymbol{\zeta})$ when it acts on a subspace of angular-momentum states with fixed quantum number j.

If we follow one rotation $R(\boldsymbol{\alpha})$ by another rotation $R(\boldsymbol{\beta})$, we can represent the total rotation by a composite rotation operator R:

$$R = R(\boldsymbol{\beta})R(\boldsymbol{\alpha}). \tag{7.140}$$

Projecting into the angular-momentum representation and using the completeness relation,

$$\langle j\ m|R|j\ m'\rangle = \sum_{m''}\langle j\ m|R(\boldsymbol{\beta})|j\ m''\rangle\langle j\ m''|R(\boldsymbol{\alpha})|j\ m'\rangle. \tag{7.141}$$

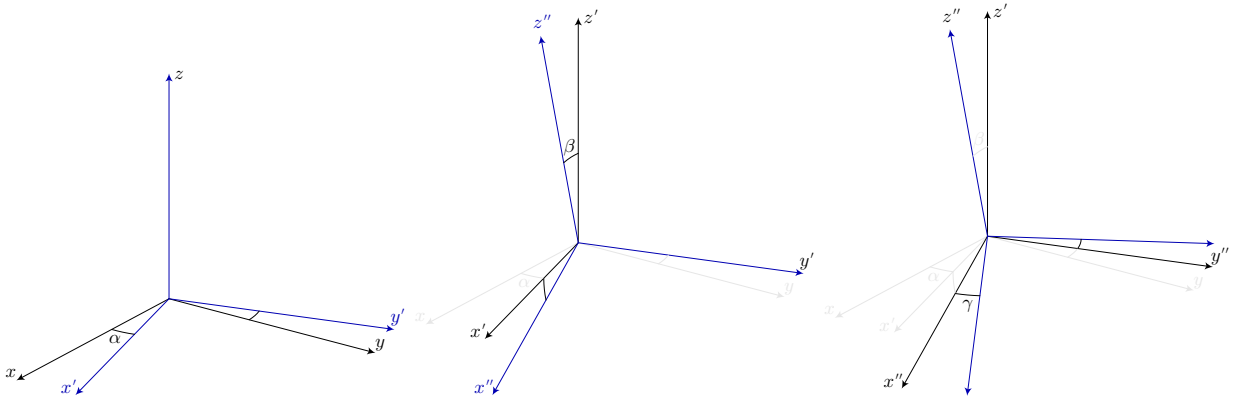
The corresponding rotation matrices thus compose by normal matrix multiplication, so long as the first rotation to operate is the rightmost:

$$\mathbf{d}^{(j)} = \mathbf{d}^{(j)}(\boldsymbol{\beta})\,\mathbf{d}^{(j)}(\boldsymbol{\alpha}). \tag{7.142}$$
(composition of rotations)

This property is very useful in decomposing arbitrary rotations, as we now discuss.

7.3.1.2 Euler Angles

As in classical mechanics, a general rotation may be represented as a composition of rotations through the three **Euler angles**: first rotate about the z-axis by angle α, then rotate about the new y-axis by angle β, and finally rotate about the new z-axis by angle γ. These angles are illustrated in the diagrams below. Note that the rotation operators act on the *state*, not the coordinate system; however, we are *also* considering rotations of the axes solely to define the second and third rotations.



Thus an arbitrary rotation R may always decomposed in the form

$$R(\alpha,\beta,\gamma) = R(\boldsymbol{\gamma})R(\boldsymbol{\beta})R(\boldsymbol{\alpha}), \tag{7.143}$$

where again $\boldsymbol{\alpha} = \alpha\hat{z}$, $\boldsymbol{\beta} = \beta\hat{y}'$, where $\hat{y}'$ is along the new y-direction after the $\boldsymbol{\alpha}$ rotation, and $\boldsymbol{\gamma} = \gamma\hat{z}''$, where $\hat{z}''$ is along the new z-direction after the $\boldsymbol{\beta}$ rotation. Clearly, $R(\boldsymbol{\alpha})$ is written to the right since it is the first rotation, and thus must operate first on the state vector. (The order is important, because in general rotation operators for different axes do not commute.) Writing these operators out explicitly,

$$R(\alpha,\beta,\gamma) = e^{-i\gamma J_{z''}/\hbar}\, e^{-i\beta J_{y'}/\hbar}\, e^{-i\alpha J_z/\hbar}. \tag{7.144}$$

But now, since $R(\boldsymbol{\beta})$ is written in terms of the coordinate system after the $R(\boldsymbol{\alpha})$ rotation, we can write this rotation as a rotated version of the operator in the original coordinate system:

$$e^{-i\beta J_{y'}/\hbar} = R(\boldsymbol{\alpha})\, e^{-i\beta J_y/\hbar}\, R^\dagger(\boldsymbol{\alpha}) = e^{-i\alpha J_z/\hbar}\, e^{-i\beta J_y/\hbar}\, e^{i\alpha J_z/\hbar}. \tag{7.145}$$

Similarly, for the last rotation, we can write

$$e^{-i\gamma J_{z''}/\hbar} = R(\boldsymbol{\beta})\, e^{-i\gamma J_{z'}/\hbar}\, R^\dagger(\boldsymbol{\beta}) = e^{-i\beta J_{y'}/\hbar}\, e^{-i\gamma J_{z'}/\hbar}\, e^{i\beta J_{y'}/\hbar}, \tag{7.146}$$

and putting this into Eq. (7.144), we find

$$R(\alpha,\beta,\gamma) = e^{-i\beta J_{y'}/\hbar}\, e^{-i\gamma J_{z'}/\hbar}\, e^{-i\alpha J_z/\hbar}. \tag{7.147}$$

Now putting in Eq. (7.145) and the analogous result with $J_{z'}$,

$$\boxed{R(\alpha,\beta,\gamma) = e^{-i\alpha J_z/\hbar}\, e^{-i\beta J_y/\hbar}\, e^{-i\gamma J_z/\hbar}.} \tag{7.148}$$
(rotation operator, Euler angles)

Conveniently, then, a rotation according to the Euler angles may be implemented solely in the original coordinate system, if the order of the rotations is *reversed*.

Now to return to the rotation matrix. Using the definition (7.139) for the matrix corresponding to this rotation operator,

$$d^{(j)}_{m'm}(\alpha,\beta,\gamma) = \langle j\ m'|R|j\ m\rangle = \langle j\ m'|e^{-i\alpha J_z/\hbar}\, e^{-i\beta J_y/\hbar}\, e^{-i\gamma J_z/\hbar}|j\ m\rangle. \tag{7.149}$$

The first and last rotations are thus easy to represent, leaving the second rotation as the only nontrivial one:

$$\boxed{d^{(j)}_{m'm}(\alpha,\beta,\gamma) = e^{-im'\alpha}\, d^{(j)}_{m'm}(\beta\hat{y})\, e^{-im\gamma}.} \tag{7.150}$$
(rotation matrix, Euler angles)

Wigner's explicit expression for the remaining rotation matrix is[20]

$$d^{(j)}_{m'm}(\beta\hat{y}) = \sqrt{(j+m)!(j-m)!(j+m')!(j-m')!} \times \sum_s \frac{(-1)^s}{(j-m'-s)!(j+m-s)!(s+m'-m)!s!} \left(\cos\frac{\beta}{2}\right)^{2j+m-m'-2s} \left(-\sin\frac{\beta}{2}\right)^{m'-m+2s}, \tag{7.151}$$

(middle rotation matrix, explicit form)

[20] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, 1957), p. 52.

where the sum is over all values of s where the factorials are nonnegative. This form is particularly useful for computer implementation of the rotation matrices. We can also see from this formula that under the replacement $\beta \longrightarrow -\beta$, only the sin factor changes sign, so that

$$d^{(j)}_{m'm}(-\beta\hat{y}) = (-1)^{m'-m} d^{(j)}_{m'm}(\beta\hat{y}), \tag{7.152}$$

since the $2s$ part never contributes a minus sign. Furthermore, this formula is invariant under the replacements $m \longrightarrow -m'$ and $m' \longrightarrow -m$,

$$d^{(j)}_{-m,-m'}(\beta\hat{y}) = d^{(j)}_{m'm}(\beta\hat{y}). \tag{7.153}$$

Finally, since the rotation by $-\beta$ is the transpose of the rotation by β (this rotation matrix is orthogonal),

$$d^{(j)}_{m,m'}(-\beta\hat{y}) = d^{(j)}_{m'm}(\beta\hat{y}). \tag{7.154}$$

Combining these last three expressions, we find

$$d^{(j)}_{mm'}(\beta\hat{y}) = (-1)^{m'-m} d^{(j)}_{-m,-m'}(\beta\hat{y}). \tag{7.155}$$

This last expression may be generalized to arbitrary axes. Combining it with Eq. (7.150) gives

$$d^{(j)}_{m'm}(\boldsymbol{\zeta}) = (-1)^{m-m'} d^{(j)*}_{-m',-m}(\boldsymbol{\zeta}), \tag{7.156}$$

(rotation matrix conjugation)

where the complex conjugation "undoes" the minus signs of m and m' in the exponents of the general rotation matrix.

7.3.1.3 Clebsch–Gordan Series

One other useful relation comes by considering the rotation matrix for an arbitrary rotation operator $\mathbf{R}$:

$$d^{(j)}_{m'm} = \langle j\ m'|R|j\ m\rangle. \tag{7.157}$$

If we regard the vector $\mathbf{J}$ associated with j to be the sum $\mathbf{J}_1 + \mathbf{J}_2$, we may write

$$d^{(j)}_{m'm} = \sum_{m_1 m_1' m_2 m_2'} \langle j\ m'|j_1\ m_1', j_2\ m_2'\rangle\langle j_1\ m_1', j_2\ m_2'|R|j_1\ m_1, j_2\ m_2\rangle\langle j_1\ m_1, j_2\ m_2|j\ m\rangle. \tag{7.158}$$

Since the rotation acts on each subspace,

$$d^{(j)}_{m'm} = \sum_{m_1 m_1' m_2 m_2'} \langle j\ m'|j_1\ m_1', j_2\ m_2'\rangle\, d^{(j_1)}_{m_1' m_1} d^{(j_2)}_{m_2' m_2} \langle j_1\ m_1, j_2\ m_2|j\ m\rangle.$$

(Clebsch–Gordan-series inverse) (7.159)

This relation acts as a recursion relation by which rotation matrices can be constructed from other rotation matrices of smaller angular momentum. It is also easy to write down the inverse relation, where we find

$$d^{(j_1)}_{m_1' m_1} d^{(j_2)}_{m_2' m_2} = \sum_{jmm'} \langle j_1\ m_1', j_2\ m_2'|j\ m'\rangle\, d^{(j)}_{m'm} \langle j\ m|j_1\ m_1, j_2\ m_2\rangle.$$

(Clebsch–Gordan series) (7.160)

This relation is called the **Clebsch–Gordan series**.[21] Obviously, the summations in both this relation and its inverse are constrained heavily by the triangularity of the Clebsch–Gordan coefficients.

[21] M. E. Rose, *op. cit.*, Eq. (4.25), p. 58.

7.3.2 Spherical Harmonics

Now we consider the physical-space representation of angular-momentum states $|j\ m\rangle$, in particular the projection into angular states $\langle\theta,\phi|j\ m\rangle$. First, consider what happens under a simple rotation, say about the z-axis. This rotation corresponds to $\beta=\gamma=0$ in the Euler angles above, giving a rotation operator of simply

$$R(\boldsymbol{\alpha})=e^{-i\alpha J_z/\hbar}, \tag{7.161}$$

or a rotation matrix

$$d^{(j)}_{m'm}(\boldsymbol{\alpha})=e^{-im\alpha}. \tag{7.162}$$

Now if $\alpha=2\pi$, then

$$d^{(j)}_{m'm}(\boldsymbol{\alpha})=e^{-im2\pi}, \tag{7.163}$$

but if j is a half-integer, then so is m, and the rotation operator amounts to a factor of -1. On the other hand, if j is an integer, then so is m, and the rotation operator is just the identity. The latter corresponds to what we expect for a vector in coordinate space: a rotation by 2π should amount to nothing. However, this is not the case for half-integer angular momenta, and so we conclude that these do not represent angular momenta of, say, particles (i.e., orbital angular momentum). However, for intrinsic particle spins, half-integer angular momenta are just fine, since we don't require 2π-periodicity in that case. Nonetheless, it seems rather strange that, say a qubit (spin-1/2 particle), under a 2π-rotation, flips its sign; only under a 4π-rotation is it invariant.[22]

Thus, for coordinate representations of angular-momenta, we will only consider the case of integer j. We will thus use the alternate notation for such ''orbital'' angular momenta of $|\ell\ m\rangle$ being a simultaneous eigenstate of $\mathbf{L}^2$ and L_z, with $\mathbf{L}:=\mathbf{r}\times\mathbf{p}$. We can thus define the **spherical harmonic** as the projection onto the usual spherical angles

$$Y^m_\ell(\theta,\phi):=\langle\theta,\phi|\ell\ m\rangle. \tag{7.164}$$
(spherical harmonic)

Later, in Section 8.4.4.1, we show that the spherical harmonics have the form

$$Y^m_\ell(\theta,\phi)=\sqrt{\frac{(2\ell+1)(\ell-m)!}{4\pi(\ell+m)!}}\,P^m_\ell(\cos\theta)\,e^{im\phi}, \tag{7.165}$$
(spherical harmonic)

where $P^m_\ell(\cos\theta)$ is an associated Legendre function, by solving the scalar wave equation in spherical coordinates, which applies to the present case of the Schrödinger equation. Some examples of low-order spherical harmonics are

$$Y^0_0(\theta,\phi)=\frac{1}{\sqrt{4\pi}},\quad Y^0_1(\theta,\phi)=\sqrt{\frac{3}{4\pi}}\,\cos\theta,\quad Y^{\pm1}_1(\theta,\phi)=\mp\sqrt{\frac{3}{8\pi}}\,\sin\theta\,e^{\pm i\phi}, \tag{7.166}$$

corresponding to monopole and dipole angular patterns.

The spherical harmonics are orthonormal, being representations of $|\ell\ m\rangle$. Thus, using $\langle\ell'\ m'|\ell\ m\rangle=\delta_{\ell\ell'}\delta_{mm'}$, we can insert the identity

$$\int d\Omega\,|\theta,\phi\rangle\langle\theta,\phi|=1 \tag{7.167}$$

in terms of angular states to obtain

$$\int d\Omega\,Y^m_\ell(\theta,\phi)\,Y^{m'}_{\ell'}(\theta,\phi)=\delta_{\ell\ell'}\delta_{mm'}, \tag{7.168}$$
(orthonormality relation)

which is simply the expicit statement of orthonormality of the spherical harmonics.

[22] Thus when the qubit is mapped to the sphere, as in the Bloch sphere, you really have to keep track of whether the particle is ''inside'' or ''outside'' the sphere, which is one representation of the minus sign. See, e.g., F. De Zela, ''Topological phase for entangled two-qubit states and the representation of the $SO(3)$ group,'' *Journal of Optics B: Quantum and Semiclassical Optics* **7**, 372 (2005) (doi: 10.1088/1464-4266/7/11/009).

7.3.2.1 Sum Rule and Addition Theorem

Another important relation comes from considering the sum

$$\sum_m \langle\theta_2,\phi_2|\ell\ m\rangle\langle\ell\ m|\theta_1,\phi_1\rangle = \sum_m Y_\ell^{m*}(\theta_1,\phi_1)\,Y_\ell^m(\theta_2,\phi_2), \tag{7.169}$$

for two spherical angles (θ_1,ϕ_1) and (θ_2,ϕ_2). We now intend to show that this expression is independent of orientation (i.e., it is a scalar under rotations) by showing it is equivalent to the rotated version

$$\sum_m \langle\theta_2,\phi_2|R|\ell\ m\rangle\,\langle\ell\ m|R^{-1}|\theta_1,\phi_1\rangle = \sum_m Y_\ell^{m*}(\theta_1',\phi_1')\,Y_\ell^m(\theta_2',\phi_2'), \tag{7.170}$$

for some rotation operator R, where $R|\ell,m\rangle$ is the rotated state. Recall that there are two ways to think of a rotation: the first is that the rotation operator acts on (and rotates) the state vector, while the other is that the rotation operator acts on the basis vectors $|\theta,\phi\rangle$ and rotates the coordinate system in the opposite sense. Thus, the rotated angles (θ_1',ϕ_1') and (θ_2',ϕ_2') are defined by $|\theta_\alpha',\phi_\alpha'\rangle := R^{-1}|\theta_\alpha,\phi_\alpha\rangle$. The rotation operator has a matrix representation that we will denote by $d^{(\ell)}_{m'm}$, which is unitary matrix. Thus, Eq. (7.170) becomes

$$\sum_m Y_\ell^{m*}(\theta_1',\phi_1')\,Y_\ell^m(\theta_2',\phi_2') = \sum_{mm'm''} d^{(\ell)}_{m'm}\,d^{(\ell)*}_{mm''}\langle\theta_2,\phi_2|\ell\ m'\rangle\,\langle\ell\ m''|\theta_1,\phi_1\rangle, \tag{7.171}$$

where we have used the unitarity of the rotation matrix. We can carry out the sum over m by again using the unitarity of the rotation matrix, which we may write as

$$\sum_m d^{(\ell)}_{m'm}\,d^{(\ell)*}_{mm''} = \delta_{m'm''}, \tag{7.172}$$

so that we arrive at

$$\sum_m Y_\ell^{m*}(\theta_1',\phi_1')\,Y_\ell^m(\theta_2',\phi_2') = \sum_m \langle\theta_2,\phi_2|\ell\ m\rangle\,\langle\ell\ m|\theta_1,\phi_1\rangle, \tag{7.173}$$

after dropping primes from the remaining dummy index. Then comparing to Eq. (7.169), we now see the independence of the sum under rotations:

$$\sum_m Y_\ell^{m*}(\theta_1',\phi_1')\,Y_\ell^m(\theta_2',\phi_2') = \sum_m Y_\ell^{m*}(\theta_1,\phi_1)\,Y_\ell^m(\theta_2,\phi_2). \tag{7.174}$$

In particular, we may choose the rotation such that (θ_1',ϕ_1') point along the z-axis, and $\phi_2'=0$. Now we use the fact that $P_l^m(\cos 0) = P_l^m(1) = \delta_{m0}$, and thus from Eq. (7.165), we have

$$Y_\ell^m(0,\phi) = \sqrt{\frac{2l+1}{4\pi}}\delta_{m0}. \tag{7.175}$$

Thus we arrive at the **spherical-harmonic addition theorem**

$$Y_\ell^0(\theta,0) = \sqrt{\frac{4\pi}{2l+1}}\sum_m Y_\ell^{m*}(\theta_1,\phi_1)\,Y_\ell^m(\theta_2,\phi_2), \tag{7.176}$$
(addition theorem)

where $\theta=\theta_2'$ is the angle between the radial vectors corresponding to the two directions (θ_1,ϕ_1) and (θ_2,ϕ_2).

Taking $\theta_1=\theta_2$ and $\phi_1=\phi_2$, so that $\theta=0$ in the addition theorem, we can drop the subscripts and write the sum rule

$$\sum_{m=-\ell}^{\ell} |Y_\ell^m(\theta,\phi)|^2 = \frac{2\ell+1}{4\pi}, \tag{7.177}$$
(sum rule)

where we have again used Eq. (7.175). This sum rule is essentially just another statement of the rotational invariance of products of spherical harmonics when summed over m. This statement indicates indirectly that the m quantum number determines the orientation of the modes; summing over it results in an isotropic angular distribution.

7.3.2.2 Relation to the Rotation Matrix

As in the previous section, when a spherical harmonic

$$\langle\theta,\phi|\ell\ m\rangle = Y_\ell^m(\theta,\phi) \tag{7.178}$$

is rotated, we can express the result in primed coordinates

$$\langle\theta,\phi|R|\ell\ m\rangle = Y_\ell^m(\theta',\phi'), \tag{7.179}$$

and then expressing the rotation operator as a matrix and using the first expression,

$$\sum_{m'} Y_\ell^{m'}(\theta,\phi)\, d^{(\ell)}_{m'm} = Y_\ell^m(\theta',\phi'). \tag{7.180}$$

In general, the rotation matrix can be specified in terms of the Euler angles. Writing this explicitly while omitting the last one (i.e., taking $\gamma = 0$),

$$Y_\ell^m(\theta',\phi') = \sum_{m'} d^{(\ell)}_{m'm}(\alpha,\beta,0)\, Y_\ell^{m'}(\theta,\phi). \tag{7.181}$$

Now we set $\theta \longrightarrow \theta_2$, $\phi \longrightarrow \phi_2$, $m = 0$, $\alpha \longrightarrow \theta_1$, $\beta \longrightarrow \theta_2$, $\theta' \longrightarrow \theta$, and we take the rotation to be such that $\phi' = 0$:

$$Y_\ell^0(\theta,0) = \sum_{m'} d^{(\ell)}_{m'0}(\phi_1,\theta_1,0)\, Y_\ell^{m'}(\theta_2,\phi_2). \tag{7.182}$$

We can now compare this result to the spherical-harmonic sum rule (7.176) and see that they have the same form if we identify

$$Y_\ell^{m*}(\theta,\phi) = \sqrt{\frac{2\ell+1}{4\pi}}\, d^{(\ell)}_{m0}(\phi,\theta,0)$$

(spherical harmonic as rotation matrix) (7.183)

Indeed, in our setup here, θ is still the angle between the vectors along (θ_1,ϕ_1) and (θ_2,ϕ_2). We have chosen the rotation ϕ_1 to bring the vector along (θ_2,ϕ_2) to the x-z plane, and thus with this particular orientation of the problem, $\phi_1 = \phi_2$. The remaining rotation indicated by θ_1 determines the separation angle between θ_1 and θ_2 via $\theta = \theta_2 - \theta_1$. (Recall that the rotation matrices acting on the *coordinate* systems induce the *opposite* rotation as on the states.) In particular, this representation of the spherical harmonics implies the conjugation relation

$$Y_l^{m*}(\theta,\phi) = (-1)^m Y_l^{-m}(\theta,\phi) \tag{7.184}$$

(spherical-harmonic conjugation)

as a direct consequence of Eq. (7.156).

Furthermore, if we use the Clebsch–Gordan series (7.160) with the second indices set to zero,

$$d^{(\ell_1)}_{m_10}\, d^{(\ell_2)}_{m_20} = \sum_{\ell m} \langle \ell_1\ m_1, \ell_2\ m_2|\ell\ m\rangle\, d^{(\ell)}_{m0} \langle \ell\ 0|\ell_1\ 0, \ell_2\ 0\rangle. \tag{7.185}$$

we can then use the representation (7.183) to write

$$Y_{\ell_1}^{m_1}(\theta,\phi)\, Y_{\ell_2}^{m_2}(\theta,\phi) = \sum_{\ell m} \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)}{4\pi(2\ell+1)}} \langle \ell_1\ m_1, \ell_2\ m_2|\ell\ m\rangle\, \langle \ell\ 0|\ell_1\ 0, \ell_2\ 0\rangle\, Y_\ell^m(\theta,\phi)$$

(recoupling relation) (7.186)

after complex conjugation. This is the **recoupling relation** for spherical harmonics. Using Eq. (7.68) to relate the Clebsch–Gordan coefficients to 3-j symbols, we find the alternate form[23]

$$Y_{\ell_1}^{m_1}(\theta,\phi)\, Y_{\ell_2}^{m_2}(\theta,\phi) = \sum_{\ell m} (-1)^m \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ m_1 & m_2 & m \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 0 & 0 & 0 \end{pmatrix} Y_\ell^{-m}(\theta,\phi)$$

(recoupling relation) (7.187)

[23]Note that this relation is referred to as the ''addition theorem'' by A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, 1957), p. 63, Eq. (4.6.5).

after letting $m \longrightarrow -m$. Again, ℓ ranges from $|\ell_1 - \ell_2|$ to $\ell_1 + \ell_2$ and $m_1 + m_2 = -m$ for the 3-j symbols to be nonvanishing.

7.3.3 Irreducible Tensor Operators

7.3.3.1 Spherical Basis

As a prelude to introducing irreducible tensor operators, we will examine the **spherical basis**, which will be important in treating dipole interactions with angular momentum. The spherical basis is simply an alternative to the Cartesian vector basis that is especially convenient when dealing with angular momentum. In terms of the Cartesian basis vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$, the spherical basis vectors are defined as

$$\begin{aligned} \hat{\mathrm{e}}_{\pm 1} &:= \mp \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y}) = -(\hat{\mathrm{e}}_{\mp 1})^* \\ \hat{\mathrm{e}}_0 &:= \hat{z} = (\hat{\mathrm{e}}_0)^*, \end{aligned} \tag{7.188}$$

(spherical basis vectors)

Likewise, if the Cartesian components of a vector $\mathbf{A}$ are defined such that $\mathbf{A} = A_x\hat{x} + A_y\hat{y} + A_z\hat{z}$, then the components of $\mathbf{A}$ in the spherical basis are given in the same way by

$$\begin{aligned} A_{\pm 1} &= \mp \frac{1}{\sqrt{2}}(A_x \pm iA_y) \\ A_0 &= A_z, \end{aligned} \tag{7.189}$$

(vector components in spherical basis)

which is to say $A_q := \hat{\mathrm{e}}_q \cdot \mathbf{A}$, and

$$\mathbf{A} = \sum_q (-1)^q A_q \hat{\mathrm{e}}_{-q} = \sum_q \hat{\mathrm{e}}_q^* A_q. \tag{7.190}$$

(vector in spherical basis)

This funny form for the vector in spherical components comes about because $\hat{\mathrm{e}}_q^* \cdot \hat{\mathrm{e}}_{q'} = \delta_{qq'}$, so we really need to think about products of vectors, where one of them is conjugated (which is, of course, a natural idea in quantum mechanics). Inverting Eqs. (7.189) gives

$$\begin{aligned} A_x &= -\frac{1}{\sqrt{2}}(A_1 - A_{-1}) \\ A_y &= \frac{i}{\sqrt{2}}(A_1 + A_{-1}) \\ A_z &= A_0, \end{aligned} \tag{7.191}$$

with, of course, the same relations for the Cartesian basis vectors in terms of the spherical basis vectors. In the spherical basis, the dot product of two vectors is given by

$$\mathbf{A} \cdot \mathbf{B} = \sum_q (-1)^q A_q \hat{\mathrm{e}}_{-q} \cdot \mathbf{B} = \sum_q (-1)^q A_q B_{-q} = \sum_q A_q (\mathbf{B}^*)_q^*.$$

(dot product in spherical basis) (7.192)

The conjugation in the last expression should be read as: conjugate the full vector $\mathbf{B}$ [by conjugation in Cartesian coordinates or via the conjugate of Eq. (7.190)], then project the qth coordinate via the inner product with $\hat{\mathrm{e}}_q$, so that $(\mathbf{B}^*)_q \equiv \hat{\mathrm{e}}_q \cdot (\mathbf{B}^*)$, and then conjugate the (scalar) result. It follows from Eqs. (7.192) that

$$\mathbf{A}^* \cdot \mathbf{B} = \sum_q (A_q)^* B_q, \qquad |\mathbf{A}|^2 = \sum_q |A_q|^2.$$

(dot product and norm in spherical basis) (7.193)

These expressions are a bit more sensible because $\mathbf{A}^* \cdot \mathbf{B}$ is a more natural inner product in this basis than $\mathbf{A} \cdot \mathbf{B}$. Finally, we note that the components of the position vector $\mathbf{r}$ can be written

$$\begin{aligned} r_{\pm 1} &= \mp \frac{r}{\sqrt{2}} \sin\theta \, e^{\pm i\phi} \\ r_0 &= r\cos\theta, \end{aligned} \qquad (7.194)$$

(position operator in spherical basis)

or more compactly,

$$r_q = r\sqrt{\frac{4\pi}{3}}\, Y_1^q(\theta,\phi). \qquad (7.195)$$

(position operator as spherical harmonic)

These forms will be useful, for example, when evaluating the dipole radiation pattern, and they show explicitly the connection of spherical-basis vector operators to the dipolar spherical harmonics.

7.3.3.2 General Definition

The position vector here (as with any Cartesian three-vector) in the spherical basis is a **vector operator**, because of the way the three components transform among each other under rotations. We will now generalize this notion of sets of operators that are closed under rotations. An **irreducible tensor operator** of rank k (specifically, a spherical tensor operator), which we denote by $\mathbf{T}^{(k)}$ is a set of $2k+1$ operators that transform among themselves under rotations in the same way as the angular-momentum states $|j\ m\rangle$, where $j = k$:

$$R(\boldsymbol{\zeta}) T_q^{(k)} R^\dagger(\boldsymbol{\zeta}) = \sum_{q'=-k}^{k} T_{q'}^{(k)} d_{q'q}^{(k)}(\boldsymbol{\zeta}). \qquad (7.196)$$

(spherical tensor operator)

Equivalently, they transform in the same way under rotations as the spherical harmonics, as in Eq. (7.180). In this context, *irreducible* means that there is no proper subset of the component operators that transform among themselves in a similar way. This is already guaranteed by the definition, as the set of angular-momentum basis states $|j\ m\rangle$ is irreducible in the same sense.

This definition actually introduces **spherical tensors** in general: that is, sets of components that transform into each other. Since again we require that they transform as the spherical harmonics, then the spherical harmonics Y_ℓ^m give a particular example of a spherical tensor of rank ℓ. The tensor *operator* comes about when we take each component of the tensor to be an operator. Since we have already seen that the position operator in the spherical basis is proportional to Y_1^m, as in Eq. (7.195), we know that $\mathbf{r}$ transforms as a spherical tensor of rank 1. Thus, $\mathbf{r}$ is an example of a rank-1 irreducible tensor operator according to the definition here, which is again also a vector operator.

7.3.3.3 Cartesian Tensors

The more familiar type of tensor is the **Cartesian tensor**, of the form $M_{\alpha\beta}$, for example, for a rank-2 tensor, where α and β range from 1 to 3 (or x to z). A rank-k Cartesian tensor is generally represented by k indices, and transforms under rotations according to

$$\tilde{M}_{\alpha_1\alpha_2\cdots\alpha_k} = R_{\alpha_1\beta_1} R_{\alpha_2\beta_2} \cdots R_{\alpha_k\beta_k} M_{\beta_1\beta_2\cdots\beta_k}, \qquad (7.197)$$

where

$$R_{\alpha\beta} = \langle\alpha|R|\beta\rangle \qquad (7.198)$$

is the rotation operator expressed in Cartesian coordinates. That is, the rotation operator is applied to each dimension, represented by each index.

How is the Cartesian tensor related to the irreducible tensors? Well, returning to the rank-2 example, the Cartesian tensor operator has nine independent component operator, whereas the irreducible, rank-2

tensor has only five. The Cartesian tensor must be reducible, and we can reduce it as follows. We may construct a scalar, or rank-0 operator, by computing the trace,

$$M^{(0)} = \mathrm{Tr}[M_{\alpha\beta}] = M_{\alpha\alpha}. \tag{7.199}$$
(scalar part)

This is invariant under rotations, since computing the trace after a rotation gives $\mathrm{Tr}[R\mathbf{M}R^\dagger] = \mathrm{Tr}[\mathbf{M}]$ after cyclic permutation under the trace. We can then form a vector (rank-1) operator as

$$M^{(1)}_\mu = \epsilon_{\mu\alpha\beta}(M_{\alpha\beta} - M_{\beta\alpha}), \tag{7.200}$$
(vector part)

which has three independent components and is clearly related to the antisymmetric part of $M_{\alpha\beta}$. To see that it transforms as a vector under rotations, we can compute the vector after rotation of the tensor, with the result

$$\begin{aligned}\tilde{M}^{(1)}_\mu &= \epsilon_{\mu\alpha\beta}[R_{\alpha\sigma}M_{\sigma\tau}(R^\dagger)_{\tau\beta} - R_{\beta\tau}M_{\tau\sigma}(R^\dagger)_{\sigma\alpha}] \\ &= \epsilon_{\mu\alpha\beta}(R_{\alpha\sigma}R_{\beta\tau}M_{\sigma\tau} - R_{\beta\tau}R_{\alpha\sigma}M_{\tau\sigma}).\end{aligned} \tag{7.201}$$

Now note that the cross product $\mathbf{A}\times\mathbf{B}$ of two vectors, after rotating each vector, is the same as the rotation of the cross product itself, or

$$(\mathbf{RA})\times(\mathbf{RB}) = \mathbf{R}(\mathbf{A}\times\mathbf{B}), \tag{7.202}$$

where $\mathbf{R}$ is the rotation matrix. Expressed in components, this becomes

$$\epsilon_{\mu\alpha\beta}(R_{\alpha\sigma}A_\sigma)(R_{\beta\tau}B_\tau) = R_{\mu\nu}(\epsilon_{\nu\sigma\tau}A_\sigma B_\tau). \tag{7.203}$$

Since this holds for any $\mathbf{A}$ and $\mathbf{B}$, we may drop them and write

$$\epsilon_{\mu\alpha\beta}R_{\alpha\sigma}R_{\beta\tau} = R_{\mu\nu}\epsilon_{\nu\sigma\tau}. \tag{7.204}$$

Putting this into Eq. (7.201),

$$\tilde{M}^{(1)}_\mu = R_{\mu\nu}\epsilon_{\nu\sigma\tau}(M_{\sigma\tau} - M_{\tau\sigma}), \tag{7.205}$$

which is the proper vector rotation of Eq. (7.200). Obviously, this vector operator is still expressed in Cartesian components, but can be transformed to a spherical tensor by Eqs. (7.189).

Finally, the reduced (now irreducible) rank-2 tensor is what remains, or is in other words the original tensor with the trace and antisymmetric parts subtracted away:

$$M^{(2)}_{\alpha\beta} = M_{(\alpha\beta)} - \frac{1}{3}M_{\mu\mu}\delta_{\alpha\beta}, \tag{7.206}$$
(irreducible tensor part)

where $M_{(\alpha\beta)} = (M_{\alpha\beta} + M_{\beta\alpha})/2$ denotes the symmetrized Cartesian tensor. The resulting tensor is clearly symmetric and traceless, and has only 5 independent components, as is consistent with the irreducible rank-2 form. It is also still obviously a rank-2 tensor, since it is a linear combination of $M_{\alpha\beta}$, $M_{\beta\alpha}$, and $\delta_{\alpha\beta}$, which are all rank-2 tensors. However, the transformation of the remaining components to a spherical rank-2 tensor is more complicated than for the vector-operator case.

In any case, we may now write the original tensor in terms of its irreducible components as

$$M_{\alpha\beta} = \frac{1}{3}M^{(0)}\delta_{\alpha\beta} + \frac{1}{4}M^{(1)}_\mu\epsilon_{\mu\alpha\beta} + M^{(2)}_{\alpha\beta}. \tag{7.207}$$
(reduced Cartensian tensor)

We can see this by using Eqs. (7.199), (7.200), and (7.206) to write

$$\begin{aligned}\frac{1}{3}M^{(0)}\delta_{\alpha\beta}+\frac{1}{4}M^{(1)}_{\mu}\epsilon_{\mu\alpha\beta}+M^{(2)}_{\alpha\beta} &= \frac{1}{3}M_{\mu\mu}\delta_{\alpha\beta}+\frac{1}{4}\Big[\epsilon_{\mu\sigma\tau}(M_{\sigma\tau}-M_{\tau\sigma})\Big]\epsilon_{\mu\alpha\beta}+\left[M_{(\alpha\beta)}-\frac{1}{3}M_{\mu\mu}\delta_{\alpha\beta}\right]\\ &= \frac{1}{4}\epsilon_{\mu\alpha\beta}\epsilon_{\mu\sigma\tau}(M_{\sigma\tau}-M_{\tau\sigma})+M_{(\alpha\beta)}\\ &= \frac{1}{4}(\delta_{\alpha\sigma}\delta_{\beta\tau}-\delta_{\alpha\tau}\delta_{\beta\sigma})(M_{\sigma\tau}-M_{\tau\sigma})+M_{(\alpha\beta)}\\ &= \frac{1}{4}\Big[(M_{\alpha\beta}-M_{\beta\alpha})-(M_{\beta\alpha}-M_{\alpha\beta})\Big]+\frac{1}{2}(M_{\alpha\beta}+M_{\beta\alpha})\\ &= M_{\alpha\beta},\end{aligned} \tag{7.208}$$

where we have used the relation

$$\epsilon_{\mu\alpha\beta}\epsilon_{\mu\sigma\tau}=\delta_{\alpha\sigma}\delta_{\beta\tau}-\delta_{\alpha\tau}\delta_{\beta\sigma}, \tag{7.209}$$

which is essentially the "bac-cab" vector identity $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B})$ written in tensor notation (Problem 7.2).

7.3.3.4 Products of Tensors

With Cartesian tensors, taking the product of two tensors to form a higher-rank tensor is straightforward: just multiply them as usual. For example, to take two vectors to form a tensor, we write

$$M_{\alpha\beta}=A_\alpha B_\beta. \tag{7.210}$$
(Cartesian tensor product)

In general, the tensor product of two Cartesian tensors of rank k_1 and k_2 will be of rank $k=k_1+k_2$. However, the case of spherical tensors is a bit more complicated. If we take the addition rules (7.40) for two angular momenta, and then project the first one into angular states $|\theta,\phi\rangle$, we find that the combination of two spherical harmonics is

$$Y^m_\ell(\theta,\phi)=\sum_{\substack{m_1m_2\\(m_1+m_2=m)}} Y^{m_1}_{\ell_1}(\theta,\phi)\,Y^{m_2}_{\ell_2}(\theta,\phi)\,\langle\ell_1\ m_1;\ell_2\ m_2|\ell\ m\rangle, \tag{7.211}$$

where $|\ell_1-\ell_2|\le\ell\le\ell_1+\ell_2$. Spherical harmonics are an example of spherical tensors, and in fact we *defined* spherical tensors to transform in the same way as spherical harmonics. Thus, we conclude that

$$T^{(k)}_q=\sum_{\substack{q_1q_2\\(q_1+q_2=q)}} T^{(k_1)}_{q_1}\,T^{(k_2)}_{q_2}\,\langle k_1\ q_1;k_2\ q_2|k\ q\rangle, \tag{7.212}$$
(spherical tensor product)

where $|k_1-k_2|\le k\le k_1+k_2$. *This* is how products of spherical tensors work: spherical tensors of rank k_1 and k_2 can be combined to form a spherical tensor with a range of different ranks. For example, suppose we want to take a product of two vector operators $\mathbf{A}$ and $\mathbf{B}$. The resulting product tensor $\mathbf{T}^{(k)}$ could have a rank of $k=0,1$, or 2. The rank-0 combination is

$$T^{(0)}_0=\sum_{q=-1}^{1}A_qB_{-q}\,\langle 1\ q;1\ -q|0\ 0\rangle=-\sum_{q=-1}^{1}\frac{(-1)^q}{\sqrt{3}}A_qB_{-q}, \tag{7.213}$$

which we see is the scalar product of the two vectors, up to a constant factor.

$$T^{(0)}_0=-\frac{\mathbf{A}\cdot\mathbf{B}}{\sqrt{3}}, \tag{7.214}$$
(rank-0 vector product)

The rank-1 combination is

$$T^{(1)}_q = \sum_{q'=-1}^{1} A_{q'} B_{q-q'} \, \langle 1 \; q'; 1 \; q-q' | 1 \; q \rangle. \tag{7.215}$$

Writing out the three components of the resulting vector,

$$\begin{aligned} T^{(1)}_1 &= \frac{1}{\sqrt{2}}(A_1 B_0 - A_0 B_1) \\ T^{(1)}_0 &= \frac{1}{\sqrt{2}}(A_1 B_{-1} - A_{-1} B_1) \\ T^{(1)}_{-1} &= \frac{1}{\sqrt{2}}(A_0 B_{-1} - A_{-1} B_0), \end{aligned} \tag{7.216}$$

and then putting in the definitions of the spherical-vector components, we see that the vector product is the usual cross product, expressed in the spherical basis:

$$T^{(1)}_q = \frac{i}{\sqrt{2}}(\mathbf{A} \times \mathbf{B})_q. \tag{7.217}$$
(rank-1 vector product)

Finally, the rank-2 combination is

$$T^{(2)}_q = \sum_{q'=-1}^{1} A_{q'} B_{q-q'} \, \langle 1 \; q'; 1 \; q-q' | 2 \; q \rangle. \tag{7.218}$$

Writing out the resulting five tensor components,

$$\begin{aligned} T^{(2)}_{\pm 2} &= A_{\pm 1} B_{\pm 1} \\ T^{(2)}_{\pm 1} &= \frac{1}{\sqrt{2}}(A_{\pm 1} B_0 + A_0 B_{\pm 1}) \\ T^{(2)}_0 &= \frac{1}{\sqrt{6}}(A_1 B_{-1} + 2 A_0 B_0 + A_{-1} B_1). \end{aligned} \tag{7.219}$$
(rank-2 vector product)

In fact, what we have rederived here is the reduction of the previous section of the rank-2 Cartesian tensor $M_{\alpha\beta} = A_\alpha B_\beta$ into its irreducible parts. Up to constant overall factors, the scalar product (7.214) is the trace (7.199), the cross product (7.217) is the antisymmetric part (7.200) of the tensor, and the rank-2 tensor (7.219) is the traceless, symmetric part (7.206) of the tensor, but here written out in spherical components (which works for arbitrary Cartesian tensors under the identification $M_{\alpha\beta} = A_\alpha B_\beta$).

Finally, we note that with Cartesian tensors, tensor products of lower rank than we have already considered are possible via **contraction**, or making two indices the same and summing over the result. For example, the scalar product of two Cartesian vectors is $A_\alpha B_\alpha$, which is of course lower rank than the tensor product $A_\alpha B_\beta$. The usual matrix product $M_{\alpha\gamma} = A_{\alpha\beta} B_{\beta\gamma}$ is the same idea, giving a rank-2 tensor as the product of two rank-2 tensors, which could give instead a rank-4 tensor without contraction. A scalar can then be obtained by a second contraction, $M_{\alpha\alpha} = A_{\alpha\beta} B_{\beta\alpha}$. In general, the product of a rank-k_1 tensor and a rank-k_2 tensor is of rank $k_1 + k_2$, and this composite rank can be reduced by 2 at a time by contraction. We have already seen how this works for spherical vectors above, and in fact we have also seen that it is possible to reduce the rank by only one, by multiplying by $\epsilon_{\alpha\beta\gamma}$ and then contracting the resulting tensor product (i.e., to give a cross product between vectors). We will simply note here that given two spherical tensors of the same rank k, it is always possible to construct a scalar product. Using Eq. (7.212),

$$T^{(0)}_0 = \sum_q T^{(k)}_q \, U^{(k)}_{-q} \, \langle k \; q; k \; -q | 0 \; 0 \rangle = \sum_q \frac{(-1)^{k+q}}{\sqrt{2k+1}} \, T^{(k)}_q \, U^{(k)}_{-q}, \tag{7.220}$$

after evaluating the Clebsch–Gordan coefficient. Usually we move the invariant factor out of the sum

$$(-1)^{-k}\sqrt{2k+1}\,T_0^{(0)} = \sum_q (-1)^q\, T_q^{(k)}\, U_{-q}^{(k)}, \tag{7.221}$$

and then define the result to be the scalar product of the two spherical tensors:

$$\mathbf{T}^{(k)} \cdot \mathbf{U}^{(k)} := \sum_{q=-k}^{k} (-1)^q\, T_q^{(k)}\, U_{-q}^{(k)}. \tag{7.222}$$

(spherical tensor product)

This extra factor is precisely the factor of $-1/\sqrt{3}$ from Eq. (7.214) beyond the usual scalar product of two vectors. Thus, the definition (7.222) of the scalar product reduces to the usual Cartesian definition for $k = 1$, as we see from Eq. (7.192) in our discussion of spherical vectors.

7.3.3.5 Commutation Rules

Consider the operator for an infinitesimal rotation $\delta\boldsymbol{\phi}$:

$$R(\delta\boldsymbol{\phi}) = 1 - \frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot\mathbf{J}. \tag{7.223}$$

The tensor operator $T_q^{(k)}$ transforms under this rotation as in Eq. (7.196), where the rotation matrix corresponding to the rotation is

$$d_{q'q}^{(k)}(\delta\boldsymbol{\phi}) = \langle k\ q'|\left(1 - \frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot\mathbf{J}\right)|k\ q\rangle. \tag{7.224}$$

Thus, Eq. (7.196) becomes

$$\left(1 - \frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot\mathbf{J}\right) T_q^{(k)} \left(1 + \frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot\mathbf{J}\right) = \sum_{q'} T_{q'}^{(k)} \langle k\ q'|\left(1 - \frac{i}{\hbar}\delta\boldsymbol{\phi}\cdot\mathbf{J}\right)|k\ q\rangle. \tag{7.225}$$

Multiplying this out and dropping second-order terms in $\delta\boldsymbol{\phi}$ gives

$$\left[\delta\boldsymbol{\phi}\cdot\mathbf{J}, T_q^{(k)}\right] = \sum_{q'} T_{q'}^{(k)} \langle k\ q'|\delta\boldsymbol{\phi}\cdot\mathbf{J}|k\ q\rangle. \tag{7.226}$$

Setting $\delta\boldsymbol{\phi} \longrightarrow \hat{z}\delta\phi$ then gives

$$\left[J_z, T_q^{(k)}\right] = \sum_{q'} T_{q'}^{(k)} \langle k\ q'|J_z|k\ q\rangle, \tag{7.227}$$

and using Eq. (7.14) leads to the commutation rule

$$\left[J_z, T_q^{(k)}\right] = \hbar q T_q^{(k)}. \tag{7.228}$$

(J_z commutator)

On the other hand, setting $\delta\boldsymbol{\phi} \longrightarrow [\mp(\hat{x} \pm i\hat{y})/\sqrt{2}]\delta\phi$ gives

$$\left[J_\pm, T_q^{(k)}\right] = \sum_{q'} T_{q'}^{(k)} \langle k\ q'|J_\pm|k\ q\rangle. \tag{7.229}$$

Then using Eq. (7.30) leads to the commutation rule

$$\left[J_\pm, T_q^{(k)}\right] = \sqrt{(k \pm q + 1)(k \mp q)}\, T_q^{(k)}. \tag{7.230}$$

($J_\pm$ commutator)

These commutation rules are the analogous relations to the effects of J_z and $J_\pm$ on kets $|j\ m\rangle$ in Eqs. (7.14) and (7.30).

7.3.4 Wigner–Eckart Theorem

Now we come to an extremely important result in angular momentum algebra. Consider the action of a tensor-operator component on an angular-momentum state,

$$T_q^{(k)}|\alpha'\ j'\ m'\rangle, \tag{7.231}$$

where α' represents other (i.e., radial) quantum numbers that do not represent angular dependence of the state. How does this state transform under a rotation? Since we may write the rotated form as

$$RT_q^{(k)}|\alpha'\ j'\ m'\rangle = (RT_q^{(k)}R^\dagger)\,R|\alpha'\ j'\ m'\rangle, \tag{7.232}$$

evidently $T_q^{(k)}$ and $|\alpha'\ j'\ m'\rangle$ transform separately. In particular, by definition the state $|\alpha'\ j'\ m'\rangle$ transforms as the ket $|j'\ m'\rangle$, while by comparing Eq. (7.196) to Eq. (7.138) we recall that $T_q^{(k)}$ transforms via the rotation matrix in the same way as the angular-momentum ket $|k\ q\rangle$. Thus, the state $T_q^{(k)}|\alpha'\ j'\ m'\rangle$ transforms as the composite state $|k\ q\rangle|j'\ m'\rangle$, or in the way we'll set it up, $|j'\ m'\rangle|k\ q\rangle$ (there is in principle a phase in making this rearrangement, and ignoring it amounts to absorbing it into the definition of the $|\alpha'\rangle$).

We can then consider the usual angular-momentum-addition relation

$$|j'\ m';k\ q\rangle = \sum_{k'q'} |k'\ q'\rangle\langle k'\ q'|j'\ m';k\ q\rangle \tag{7.233}$$

and write in analogy to it the same superposition

$$T_q^{(k)}|\alpha'\ j'\ m'\rangle = \sum_{k'q'} |\tilde{\alpha}\ k'\ q'\rangle\langle k'\ q'|j'\ m';k\ q\rangle, \tag{7.234}$$

where $\tilde{\alpha}$ is some set of transformed radial quantum numbers, since the states in the two relations transform equivalently. This is the crux of the argument: since we know that the tensor operator transforms like angular-momentum states, the action of the tensor operator on an angular-momentum state is just like a mixing of two angular momenta. In fact, the form of the transformed vector is quite constrained: Note that the vector in Eq. (7.234) is an eigenvector of J_z with eigenvalue $\hbar(m'+q)$—although it is a linear combination of vectors of the form $|k'\ q'\rangle$, the Clebsch–Gordan coefficients constrain these such that $q' = m' + q$.

Now we can operate from the left on Eq. (7.234) with $\langle\alpha\ j\ m|$, we then find the matrix element

$$\begin{aligned}\langle\alpha\ j\ m|T_q^{(k)}|\alpha'\ j'\ m'\rangle &= \sum_{k'q'}\langle\alpha\ j\ m|\tilde{\alpha}\ k'\ q'\rangle\,\langle k'\ q'|j'\ m';k\ q\rangle \\ &= \langle\alpha|\tilde{\alpha}\rangle\,\langle j\ m|j'\ m';k\ q\rangle,\end{aligned} \tag{7.235}$$

where we have used the orthogonality of the angular-momentum states to obtain the second equality. Now we note that the inner product $\langle\alpha\ j\ m|\tilde{\alpha}\ j\ m\rangle$ is, in fact, independent of m, just as the inner product $\langle j\ m|j\ m\rangle = 1$ is m-independent. We may thus define the m-independent **reduced matrix element**

$$\langle\alpha\ j\|\mathbf{T}^{(k)}\|\alpha'\ j'\rangle := (-1)^{2k}\langle\alpha|\tilde{\alpha}\rangle, \tag{7.236}$$
(reduced matrix element)

where the dependence on j, j', and $\mathbf{T}^{(k)}$ comes in via the way α transforms into $\tilde{\alpha}$. (The transformation $\alpha \longrightarrow \tilde{\alpha}$ of course introduces no m-dependence because by assumption α represented the radial and thus orientation-independent part of the quantum state.) Note that the reduced matrix element, while using the notation of a tensor, is in fact a scalar quantity, as is clear from the right-hand side of the definition.

Finally, using the reduced matrix element in Eq. (7.235), we arrive at the **Wigner–Eckart theorem**[24]

$$\langle\alpha\ j\ m|T_q^{(k)}|\alpha'\ j'\ m'\rangle = (-1)^{2k}\langle\alpha\ j\|\mathbf{T}^{(k)}\|\alpha'\ j'\rangle\,\langle j\ m|j'\ m';k\ q\rangle.$$
(Wigner–Eckart theorem) (7.237)

[24] Carl Eckart, "The Application of Group Theory to the Quantum Dynamics of Monatomic Systems," *Reviews of Modern Physics* **2**, 305 (1930) (doi: 10.1103/RevModPhys.2.305); Eugene P. Wigner, "Group Theory and Its Application to Quantum Mechanics of Atomic Spectra," (Academic Press, 1959).

Many sign and normalization conventions abound, particularly for the Wigner–Eckart theorem and the reduced matrix elements. By using the orthogonality relation (7.46), we can invert (7.237) to give

$$\langle \alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j' \rangle = (-1)^{2k} \sum_{m' q} \langle \alpha\ j\ m | T_q^{(k)} | \alpha'\ j'\ m' \rangle\ \langle j\ m | j'\ m'; k\ q \rangle.$$

(reduced matrix element) (7.238)

as an expression for the reduced matrix element in terms of a sum over matrix elements. The Wigner–Eckart theorem thus factors a matrix element of a component of an irreducible tensor operator into an orientation-independent part (the reduced matrix element) and a Clebsch–Gordan coefficient (which encapsulates *all* the orientation dependence of the matrix element).

Note that for the reduced matrix elements, we are following here the normalization convention of Brink and Satchler.[25] A common alternate convention for the Wigner–Eckart theorem may be written as[26]

$$\langle \alpha\ j\ m | T_q^{(k)} | \alpha'\ j'\ m' \rangle = \frac{(\alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j')}{\sqrt{2j+1}} \langle j\ m | j'\ m'; k\ q \rangle, \tag{7.239}$$

where the alternate reduced matrix element is related to the first one by

$$(\alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j') = (-1)^{2k} \sqrt{2j+1}\ \langle \alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j' \rangle. \tag{7.240}$$

However, we shall stick exclusively to the matrix element $\langle \alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j' \rangle$. This normalization convention is thus defined by

$$|\langle \alpha\ j \| \mathbf{T}^{(k)} \| \alpha'\ j' \rangle|^2 = \sum_{m' q} |\langle \alpha\ j\ m | T_q^{(k)} | \alpha'\ j'\ m' \rangle|^2 = \sum_{m'} |\langle \alpha\ j\ m | \mathbf{T}^{(k)} | \alpha'\ j'\ m' \rangle|^2,$$

(normalization convention for reduced matrix element) (7.241)

which follows from squaring the Wigner–Eckart theorem (7.237) and then summing over all m' and q, along with the orthogonality relation (7.46) to eliminate the Clebsch–Gordan coefficient.

The Clebsch–Gordan coefficient in Eq. (7.237) requires that j take values between $|k - j'|$ and $k + j'$. In particular, this indicates that $T_q^{(k)}$ can only be of integer rank k. A tensor of half-integer rank would have the awkward consequence of inducing transitions between integer and half-integer states (i.e., between bosonic and fermionic states, which would only be acceptable if a particles are changing or being created/destroyed).

7.3.4.1 Dipole Operator

As we mentioned above, the Wigner–Eckart theorem is so powerful, because it completely pins down the angular part of a matrix element of a tensor operator: the angular dependence of the matrix element can be factored out completely and written solely in terms of a Clebsch–Gordan coefficient (or equivalently, a Wigner 3-j symbol). Of course, this is because of an implicit spherical symmetry to the problem, since we assumed the existence of radial and angular-momentum quantum numbers.

The real utility of the Wigner–Eckart theorem in quantum optics comes from its application to the dipole operator. Recalling that the dipole operator is proportional to the position operator, we know that it transforms as a tensor of rank $k = 1$. Thus, the Wigner–Eckart theorem (7.237) becomes

$$\begin{aligned} \langle J\ m_J | d_q | J'\ m'_J \rangle &= \langle J \| \mathbf{d} \| J' \rangle\ \langle J\ m_J | J'\ m'_J; 1\ q \rangle \\ &= \langle J \| \mathbf{d} \| J' \rangle\ (-1)^{J' - J + m' - m} \sqrt{\frac{2J+1}{2J'+1}} \langle J'\ m' | J\ m; 1\ -q \rangle, \end{aligned}$$

(Wigner–Eckart theorem, dipole operator) (7.242)

[25] D. M. Brink and G. R. Satchler, *Angular Momentum*, 2nd ed. (Oxford, 1968), Section 4.7, p. 56 (ISBN: 0198514190).

[26] See, for example, Gordon Baym, *Lectures on Quantum Mechanics* (Westview Press, 1969); or A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton, 1957).

where the second form follows upon application of the symmetry relation (7.65) followed by an application of (7.54). Again, the orientation dependence of the dipole matrix element appears simply as a Clebsch–Gordan coefficient, while the radial dependence appears in the reduced matrix element. The reduced matrix elements of an atom may, as indicated above, be calculated from the radial parts of the atomic wave functions. However, the simplest way to obtain the reduced matrix element is via the following relation to the atomic spontaneous decay rate from the J_e fine-structure level to the J_g level

$$\Gamma_{J_g J_e} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}\frac{2J_g+1}{2J_e+1}|\langle J_g\|\mathbf{d}\|J_e\rangle|^2,$$

(spontaneous decay rate and reduced dipole matrix element) (7.243)

as we show later in Chapter 11. This relates the reduced matrix element to a quantity readily accessible to experiment.

We will return to the implications of the Wigner–Eckart theorem in detail below in Section 7.3.7.

7.3.4.2 Projection Theorem

A useful special case of the Wigner–Eckart theorem arises if we consider the scalar product of a vector operator $\mathbf{A}$ with angular momentum $\mathbf{J}$.[27] Considering a matrix element of the scalar combination and expanding in the spherical basis using the scalar product (7.192),

$$\begin{aligned}\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle &= \sum_q (-1)^q \langle \alpha\ j\ m|J_q A_{-q}|\alpha'\ j\ m\rangle \\ &= \langle \alpha\ j\ m|\Big(J_0A_0 - J_1A_{-1} - J_{-1}A_1\Big)|\alpha'\ j\ m\rangle\end{aligned} \tag{7.244}$$

Then comparing Eqs. (7.7) and (7.189) to identify the spherical components of $\mathbf{J}$ in terms of the ladder operators,

$$J_{\pm1} = \mp\frac{1}{\sqrt{2}}J_\pm, \tag{7.245}$$

we can use the action (7.30) of the ladder operators on angular-momentum eigenstates (with $J_\pm^\dagger = J_\mp$) to write

$$\begin{aligned}\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle = m\hbar\langle \alpha\ j\ m|A_0|\alpha'\ j\ m\rangle &+ \frac{\hbar}{\sqrt{2}}\sqrt{(j+m)(j-m+1)}\langle \alpha\ j\ m-1|A_{-1}|\alpha'\ j\ m\rangle \\ &- \frac{\hbar}{\sqrt{2}}\sqrt{(j-m)(j+m+1)}\langle \alpha\ j\ m+1|A_{-1}|\alpha'\ j\ m\rangle.\end{aligned} \tag{7.246}$$

Although this result is actually straying a bit from the main point, it shows how the matrix elements here can be expressed solely in terms of matrix elements of $\mathbf{A}$. For our purposes, the more useful result comes from applying the Wigner–Eckart theorem (7.237) to the scalar operator $\mathbf{J}\cdot\mathbf{A}$, for which the matrix element must be orientation-independent (and thus m-independent):

$$\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle =: c_j\langle \alpha\ j\|\mathbf{V}\|\alpha'\ j\rangle\,. \tag{7.247}$$

Here, c_j is again independent of m, and also independent of $\mathbf{A}$ (except that it is a vector operator) and α and α'. Now to compute c_j, we take advantage of the $\mathbf{A}$-independence of this last relation, setting $\mathbf{A}\longrightarrow\mathbf{J}$ and $\alpha'\longrightarrow\alpha$ to obtain

$$c_j\langle \alpha\ j\|\mathbf{J}\|\alpha'\ j\rangle = \langle \alpha\ j\ m|J^2|\alpha'\ j\ m\rangle = \hbar^2 j(j+1). \tag{7.248}$$

Using this relation to eliminate c_j in Eq. (7.247), we have

$$\frac{\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle}{\hbar^2 j(j+1)} = \frac{\langle \alpha\ j\|\mathbf{A}\|\alpha'\ j\rangle}{\langle \alpha\ j\|\mathbf{J}\|\alpha\ j\rangle}. \tag{7.249}$$

[27] Here we are following J. J. Sakurai and Jim Napolitano, *Modern Quantum Mechanics*, 2nd ed. (Cambridge, 2017), pp. 254-5 (ISBN: 9781108422413).

Applying the Wigner–Eckart theorem (7.237) again but now for $k = 1$, the last ratio transforms to

$$\frac{\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle}{\hbar^2 j(j+1)} = \frac{\langle \alpha\ j\ m|A_q|\alpha'\ j\ m'\rangle}{\langle \alpha\ j\ m|J_q|\alpha\ j\ m'\rangle}, \tag{7.250}$$

assuming that the J_q matrix element does not vanish. Rearranging, we come to

$$\langle \alpha\ j\ m|A_q|\alpha'\ j\ m'\rangle = \frac{\langle \alpha\ j\ m|\mathbf{J}\cdot\mathbf{A}|\alpha'\ j\ m\rangle}{\hbar^2 j(j+1)}\langle \alpha\ j\ m|J_q|\alpha\ j\ m'\rangle, \tag{7.251}$$

(projection theorem)

which is known as the **projection theorem**. This is sometimes a handy special case of the Wigner–Eckart theorem where the reduced matrix element has already been eliminated. This will be useful later, for example, when we discuss the Zeeman effect (Section 7.4.1).

7.3.5 Hermitian Conjugates of Tensor Operators

Now we can ask, what is the Hermitian conjugate of an irreducible tensor operator? This is not too hard to establish, given the commutation relations (7.228) and (7.230). First, we can establish the commutator of J_z with the conjugate of $T_q^{(k)}$, using the fact that J_z is Hermitian and Eq. (7.228):

$$\left[J_z, \left(T_q^{(k)}\right)^\dagger\right] = -\left[J_z, T_q^{(k)}\right]^\dagger = -\hbar q \left(T_q^{(k)}\right)^\dagger. \tag{7.252}$$

Similarly, we find the commutator of $J_\pm$ with the conjugate of $T_q^{(k)}$, using the fact that $J_\pm^\dagger = J_\mp$ is Hermitian and Eq. (7.230):

$$\left[J_\pm, \left(T_q^{(k)}\right)^\dagger\right] = -\left[J_\mp, T_q^{(k)}\right]^\dagger = -\sqrt{(k\mp q+1)(k\pm q)}\ \left(T_{q\mp1}^{(k)}\right)^\dagger. \tag{7.253}$$

Notice that if we introduce the operator

$$\tilde{T}_q^{(k)} := (-1)^q \left(T_{-q}^{(k)}\right)^\dagger, \tag{7.254}$$

then the above commutation relations take the form (after letting $q \longrightarrow -q$)

$$\begin{aligned}\left[J_z, \tilde{T}_q^{(k)}\right] &= \hbar q \tilde{T}_q^{(k)} \\ \left[J_\pm, \tilde{T}_q^{(k)}\right] &= \sqrt{(k\pm q+1)(k\mp q)}\ \tilde{T}_{q\pm1}^{(k)}.\end{aligned} \tag{7.255}$$

These are precisely the commutation rules (7.228) and (7.230) for $T_q^{(k)}$. From our derivation of the commutators, we recall that they determine the behavior of the operators under rotations, and since the operators are irreducible we can identify $\tilde{T}_q^{(k)}$ with $T_q^{(k)}$. Thus, the Hermitian conjugate of $T_q^{(k)}$ is (up to an arbitrary phase)

$$\left(T_q^{(k)}\right)^\dagger = (-1)^q\, T_{-q}^{(k)}. \tag{7.256}$$

(tensor operator conjugate)

Evidently, only the $q = 0$ component of a tensor operator is Hermitian.

7.3.5.1 Conjugates of Reduced Matrix Elements

By considering the Wigner–Eckart theorem from both Eqs. (7.237) and (7.242), we can write

$$\begin{aligned}\langle J\ m|T_q^{(k)}|J'\ m'\rangle &= (-1)^{2k}\langle J\|\mathbf{T}^{(k)}\|J'\rangle\ \langle J\ m|J'\ m';1\ q\rangle \\ &= (-1)^{2k}\langle J\|\mathbf{T}^{(k)}\|J'\rangle\ (-1)^{J'-J+m'-m}\sqrt{\frac{2J+1}{2J'+1}}\langle J'\ m'|J\ m;1\ -q\rangle.\end{aligned} \tag{7.257}$$

Clearly the reduced matrix element is not symmetric in J and J', for we may exchange the primed and unprimed numbers and let $q \longrightarrow -q$ to write

$$\langle J'\ m'|T^{(k)}_{-q}|J\ m\rangle = (-1)^{2k}\langle J'\|\mathbf{T}^{(k)}\|J\rangle\,\langle J'\ m'|J\ m;1\ -q\rangle. \tag{7.258}$$

Noting from Eq. (7.256) that $\langle J'\ m'|T^{(k)}_{-q}|J\ m\rangle = (-1)^q\,\langle J\ m|T^{(k)}_q|J'\ m'\rangle^*$, we can compare the above two expressions, using $q = m - m'$, to write the following relation between the reduced matrix elements $\langle J'\|\mathbf{T}^{(k)}\|J\rangle$ and $\langle J\|\mathbf{T}^{(k)}\|J'\rangle$:

$$\langle J'\|\mathbf{T}^{(k)}\|J\rangle = (-1)^{J'-J}\sqrt{\frac{2J+1}{2J'+1}}\,\langle J\|\mathbf{T}^{(k)}\|J'\rangle^*.$$

(reduced matrix element conjugate) (7.259)

Of course, this relation applies as well to reduced matrix elements of the dipole operator, and thus when using reduced matrix elements to compute transition probabilities, it is important to pay attention to the ordering of the J and J' (F and F' for a hyperfine transition) quantum numbers.

7.3.6 Relations Between Reduced Matrix Elements of Tensor Operators

7.3.6.1 Tensor Operator Acting on One Component

Suppose we have a reduced matrix element

$$\langle j\|\mathbf{T}^{(k)}\|j'\rangle \equiv \langle j_1, j_2; j\|\mathbf{T}^{(k)}\|j_1', j_2'; j'\rangle \tag{7.260}$$

between angular-momentum states of the composite angular momentum $\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2$. Suppose further that $\mathbf{T}^{(k)}$ acts only on the states associated with $\mathbf{J}_1$, but not those of $\mathbf{J}_2$. We can then reduce this matrix element to a form in terms of an uncoupled matrix element:

$$\langle j\|\mathbf{T}^{(k)}\|j'\rangle = \delta_{j_2 j_2'}(-1)^{j'+j_1+k+j_2}\sqrt{(2j'+1)(2j_1+1)}\begin{Bmatrix} j_1 & j_1' & k \\ j' & j & j_2 \end{Bmatrix}\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle.$$

(reduced matrix element, single subsystem) (7.261)

Obviously, if $\mathbf{T}^{(k)}$ doesn't couple at all to the $\mathbf{J}_2$ space, the matrix element should only be determined in terms of $\mathbf{J}_1$ matrix elements. Further, this result sensibly says that states of different j_2 are not coupled by this operator.

To prove this result, we start with the expression (7.238) for the reduced matrix element, and then transform into the uncoupled states:

$$\begin{aligned}
\langle j\|\mathbf{T}^{(k)}\|j'\rangle &= (-1)^{2k}\sum_{m'q}\langle j\ m|T^{(k)}_q|j'\ m'\rangle\,\langle j\ m|j'\ m';k\ q\rangle \\
&= (-1)^{2k}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j\ m|j_1\ m_1;j_2\ m_2\rangle\,\langle j_1\ m_1;j_2\ m_2|T^{(k)}_q|j_1'\ m_1';j_2'\ m_2'\rangle\,\langle j_1'\ m_1';j_2'\ m_2'|j'\ m'\rangle \\
&\qquad\qquad \times\langle j\ m|j'\ m';k\ q\rangle \\
&= (-1)^{2k}\delta_{j_2j_2'}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j\ m|j_1\ m_1;j_2\ m_2\rangle\,\langle j_1'\ m_1';j_2'\ m_2'|j'\ m'\rangle\,\langle j\ m|j'\ m';k\ q\rangle \\
&\qquad\qquad \times\langle j_1\ m_1|T^{(k)}_q|j_1'\ m_1'\rangle\,\delta_{m_2m_2'} \\
&= (-1)^{2k}\delta_{j_2j_2'}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j\ m|j_1\ m_1;j_2\ m_2\rangle\,\langle j_1'\ m_1';j_2\ m_2|j'\ m'\rangle\,\langle j\ m|j'\ m';k\ q\rangle \\
&\qquad\qquad \times\langle j_1\ m_1|T^{(k)}_q|j_1'\ m_1'\rangle.
\end{aligned} \tag{7.262}$$

Now applying the Wigner–Eckart theorem (7.237) to the matrix element,

$$\begin{aligned}\langle j\|\mathbf{T}^{(k)}\|j'\rangle &= (-1)^{2k}\delta_{j_2j_2'}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j\ m|j_1\ m_1;j_2\ m_2\rangle\langle j_1'\ m_1';j_2\ m_2|j'\ m'\rangle\langle j\ m|j'\ m';k\ q\rangle\\ &\qquad\times(-1)^{2k}\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle\langle j_1\ m_1|j_1'\ m_1';k\ q\rangle\\ &= \delta_{j_2j_2'}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j_1\ m_1;j_2\ m_2|j\ m\rangle\langle j_1'\ m_1';j_2\ m_2|j'\ m'\rangle\\ &\qquad\times\langle j'\ m';k\ q|j\ m\rangle\langle j_1'\ m_1';k\ q|j_1\ m_1\rangle\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle\\ &= \delta_{j_2j_2'}\sum_{\substack{m'q\\ m_1m_2m_1'm_2'}}\langle j_1\ m_1;j_2\ m_2|j\ m\rangle\langle j_1'\ m_1';j_2\ m_2|j'\ m'\rangle\\ &\qquad\times(-1)^{j'+k-j}(-1)^{j_1'+k-j_1}\langle k\ q;j'\ m'|j\ m\rangle\langle k\ q;j_1'\ m_1'|j_1\ m_1\rangle\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle,\end{aligned}\tag{7.263}$$

where in the last step we used the symmetry relation (7.54) to exchange the first two angular momenta in each of the last two Clebsch–Gordan coefficients. The combination of Clebsch–Gordan coefficients here, if we make the identifications $k\longrightarrow j_1$, $j_1'\longrightarrow j_2$, $j_2\longrightarrow j_3$, $j\longrightarrow j$, $j_1\longrightarrow j_{12}$, and $j'\longrightarrow j_{23}$, has the same form as in the expression (7.86) for the 6-j symbol, and thus

$$\begin{aligned}\langle j\|\mathbf{T}^{(k)}\|j'\rangle &= \delta_{j_2j_2'}(-1)^{j'+k-j}(-1)^{j_1'+k-j_1}(-1)^{-k-j_1'-j_2-j}\sqrt{(2j_1+1)(2j'+1)}\begin{Bmatrix}k & j_1' & j_1\\ j_2 & j & j'\end{Bmatrix}\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle\\ &= \delta_{j_2j_2'}(-1)^{j'-j_1+k-j_2-2j}\sqrt{(2j'+1)(2j_1+1)}\begin{Bmatrix}j_1 & j_1' & k\\ j' & j & j_2\end{Bmatrix}\langle j_1\|\mathbf{T}^{(k)}\|j_1'\rangle,\end{aligned}\tag{7.264}$$

after exchanging the first and last rows of the 6-j symbol. Finally, we use the fact from Eq. (7.90) that j_1+j_2+j is an integer, and thus we can add $2(j_1+j_2+j)$ to the exponent of the (-1), and thus we arrive at the result (7.261).

7.3.6.2 Scalar Products of Tensor Operators

Suppose we have *two* tensor operators, $\mathbf{T}^{(k)}$ and $\mathbf{U}^{(k)}$. We will assume that components of the different tensors commute, $[T_q^{(k)},U_q^{(k)}]=0$, so that the two tensors represent independent systems and thus can support simultaneous eigenstates of each system. However, we can suppose that the two systems are coupled according to the product of the two operators,

$$H_{\text{int}}=\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}=\sum_q(-1)^qT_q^{(k)}U_{-q}^{(k)}.\tag{7.265}$$

Now consider the composite angular momentum $\mathbf{J}=\mathbf{J}_1+\mathbf{J}_2$, where $\mathbf{T}^{(k)}$ is diagonal in the eigenstates $|j_1\ m_1\rangle$ of $\mathbf{J}_1^2$ and J_{1z}, $\mathbf{U}^{(k)}$ is diagonal in the eigenstates $|j_2\ m_2\rangle$ of $\mathbf{J}_2^2$ and J_{2z}, and the interaction H_{int} is diagonal in the coupled eigenstates $|j\ m\rangle$ of $\mathbf{J}^2$ and J_z. We can treat this problem essentially just as in the previous section. But first, if we apply the Wigner–Eckart theorem (7.237), we obtain

$$\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j'\ m'\rangle=\langle j\|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}\|j'\rangle\langle j\ m|j'\ m';0\ 0\rangle=\langle j\|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}\|j'\rangle\,\delta_{jj'}\delta_{mm'}.\tag{7.266}$$

Thus, we need only consider diagonal matrix elements of the scalar product. What we will show is the result

$$\begin{aligned}\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j'\ m'\rangle &= \langle j\|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}\|j'\rangle\,\delta_{mm'}\\ &= (-1)^{j_1+j_2+j}\sqrt{(2j_1+1)(2j_2+1)}\begin{Bmatrix}j & j_2 & j_1\\ k & j_1 & j_2\end{Bmatrix}\langle j_1\|\mathbf{T}^{(k)}\|j_1\rangle\langle j_2\|\mathbf{U}^{(k)}\|j_2\rangle\\ &\qquad\times\delta_{jj'}\delta_{mm'},\end{aligned}$$

(matrix element of scalar product) (7.267)

where the interaction is represented by a product of reduced matrix elements on each subspace and then coupled by a 6-j symbol.

To prove this, we start by taking matrix elements of the interaction in the coupled basis and transforming to the uncoupled basis,

$$\begin{aligned}\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j\ m\rangle &= \sum_q (-1)^q \langle j\ m|T^{(k)}_q U^{(k)}_{-q}|j\ m\rangle \\ &= \sum_{qm_1m_2m_1'm_2'} (-1)^q \langle j\ m|j_1\ m_1; j_2\ m_2\rangle \\ &\qquad \times \langle j_1\ m_1; j_2\ m_2|T^{(k)}_q U^{(k)}_{-q}|j_1\ m_1'; j_2\ m_2'\rangle \langle j_1\ m_1'; j_2\ m_2'|j\ m\rangle \\ &= \sum_{qm_1m_2m_1'm_2'} (-1)^q \langle j_1\ m_1; j_2\ m_2|j\ m\rangle \langle j_1\ m_1'; j_2\ m_2'|j\ m\rangle \\ &\qquad \times \langle j_1\ m_1|T^{(k)}_q|j_1\ m_1'\rangle \langle j_2\ m_2|U^{(k)}_{-q}|j_2\ m_2'\rangle.\end{aligned} \tag{7.268}$$

Applying the Wigner–Eckart theorem (7.237) twice,

$$\begin{aligned}\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j'\ m'\rangle = \sum_{qm_1m_2m_1'm_2'} &(-1)^q \langle j_1\ m_1; j_2\ m_2|j\ m\rangle \langle j_1\ m_1'; j_2\ m_2'|j\ m\rangle \\ &\times \langle j_1\ m_1'; k\ q|j_1\ m_1\rangle \langle j_2\ m_2'; k\ -q|j_2\ m_2\rangle \langle j_1\|\mathbf{T}^{(k)}\|j_1\rangle \langle j_2\|\mathbf{U}^{(k)}\|j_2\rangle.\end{aligned} \tag{7.269}$$

Permuting the symbols in the last Clebsch–Gordan coefficient via (7.65),

$$\begin{aligned}\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j'\ m'\rangle = \sum_{qm_1m_2m_1'm_2'} &(-1)^q(-1)^{k-q} \langle j_1\ m_1; j_2\ m_2|j\ m\rangle \langle j_1\ m_1'; j_2\ m_2'|j\ m\rangle \\ &\times \langle j_1\ m_1'; k\ q|j_1\ m_1\rangle \langle k\ q; j_2\ m_2|j_2\ m_2'\rangle \langle j_1\|\mathbf{T}^{(k)}\|j_1\rangle \langle j_2\|\mathbf{U}^{(k)}\|j_2\rangle.\end{aligned} \tag{7.270}$$

Now again if we identify $j_1 \longrightarrow j_{12}$, $j_2 \longrightarrow j_3$, $j_1' \longrightarrow j_1$, $j_2' \longrightarrow j_{23}$, $j \longrightarrow j$, and $k \longrightarrow j_2$, we can again use Eq. (7.86) for the 6-j symbol, with the result

$$\langle j\ m|\mathbf{T}^{(k)}\cdot\mathbf{U}^{(k)}|j'\ m'\rangle = (-1)^k(-1)^{-(j_1+j_2+j+k)}\sqrt{(2j_1+1)(2j_2+1)} \begin{Bmatrix} j_1 & k & j_1 \\ j_2 & j & j_2 \end{Bmatrix} \langle j_1\|\mathbf{T}^{(k)}\|j_1\rangle \langle j_2\|\mathbf{U}^{(k)}\|j_2\rangle. \tag{7.271}$$

Using the fact that $j_1 + j_2 + j \in \mathbb{Z}$, as required for the 6-$j$ symbol, and permuting the elements of the 6-j symbol as permitted by its symmetries, we obtain the result (7.267).

The general case of the tensor product of two commuting tensor operators is more complicated, as it involves a 9-j symbol.[28] Since we will not use this case, we will avoid it here.

7.3.6.3 Matrix Elements of Tensor Products Operating on the Same System

One last variation on the above theme is to consider a tensor product

$$\mathbf{T}^{(k)} = \mathbf{U}^{(k_1)}\mathbf{V}^{(k_2)}, \tag{7.272}$$

where as usual $|k_1 - k_2| \le k \le k_1 + k_2$, but now *both* component tensors $\mathbf{U}^{(k_1)}$ and $\mathbf{V}^{(k_2)}$ act on the *same* angular-momentum space of $\mathbf{J}$. In this case, we have the reduced matrix element[29]

$$\langle J\|\mathbf{T}^{(k)}\|J'\rangle = (-1)^{k+J+J'}\sum_{J''}\sqrt{(2J''+1)(2k+1)} \begin{Bmatrix} k_1 & k_2 & k \\ J' & J & J'' \end{Bmatrix} \langle J\|\mathbf{U}^{(k_1)}\|J''\rangle \langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle, \tag{7.273}$$

(matrix element of operators on same space)

so that we have a rule for splitting reduced matrix elements of operator products into products of reduced matrix elements.

[28] D. M. Brink and G. R. Satchler, *Angular Momentum*, 2nd ed. (Oxford, 1968), Section 5.3, p. 80.
[29] Brink and Satchler, *op. cit.*, Eq. (5.5).

To prove this, we start with the matrix elements of $\mathbf{T}^{(k)}$, as given by the inverse (7.238) of the Wigner–Eckart theorem:

$$\langle J\|\mathbf{T}^{(k)}\|J'\rangle = (-1)^{2k}\sum_{m'q}\langle J\ m|T^{(k)}_q|J'\ m'_J\rangle\,\langle J\ m_J|J'\ m'_J;k\ q\rangle. \tag{7.274}$$

Now using Eq. (7.272) in the product-component form of Eq. (7.212),

$$T^{(k)}_q = \sum_{\substack{q_1q_2\\(q_1+q_2=q)}} U^{(k_1)}_{q_1}\,V^{(k_2)}_{q_2}\,\langle k_1\ q_1;k_2\ q_2|k\ q\rangle, \tag{7.275}$$

we can replace $T^{(k)}_q$ and introducing the identity to find

$$\begin{aligned}\langle J\|\mathbf{T}^{(k)}\|J'\rangle &= (-1)^{2k}\sum_{m'qq_1q_2}\langle J\ m_J|U^{(k_1)}_{q_1}\,V^{(k_2)}_{q_2}|J'\ m'_J\rangle\,\langle k_1\ q_1;k_2\ q_2|k\ q\rangle\,\langle J\ m_J|J'\ m'_J;k\ q\rangle\\ &= (-1)^{2k}\sum_{\substack{m'qq_1q_2\\J''m''}}\langle J\ m_J|U^{(k_1)}_{q_1}|J''\ m''_J\rangle\,\langle J''\ m''_J|V^{(k_2)}_{q_2}|J'\ m'_J\rangle\,\langle k_1\ q_1;k_2\ q_2|k\ q\rangle\,\langle J\ m_J|J'\ m'_J;k\ q\rangle.\end{aligned} \tag{7.276}$$

Using the Wigner–Eckart theorem (7.237) twice, and introducing the identity to find

$$\begin{aligned}\langle J\|\mathbf{T}^{(k)}\|J'\rangle &= (-1)^{2(k+k_1+k_2)}\sum_{\substack{m'qq_1q_2\\J''m''}}\langle J\|\mathbf{U}^{(k_1)}\|J''\rangle\,\langle J\ m_J|J''\ m''_J;k_1\ q_1\rangle\,\langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle\,\langle J''\ m''_J|J'\ m'_J;k_2\ q_2\rangle\\ &\qquad\times\langle k_1\ q_1;k_2\ q_2|k\ q\rangle\,\langle J\ m_J|J'\ m'_J;k\ q\rangle\\ &= (-1)^{2(k+k_1+k_2)}\sum_{\substack{m'qq_1q_2\\J''m''}}\langle J''\ m''_J;k_1\ q_1|J\ m_J\rangle\,\langle J'\ m'_J;k_2\ q_2|J''\ m''_J\rangle\\ &\qquad\times\langle J'\ m'_J;k\ q|J\ m_J\rangle\,\langle k_1\ q_1;k_2\ q_2|k\ q\rangle\,\langle J\|\mathbf{U}^{(k_1)}\|J''\rangle\,\langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle\\ &= (-1)^{3k+k_1+k_2}\sum_{\substack{m'qq_1q_2\\J''m''}}\langle J''\ m''_J;k_1\ q_1|J\ m_J\rangle\,\langle J'\ m'_J;k_2\ q_2|J''\ m''_J\rangle\\ &\qquad\times\langle J'\ m'_J;k\ q|J\ m_J\rangle\,\langle k_2\ q_2;k_1\ q_1|k\ q\rangle\,\langle J\|\mathbf{U}^{(k_1)}\|J''\rangle\,\langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle,\end{aligned} \tag{7.277}$$

where we used the symmetry rule (7.54) for the last Clebsch–Gordan coefficient. Again identifying $J'\longrightarrow j_1$, $J''\longrightarrow j_{12}$, $J\longrightarrow j$, $k_1\longrightarrow j_3$, $k_2\longrightarrow j_2$, and $k\longrightarrow j_{23}$, we can again use Eq. (7.86) to introduce the 6-j symbol, with the result

$$\begin{aligned}\langle J\|\mathbf{T}^{(k)}\|J'\rangle &= (-1)^{3k+k_1+k_2}(-1)^{-(J'+k_1+k_2+J)}\\ &\qquad\times\sum_{J''}\sqrt{(2J''+1)(2k+1)}\begin{Bmatrix}J' & k_2 & J''\\ k_1 & J & k\end{Bmatrix}\langle J\|\mathbf{U}^{(k_1)}\|J''\rangle\,\langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle\\ &= (-1)^{3k-J-J'}\sum_{J''}\sqrt{(2J''+1)(2k+1)}\begin{Bmatrix}k_1 & k_2 & k\\ J' & J & J''\end{Bmatrix}\langle J\|\mathbf{U}^{(k_1)}\|J''\rangle\,\langle J''\|\mathbf{V}^{(k_2)}\|J'\rangle.\end{aligned} \tag{7.278}$$

Finally, using $(-1)^{4k}=1$ and the fact that $k+J+J'\in\mathbb{Z}$, we arrive at the result (7.273).

7.3.7 Application to Atomic Transitions

7.3.7.1 Decomposition and Calculation of Reduced Matrix Elements

The Wigner–Eckart theorem (7.237) and the decomposition rule (7.261) for reduced matrix elements apply immediately to the matrix elements of the dipole operator that govern atomic electric-dipole transitions. Obviously the dipole operator has rank $k=1$, and so applying the Wigner–Eckart theorem (7.237) to a

transition between two fine-structure sublevels $|J\ m_J\rangle \longrightarrow |J'\ m'_J\rangle$, we find

$$\begin{aligned}\langle J\ m_J|d_q|J'\ m'_J\rangle &= \langle J\|\mathbf{d}\|J'\rangle\,\langle J\ m_J|J'\ m'_J;1\ q\rangle \\ &= \langle J\|\mathbf{d}\|J'\rangle\,(-1)^{J'-J+m'_J-m_J}\sqrt{\frac{2J+1}{2J'+1}}\langle J'\ m'_J|J\ m_J;1\ -q\rangle,\end{aligned}$$

(Wigner–Eckart theorem, fine-structure transition) (7.279)

where to write the last expression, we used the symmetry rules (7.65) and (7.54) to write

$$\langle J'\ m'_J;1\ q|J\ m_J\rangle = (-1)^{1+q}\sqrt{\frac{2J+1}{2J'+1}}\langle 1\ -q;J\ m_J|J'\ m'_J\rangle = (-1)^{J-J'+q}\sqrt{\frac{2J+1}{2J'+1}}\langle J\ m_J;1\ -q|J'\ m'_J\rangle, \tag{7.280}$$

with $(-1)^q = (-1)^{m_J-m'_J} = (-1)^{m'_J-m_J}$ for nonvanishing coefficients. The Wigner–Eckart theorem applies in exactly the same way to a hyperfine transition $|F\ m_F\rangle \longrightarrow |F'\ m'_F\rangle$, so that

$$\begin{aligned}\langle F\ m_F|d_q|F'\ m'_F\rangle &= \langle F\|\mathbf{d}\|F'\rangle\,\langle F\ m_F|F'\ m'_F;1\ q\rangle \\ &= \langle F\|\mathbf{d}\|F'\rangle\,(-1)^{F'-F+m'_F-m_F}\sqrt{\frac{2F+1}{2F'+1}}\langle F'\ m'_F|F\ m_F;1\ -q\rangle.\end{aligned}$$

(Wigner–Eckart theorem, hyperfine transition) (7.281)

In both cases, the Wigner–Eckart theorem gives the rather convenient result that the dependence on the two m levels of the matrix element (which, for example, measures the relative transition rate of the transition) is given entirely by a Clebsch–Gordan coefficient. Stated another way: *the entire angular dependence of the dipole matrix elements is given simply by a Clebsch–Gordan coefficient.* Of course, the dependence on the J or F quantum numbers still appears in what remains in the reduced matrix element.

However, for the reduced matrix element we can make further progress according to Eq. (7.261). The crucial point is that the dipole operator refers to the position of the *electron*. However, a *hyperfine* transition is a coupling between two states corresponding to different $\mathbf{F} = \mathbf{J}+\mathbf{I}$, where in terms of the uncoupled states $|J\ m_J\rangle|I\ m_I\rangle$, the dipole operator acts *only* on the electron angular-momentum state $|J\ m_J\rangle$, not the nuclear state $|I\ m_I\rangle$. Thus, applying the decompostion (7.261) to the reduced hyperfine matrix element,

$$\begin{aligned}\langle F\|\mathbf{d}\|F'\rangle &\equiv \langle J\ I\ F\|\mathbf{d}\|J'\ I\ F'\rangle \\ &= \langle J\|\mathbf{d}\|J'\rangle\,(-1)^{F'+J+1+I}\sqrt{(2F'+1)(2J+1)}\begin{Bmatrix} J & J' & 1 \\ F' & F & I\end{Bmatrix}.\end{aligned}$$

(decomposition of hyperfine reduced matrix element) (7.282)

Note again that since the dipole operator doesn't refer to the nucleus, the nuclear spin I is preserved in the transition. Thus we see that the hyperfine reduced matrix element is just given in terms of the fine-structure (electronic) reduced matrix element, multiplied by a factor that essentially represents the orientation of the electron with respect to the nucleus. Recalling that the 6-j symbol represents a transformation between two different ways to couple three angular momenta, the interpretation in this sense is a bit more murky. However, the basic idea is that one can view the photon as either changing J or F; this amounts to coupling the photon (of unit angular momentum) to the electron either before or after coupling the electron to the nucleus, and thus the appearance of the 6-j symbol.

By exactly the same procedure, the fine-structure reduced matrix element can be further factored into another 6-j symbol and a reduced matrix element involving only the L quantum number:

$$\begin{aligned}\langle J\|\mathbf{d}\|J'\rangle &\equiv \langle L\ S\ J\|\mathbf{d}\|L'\ S\ J'\rangle \\ &= \langle L\|\mathbf{d}\|L'\rangle\,(-1)^{J'+L+1+S}\sqrt{(2J'+1)(2L+1)}\begin{Bmatrix} L & L' & 1 \\ J' & J & S\end{Bmatrix}.\end{aligned}$$

(decomposition of fine-structure reduced matrix element) (7.283)

This, of course, works out because the dipole operator again represents the atomic *position*, but does not refer to its spin. Thus, the dipole operator couples states of different orbital angular momentum L, but doesn't touch the spin S of the electron. Essentially the same interpretations as for the hyperfine case apply here, with the nuclear spin replaced by the electron spin.

7.3.7.2 Fine-Structure Selection Rules

The Wigner–Eckart theorem immediately leads to **selection rules** for ''dipole-allowed'' transitions. That is, for a transition $|J\ m_J\rangle \longrightarrow |J'\ m'_J\rangle$ represented by the matrix element $\langle J\ m_J|d_q|J'\ m'_J\rangle$ to be a non-vanishing matrix element, several conditions are required. In particular, the Clebsch–Gordan coefficient $\langle J\ m_J|J'\ m'_J;1\ q\rangle$, representing the angular dependence of the matrix element according to the Wigner–Eckart theorem, represents the addition of angular momenta $|J'\ m'_J\rangle$ and $|1\ q\rangle$ to form the composite state $|J\ m_J\rangle$. Recall from the triangular condition (7.41) that for such an addition the allowed range of J is bounded below by $|J'-1|$ and above by $J'+1$. This leads to the first selection rule

$$J' = J \qquad \text{or} \qquad J' = J \pm 1. \tag{7.284}$$
(first selection rule)

Next, the addition of angular momentum requires first requires

$$m_J = m'_J + q \tag{7.285}$$
(angular-momentum conservation)

to conserve angular momentum. Since the dipole component index q can take on the values -1, 0, or $+1$ (we will see these correspond to interactions with different polarizations of the electromagnetic field), the second selection rule becomes

$$m'_J = m_J \qquad \text{or} \qquad m'_J = m_J \pm 1. \tag{7.286}$$
(second selection rule)

Finally, one can show from Eq. (7.67) that the Clebsch–Gordan coefficient $\langle J\ 0|J\ 0;1\ q\rangle$ vanishes for any J and q, leading to the final selection rule

$$J' \neq J \qquad \text{if} \qquad m'_J = m_J = 0. \tag{7.287}$$
(third selection rule)

In particular, $J = J' = 0$ represents a forbidden transition; intuitively, this is because the atom must ''absorb'' the angular momentum of the photon (a spin-1 particle), which is not possible if $J = 0$ both before and after absorbing the photon.

In all other cases, the corresponding dipole matrix element vanishes, or in other words the transition is **dipole forbidden**.

As a final note, the dipole interaction couples the electric field to the electron's *position* and not to its spin, and thus the electron spin S (and m_S) should not change in an electric-dipole transition.

$$S' = S, \qquad m'_S = m_S. \tag{7.288}$$
(electric-dipole spin selection rules)

Furthermore, we can note that the above selection rules for J also apply to the orbital angular momentum L, so we may write

$$\begin{aligned} &L' = L \quad \text{or} \quad L' = L \pm 1 \\ &L \neq 0 \quad \text{or} \quad L' \neq 0. \end{aligned} \tag{7.289}$$
(orbital selection rules)

In particular, from the second rule any fine-structure transition of the form $n\mathrm{S}_{1/2} \longrightarrow n'\mathrm{S}_{1/2}$ is dipole forbidden. Of course, analogous rules should hold for m_L, but this often isn't a useful quantum number, so we'll skip it.

We recall also that the dipole operator only couples states of opposite parity (Section 5.1.1). A hydrogen-atom state $|nlm\rangle$ has an angular dependence given by the spherical harmonic $Y_\ell^m(\theta,\phi)$. But under a parity transformation

$$Y_\ell^m(\pi-\theta,\phi+\pi) = (-1)^\ell Y_\ell^m(\theta,\phi), \tag{7.290}$$

so evidently

$$\Delta\ell = \pm 1 \tag{7.291}$$
(single-electron orbital selection rules)

for exactly one electron in the atom. (Recall that l is the orbital quantum number for a single electron, while L is the combined orbital quantum number for all the electrons.) This selection rule is often interpreted as conservation of angular momentum when a photon is absorbed or emitted. Thus for a single electron atom, or in two-electron atoms in low-energy transitions where only one electron is active, this rule implies that $\Delta L = \pm 1$ as well.

One final selection rule is a bit more subtle, and is specific to fine-structure transitions where $S = 1/2$. Because $\mathbf{J} = \mathbf{L} + \mathbf{S}$, we have

$$|L - 1/2| \leq J \leq L + 1/2, \tag{7.292}$$

and since $\mathbf{J}' = \mathbf{L}' + \mathbf{S}$, we have

$$|L' - 1/2| \leq J' \leq L' + 1/2. \tag{7.293}$$

Now consider a transition where $L' = L + 1$ but $J' = J - 1$. Then the second condition becomes

$$L + 3/2 \leq J \leq L + 5/2, \tag{7.294}$$

but this contradicts $J \leq L + 1/2$ from Eq. (7.292). Similarly, if $L' = L - 1$ but $J' = J + 1$, then (7.293) becomes

$$|L - 3/2| - 1 \leq J \leq L - 3/2, \tag{7.295}$$

but this contradicts $J \geq L - 1/2$ from Eq. (7.292). Thus we may write the fourth fine-structure selection rule

$$\text{if} \quad L' = L \pm 1 \quad \text{then} \quad J' \neq J \mp 1. \tag{7.296}$$

(fourth selection rule)

This argument assumed $S = 1/2$, and so will not in general carry over to, for example, the hyperfine transitions we consider later.

In any case, the selection rules for l, L, and S are approximate, since they assume that these are good quantum numbers. In heavy, many-electron atoms, this may not be the case, and so these rules may be violated to some extent.[30]

7.3.7.3 Hyperfine Selection Rules

Just as in Section 7.3.7.2, the constraints on the Clebsch–Gordan coefficient in the hyperfine Wigner–Eckart theorem (7.281) induce selection rules on the hyperfine quantum numbers. Of course, the selection rules on J and m_J of Section 7.3.7.2 still apply, but because the involved Clebsch–Gordan coefficients are the same, the same constraints apply to F and m_F. In particular, the rules

$$\begin{array}{ccc} F' = F & \text{or} & F' = F \pm 1 \\ m'_F = m_F & \text{or} & m'_F = m_F \pm 1 \\ F' \neq F & \text{if} & m'_F = m_F = 0 \end{array} \tag{7.297}$$

(hyperfine selection rules)

apply to the hyperfine transition $|F\ m_F\rangle \longrightarrow |F'\ m'_F\rangle$.

Finally, since the dipole interaction couples only the field to the electron dipole, there is no nuclear coupling and thus the nuclear spin I should not change in an electric-dipole transition (nor should m_I, if that is the preferred basis). Thus, we may write the extra selection rules

$$I' = I, \qquad m'_I = m_I, \tag{7.298}$$

(electric-dipole nuclear-spin selection rules)

Of course, the nuclear spin may effectively change *relative to the electron angular momentum*, which is why it is meaningful to have hyperfine states and, say, to optically pump into particular hyperfine sublevels.

[30] See Alan Corney, *Atomic and Laser Spectroscopy* (Oxford, 1977), Chapters 5-7.

7.3.7.4 Decay Rate and the Reduced Matrix Element

But now the question remains, how do we compute the dipole matrix elements? The basic answer is to know the decay rate Γ (equivalently, the lifetime) of the excited level, and here we will relate the decay rate to the reduced matrix elements. For the two level atom, the spontaneous emission rate from the quantum treatment of the atom–field interaction, Eq. (11.29), from $|\mathrm{e}\rangle \longrightarrow |\mathrm{g}\rangle$ (with transition frequency ω_0) is

$$\Gamma = \frac{\omega_0^3 |\langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle|^2}{3\pi\epsilon_0\hbar c^3}. \tag{7.299}$$

This result is only for two levels, but now we are confronted with the physically important case of decay between levels with angular-momentum degeneracy.

Consider the decay of the $J_\mathrm{g} \longrightarrow J_\mathrm{e}$ fine-structure transition (with J_e being the excited state as usual). Then the decay rate from sublevel $|J_\mathrm{e}\ m_\mathrm{e}\rangle \longrightarrow |J_\mathrm{g}\ m_\mathrm{g}\rangle$ is just given by Eq. (7.299) with the appropriate matrix element:

$$\Gamma_{J_\mathrm{g},m_\mathrm{g};J_\mathrm{e},m_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3} |\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2. \tag{7.300}$$

The quantum vacuum is isotropic, and so spontaneous emission is as well. Thus, all magnetic sublevels of the excited manifold decay at the same rate, and we can drop the explicit dependence on m_e. The total decay rate out of any excited sublevel (and thus of the whole excited manifold) is this decay rate summed over all ground sublevels:

$$\Gamma_{J_\mathrm{g}J_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3} \sum_{m_\mathrm{g}} |\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2. \tag{7.301}$$

Summing over the excited sublevels as well, we can write

$$\Gamma_{J_\mathrm{g}J_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3} \frac{1}{2J_\mathrm{e}+1} \sum_{m_\mathrm{g},m_\mathrm{e}} |\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2. \tag{7.302}$$

Recalling the normalization convention (7.241) for the reduced dipole matrix element,

$$\sum_{m_\mathrm{e}} |\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2 = \sum_{m_\mathrm{e}q} |\langle J_\mathrm{g}\ m_\mathrm{g}|d_q|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2 = |\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2\,, \tag{7.303}$$

we can eliminate the excited-state sum:

$$\Gamma_{J_\mathrm{g}J_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3} \frac{1}{2J_\mathrm{e}+1} \sum_{m_\mathrm{g}} |\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2. \tag{7.304}$$

The summand no longer depends on m_g, so we arrive at the final result:

$$\Gamma_{J_\mathrm{g}J_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3} \frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1} |\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2.$$

(spontaneous decay rate, fine-structure transition) (7.305)

This expression can also effectively define the normalization convention of the matrix elements. More importantly, *this* is generally how you compute the dipole matrix elements, since lifetimes are a commonly measured quantity used to define the strength of transitions (and even better, the lifetimes are free of ambiguities from normalization conventions!).

A similar argument applies for hyperfine levels. However, when summing the dipole matrix elements over all ground sublevels $|F_\mathrm{g}\ m_\mathrm{g}\rangle$ connected to an excited sublevel, the explicit dependence on all hyperfine quantum numbers vanishes:

$$\sum_{F_\mathrm{g}m_\mathrm{g}} |\langle F_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|F_\mathrm{e}\ m_\mathrm{e}\rangle|^2 = \frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1} |\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2\,. \tag{7.306}$$

Thus, the only dependence of the decay rate on the hyperfine quantum numbers enters through differences in the transition frequencies. Even for the larger alkali atoms, this amounts to a difference of less than 0.1% in the decay rate, too small to be resolved so far in lifetime measurments. We can therefore ignore the hyperfine splittings and use Eq. (7.305) as the decay rate for hyperfine levels.

Of course, these arguments establish the *magnitude* of the reduced matrix elements, but not the *phase*. By convention, all dipole matrix elements are real, but may be positive or negative. For the interaction of a field with only a single fine-structure transition, the reduced matrix element $\langle J\|\mathbf{d}\|J'\rangle$ can be assigned an arbitrary sign, say positive. The correct signs for the hyperfine matrix elements are then assured by the decomposition formula (7.282). If the interactions with multiple fine-structure levels are important, the relative signs of the different matrix elements must be set according to the further decomposition rule (7.283).

7.4 Interaction with Static Fields

7.4.1 Static Magnetic Fields: Zeeman Effect

Each of the fine-structure (J) energy levels contains $2J+1$ magnetic sublevels that determine the angular distribution of the electron wave function. These states are labeled by the quantum numbers m_J associated with the J_z operator, satisfying $-J \le m_J \le J$. In the absence of external fields, these sublevels are degenerate: as we saw above, the shift only depended on the J quantum number, not m_J.

However, when an external magnetic field is applied, their degeneracy is broken. Both the magnetic moments due to the electron spin and orbit couple to the field, and the Hamiltonian describing the atomic interaction with the magnetic field is simply the magnetic-dipole-interaction Hamiltonian

$$\begin{aligned} H_B^{(\mathrm{fs})} &= -\boldsymbol{\mu}_S \cdot \mathbf{B} - \boldsymbol{\mu}_L \cdot \mathbf{B} \\ &= \frac{\mu_\mathrm{B}}{\hbar}(g_S \mathbf{S} + g_L \mathbf{L}) \cdot \mathbf{B} \\ &= \frac{\mu_\mathrm{B}}{\hbar}(g_S S_z + g_L L_z) B, \end{aligned} \tag{7.307}$$

if we take the magnetic field to be along the z-direction (i.e., along the atomic quantization axis), $\mathbf{B} = B\hat{z}$. Again, the quantities g_S and g_L are respectively the electron spin and orbital "g-factors" or fudge-factors that account for various modifications to the corresponding magnetic dipole moments. We already saw g_S in Section 7.2.1; the value for g_L is approximately 1, as we assumed in the same section, but to account for the finite nuclear mass, we can take to lowest order

$$g_L = \frac{m}{m_\mathrm{e}} = \frac{1}{1 + m_\mathrm{e}/m_\mathrm{nuc}} \approx 1 - \frac{m_\mathrm{e}}{m_\mathrm{nuc}}, \tag{7.308}$$

where again m_e is the electron mass and m_nuc is the nuclear mass to account for the fact that the electron's mass in the expression (7.109) for the orbital angular momentum is really its reduced mass, while the Bohr magneton $\mu_\mathrm{B} = e\hbar/2m_\mathrm{e}$ instead uses the electron mass.

7.4.1.1 Anomalous Zeeman Effect: Weak Fields

If we assume the magnetic-field interaction to be a small perturbation to the fine-structure Hamiltonian, then to first order in perturbation theory we ignore any mixing of the fine-structure states $|J\,m_J\rangle$ and the energy shift is

$$\begin{aligned} \Delta E_\mathrm{B}^{(\mathrm{fs})}(J, m_J) &= \langle J\,m_J|H_B^{(\mathrm{fs})}|J\,m_J\rangle \\ &= \frac{\mu_\mathrm{B} B}{\hbar}\langle J\,m_J|(g_S S_z + g_L L_z)|J\,m_J\rangle \\ &= \frac{\mu_\mathrm{B} B g_L}{\hbar}\langle J\,m_J|J_z|J\,m_J\rangle + \frac{\mu_\mathrm{B} B(g_S - g_L)}{\hbar}\langle J\,m_J|S_z|J\,m_J\rangle \\ &= \mu_\mathrm{B} B g_L m_J + \frac{\mu_\mathrm{B} B(g_S - g_L)}{\hbar}\langle J\,m_J|S_z|J\,m_J\rangle, \end{aligned} \tag{7.309}$$

where we used $J_z = L_z + S_z$ and of course $J_z|J\,m_J\rangle = m_J\hbar|J\,m_J\rangle$. To evaluate the second term, we first note that since $\mathbf{J} = \mathbf{L} + \mathbf{S}$, squaring this relation gives

$$\mathbf{S}\cdot\mathbf{J} = \frac{1}{2}\left(\mathbf{J}^2 + \mathbf{S}^2 - \mathbf{L}^2\right). \tag{7.310}$$

Using the projection theorem (7.251), we can write

$$\begin{aligned}\langle J\,m_J|S_z|J\,m_J\rangle &= \frac{m_J}{J(J+1)\hbar}\langle J\,m_J|\mathbf{S}\cdot\mathbf{J}|J\,m_J\rangle \\ &= \frac{m_J}{2J(J+1)\hbar}\langle J\,m_J|\left(\mathbf{J}^2 + \mathbf{S}^2 - \mathbf{L}^2\right)|J\,m_J\rangle \\ &= \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}m_J\hbar,\end{aligned} \tag{7.311}$$

where we used Eq. (7.310) for the dot product, and we recall that $|J\,m_J\rangle \equiv |L\,S\,J\,m_J\rangle$. Putting this expectation value into Eq. (7.309), we obtain the perturbative shift

$$\Delta E_{\text{B}}^{(\text{fs})} = \mu_{\text{B}} g_J m_J B, \tag{7.312}$$

(fine-structure Zeeman shift, small B)

where the **Landé g_J factor**[31] is

$$g_J := g_L + (g_S - g_L)\frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \tag{7.313}$$

(Landé g_J factor)

Note that this expression does not include corrections due to multielectron[32] and QED[33] effects, and thus measured values may differ slightly from this. Note also that since $g_L \approx 1$ and $g_S \approx 2$, the g_J factor is commonly written as

$$g_J \approx 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)}. \tag{7.314}$$

The shift in this regime is thus proportional to both the m_J quantum number and the magnetic field. Recall that this is *only* valid when the magnetic field is along the $\hat{z}$ direction; otherwise, you must compute the energy shifts according to this method for the states quantized along the magnetic field, and then use the rotation operators below to obtain the states with the desired quantization axis, which will no longer be eigenstates of the system. Thus, with other quantization axes, a given state will "remix" with the others due to precession along the magnetic-field axis.

For the hyperfine case, if the energy shift due to the magnetic field is small compared to the fine-structure splitting, then as we just argued J is a good quantum number. Then the interaction Hamiltonian can be written as the fine-structure interaction plus the magnetic-dipole interaction of the *nuclear* magnetic moment with the magnetic field:

$$H_B^{(\text{hfs})} = H_B^{(\text{fs})} - \boldsymbol{\mu}_I\cdot\mathbf{B} = \frac{\mu_{\text{B}}}{\hbar}(g_J J_z + g_I I_z)B_z. \tag{7.315}$$

Again, if the energy shift due to the magnetic field is small compared to the hyperfine splittings, then the fine-structure treatment carries through with $J \longrightarrow F$, $S \longrightarrow I$, and $L \longrightarrow J$, so that

$$\Delta E_{\text{B}}^{(\text{hfs})} = \mu_{\text{B}} g_F m_F B, \tag{7.316}$$

(hyperfine Zeeman shift, small B)

[31] S. Goudsmit, "Nuclear Magnetic Moments," *Physical Review* **43**, 636 (1933) (doi: 10.1103/PhysRev.43.636); Alfred Landé, "The Magnetic Moment of the Proton," *Physical Review* **44**, 1028 (1933) (doi: 10.1103/PhysRev.44.1028); Alfred Landé, "Nuclear Magnetic Moments and Their Origin," *Physical Review* **46**, 477 (1934) (doi: 10.1103/PhysRev.46.477); Paul Forman, "Alfred Landé and the anomalous Zeeman Effect, 1919-1921," *Historical Studies in the Physical Sciences* **2**, 153 (1970).

[32] Hans A. Bethe and Edwin E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer-Verlag, 1957).

[33] Leonti Labzowsky, Igor Goidenko, and Pekka Pyykkö, "Estimates of the bound-state QED contributions to the g-factor of valence ns electrons in alkali metal atoms," *Physics Letters A* **258**, 31 (1999).

where the g_F factor is

$$g_F := g_J + (g_I - g_J)\frac{F(F+1) + I(I+1) - J(J+1)}{2F(F+1)}, \tag{7.317}$$

or in a more symmetric form,

$$g_F := g_J \frac{F(F+1) - I(I+1) + J(J+1)}{2F(F+1)} + g_I \frac{F(F+1) + I(I+1) - J(J+1)}{2F(F+1)}. \tag{7.318}$$

(Landé g_F factor)

Recalling that g_I is much smaller than g_J, this is commonly written

$$g_F \approx g_J \frac{F(F+1) - I(I+1) + J(J+1)}{2F(F+1)}, \tag{7.319}$$

which is correct at the 0.1% level.

The shifts proportional to the magnetic fields in this weak-field regime were historically referred to as the **anomalous Zeeman effect**, after Zeeman's observation of the splitting of spectral lines.[34] The "normal" case was based on the predicted splittings due only to orbital angular momentum. However, "anomalous" cases—which included spin effects—were observed before spin was known.

7.4.1.2 Paschen–Back Effect: Strong Fields

In the case of fine structure, when the field is large enough that the shifts due to the magnetic-field interaction Hamiltonian (7.307) dominate the fine-structure splittings, the interaction is again simple, but J is no longer a good quantum number. Ignoring the fine-structure Hamiltonian, the eigenstates of the interaction are the uncoupled, or "high-field" fine-structure states $|L\, m_L; S\, m_S\rangle$. The energy shifts are thus given by the expectation value of the interaction Hamiltonian, or

$$\Delta E_{|L\, m_L; S\, m_S\rangle} = \left\langle H_B^{(\mathrm{fs})} \right\rangle = \frac{\mu_\mathrm{B}}{\hbar} \langle (g_S S_z + g_L L_z) \rangle\, B = \mu_\mathrm{B}(g_S m_S + g_L m_L) B. \tag{7.320}$$

The energies again shift linearly with the applied field amplitude, but now the shifts have contributions proportional to m_S and m_L rather than simply being proportional to m_J. The shift in this large-field regime is called the **Paschen–Back effect**.[35]

We will consider in a bit more detail the hyperfine case, since it is easier to enter the Paschen–Back regime for the much smaller hyperfine splittings. For strong fields where the appropriate interaction is described by Eq. (7.315), the interaction Hamiltonian dominates the hyperfine Hamiltonian (7.132), so that the hyperfine Hamiltonian perturbs the strong-field eigenstates $|J\, m_J; I\, m_I\rangle$. For this treatment to hold, the energy perturbations must still be small compared to the fine-structure splitting, otherwise we need to account for that effect as well. We can compute the energies to first order in perturbation theory (lowest order in $1/B$) by computing the expectation value

$$E_{|J\, m_J; I\, m_I\rangle} = \left\langle H_\mathrm{hfs} + H_B^{(\mathrm{hfs})} \right\rangle \tag{7.321}$$

with respect to the strong-field states $|J\, m_J\, I\, m_I\rangle$. To do this, we first invert the defining relations (7.7) for the ladder operators to find $J_x = (J_+ + J_-)/2$ and $J_y = (J_+ - J_-)/2i$, so that

$$\begin{aligned} \mathbf{I}\cdot\mathbf{J} &= I_z J_z + I_x J_x + I_y J_y \\ &= I_z J_z + \frac{(I_+ + I_-)(J_+ + J_-)}{4} - \frac{(I_+ - I_-)(J_+ - J_-)}{4} \\ &= I_z J_z + \frac{I_+ J_- + I_- J_+}{2}. \end{aligned} \tag{7.322}$$

[34] P. Zeeman, "The Effect of Magnetisation on the Nature of Light Emitted by a Substance," *Nature* **55**, 347 (1897) (doi: 10.1038/055347a0); P. Zeeman, "On the influence of Magnetism on the Nature of the Light emitted by a Substance," *Philosophical Magazine* **43**, 226 (1897); P. Zeeman, "Doubles and triplets in the spectrum produced by external magnetic forces," *Philosophical Magazine* **44** 55 (1897).

[35] Named for F. Paschen and E. Back, "Normale und anomale Zeemaneffekte," *Annalen der Physik* **344**, 897 (1912) (doi: 10.1002/andp.19123441502).

In the expectation value, the second term vanishes, leaving

$$\langle J\, m_J; I\, m_I|(\mathbf{I}\cdot\mathbf{J})|J\, m_J; I\, m_I\rangle = m_I m_J \hbar^2. \tag{7.323}$$

Squaring Eq. (7.322) then gives

$$\begin{aligned}(\mathbf{I}\cdot\mathbf{J})^2 &= (I_z J_z)^2 + \frac{1}{2}[(I_z J_z),(I_+ J_- + I_- J_+)]_+ + \frac{(I_+ J_- + I_- J_+)^2}{4} \\ &= (I_z J_z)^2 + \frac{1}{2}[(I_z J_z),(I_+ J_- + I_- J_+)]_+ + \frac{(I_+ J_-)^2 + (I_- J_+)^2}{4} + \frac{I_+ I_- J_- J_+ + I_- I_+ J_+ J_-}{4},\end{aligned} \tag{7.324}$$

and then computing the expectation value, the middle terms vanish, while the last term can be computed from Eq. (7.27) to obtain[36]

$$\begin{aligned}\langle J\, m_J; I\, m_I|(\mathbf{I}\cdot\mathbf{J})^2|J\, m_J; I\, m_I\rangle &= \hbar^4 (m_I m_J)^2 + \frac{\hbar^4}{4}[I(I+1) - m_I(m_I-1)][J(J+1) - m_J(m_J+1)] \\ &\quad + \frac{\hbar^4}{4}[I(I+1) - m_I(m_I+1)][J(J+1) - m_J(m_J-1)] \\ &= \hbar^4 (m_I m_J)^2 + \frac{\hbar^4}{2}[I(I+1) - m_I^2][J(J+1) - m_J^2] - \frac{\hbar^4}{2} m_I m_J.\end{aligned} \tag{7.325}$$

Now we can evaluate Eq. (7.321) by putting the above expectation values into (7.132), while dropping the small C_{hfs} term for (relative) simplicity, with the result

$$\begin{aligned}E_{|J\, m_J; I\, m_I\rangle} \approx A_{\text{hfs}} m_I m_J + B_{\text{hfs}} &\frac{9(m_I m_J)^2 - 3J(J+1)m_I^2 - 3I(I+1)m_J^2 + I(I+1)J(J+1)}{4J(2J-1)I(2I-1)} \\ &+ \mu_{\text{B}}(g_J\, m_J + g_I\, m_I)B.\end{aligned}$$

(hyperfine Paschen–Back effect) (7.326)

The expectation of the interaction Hamiltonian is trivial in this case. Clearly, the hyperfine Hamiltonian, while not contributing a B-dependent energy, is important in determining the correct splittings between the states, even for strong fields. The energy shift in this regime is called the **hyperfine Paschen-Back effect**.

Note that for both instances of the Paschen–Back effect, we are considering fields that cause large shifts on the scale of the unperturbed splittings. However, we don't want the fields to be *too* large, where for example the diamagnetic interaction, which in the dipole approximation is $H_{\text{diamagnetic}} = [\mathbf{d}\times\mathbf{B}(0)]^2/8m_{\text{e}}$, as we recall from Eq. (9.119), which leads to a **quadratic Zeeman effect**.

7.4.1.3 Incomplete Paschen–Back Effect: Intermediate Fields

For intermediate fields, where for example in hyperfine structure the interaction Hamiltonian neither weakly perturbs nor dominates the hyperfine Hamiltonian, the energy shift is more difficult to calculate, and in general one must numerically diagonalize $H_{\text{hfs}} + H_B^{(\text{hfs})}$. A notable exception comes about in hyperfine structure when either $J = 1/2$ or $I = 1/2$. In this case, the hyperfine Hamiltonian (7.132) is given only by the magnetic-dipole term,

$$H_{\text{hfs}} = A_{\text{hfs}} \frac{\mathbf{I}\cdot\mathbf{J}}{\hbar^2}. \tag{7.327}$$

In the strong-field basis, we have from Eq. (7.322) again the diagonal matrix elements

$$\langle J\, m_J; I\, m_I|H_{\text{hfs}}|J\, m_J; I\, m_I\rangle = A_{\text{hfs}} m_I m_J \tag{7.328}$$

[36] This expression differs from that of E. B. Alexandrov, M. P. Chaika, and G. I. Khvostenko, *Interference of Atomic States* (Springer–Verlag, 1993), p. 222, Eq. (5.160), where the authors neglected the contributions from the last term in Eq. (7.324). Comparison of the two expressions to energies from numerical diagonalization confirms that the expression shown here is a better approximation.

and also from Eq. (7.30) the off-diagonal matrix elements

$$\begin{aligned}\langle J\,(m_J-1);I\,(m_I+1)|H_{\rm hfs}|J\,m_J;I\,m_I\rangle &= \frac{A_{\rm hfs}}{2}\sqrt{(J-m_J+1)(J+m_J)(I+m_I+1)(I-m_I)}\\ \langle J\,(m_J+1);I\,(m_I-1)|H_{\rm hfs}|J\,m_J;I\,m_I\rangle &= \frac{A_{\rm hfs}}{2}\sqrt{(J+m_J+1)(J-m_J)(I-m_I+1)(I+m_I)},\end{aligned} \tag{7.329}$$

with all other matrix elements vanishing. The interaction Hamiltonian is diagonal in the strong-field basis, with diagonal matrix elements

$$\langle J\,m_J;I\,m_I|H_B^{(\rm hfs)}|J\,m_J;I\,m_I\rangle = \mu_{\rm B}(g_J m_J + g_I m_I)B. \tag{7.330}$$

Suppose now for concreteness that $J = 1/2$. Then states with $m_J = 1/2$ are only coupled to states with $m_J = -1/2$ (and m_I increased by 1), and states with $m_J = -1/2$ are only coupled to states with $m_J = 1/2$ (and m_I *decreased* by 1). Thus, the combined Hamiltonian is block diagonal, with blocks of the form

$$\left[\begin{array}{cc} A_{\rm hfs}\dfrac{m_I}{2} + \mu_{\rm B}\left(\dfrac{g_J}{2}+g_I m_I\right)B & \dfrac{A_{\rm hfs}}{2}\sqrt{(I+m_I+1)(I-m_I)} \\ \dfrac{A_{\rm hfs}}{2}\sqrt{(I+m_I+1)(I-m_I)} & -A_{\rm hfs}\dfrac{(m_I+1)}{2} + \mu_{\rm B}\left(-\dfrac{g_J}{2}+g_I(m_I+1)\right)B \end{array}\right]. \tag{7.331}$$

Then we have a matrix with eigenvalues of the form

$$\left[\begin{array}{cc} A & C \\ C & D \end{array}\right] \quad\longrightarrow\quad \text{eigenvalues:} \quad \frac{A+D}{2} \pm \frac{1}{2}\sqrt{(A-D)^2+4C^2}, \tag{7.332}$$

so that the new hyperfine eigenvalues are

$$\begin{aligned}E_{|J=1/2\,m_J;I\,m_I\rangle} = &-\frac{A_{\rm hfs}}{4} + \mu_{\rm B}g_I\left(m_I+\frac{1}{2}\right)B \\ &\pm\frac{1}{2}\left\{\left[\frac{A_{\rm hfs}}{2}(2m_I+1)+\mu_{\rm B}(g_J-g_I)B\right]^2 + A_{\rm hfs}^2[I(I+1)-m_I(m_I+1)]\right\}^{1/2}.\end{aligned} \tag{7.333}$$

Introducing the notations

$$\begin{aligned}\Delta E_{\rm hfs} &= A_{\rm hfs}\left(I+\frac{1}{2}\right)\\ x &= \frac{\mu_{\rm B}(g_J-g_I)B}{\Delta E_{\rm hfs}},\end{aligned} \tag{7.334}$$

where $\Delta E_{\rm hfs}$ is the hyperfine splitting and x is a scaled magnetic-field strength,

$$\begin{aligned}E_{|J=1/2\,m_J;I\,m_I\rangle} &= -\frac{\Delta E_{\rm hfs}}{2(2I+1)} + \mu_{\rm B}g_I\left(m_I+\frac{1}{2}\right)B \\ &\qquad \pm\frac{\Delta E_{\rm hfs}}{2}\left[\left(\frac{(2m_I+1)}{(2I+1)}+x\right)^2 + \frac{4I(I+1)}{(2I+1)^2} - \frac{4m_I(m_I+1)}{(2I+1)^2}\right]^{1/2}\\ &= -\frac{\Delta E_{\rm hfs}}{2(2I+1)} + \mu_{\rm B}g_I\left(m_I+\frac{1}{2}\right)B \pm \frac{\Delta E_{\rm hfs}}{2}\left[\frac{1}{(2I+1)} + 2x\frac{(2m_I+1)}{(2I+1)} + \frac{4I(I+1)}{(2I+1)^2} + x^2\right]^{1/2}\\ &= -\frac{\Delta E_{\rm hfs}}{2(2I+1)} + \mu_{\rm B}g_I\left(m_I+\frac{1}{2}\right)B \pm \frac{\Delta E_{\rm hfs}}{2}\left[1 + 2x\frac{(2m_I+1)}{(2I+1)} + x^2\right]^{1/2}\end{aligned} \tag{7.335}$$

Now recall that without the couplings (off-diagonal matrix elements), the $m_J = +1/2$ state is of higher energy than the $m_J = -1/2$ state (recall g_I is small compared to g_J). The couplings, just as was the case for the dressed states of the two-level atom, will simply be spread further due to the interaction. Thus, we can associate the $\pm$ energy with the $m_J = \pm 1/2$ state. Further, we can define $m = m_I + 1/2$; noting that we labeled the upper $(+)$ state as having quantum number m_I while the lower $(-)$ state has quantum number

$m_I + 1$, we can thus also interpret $m = m_I \pm m_J$, where m_I and m_J are the literal quantum numbers for each state, and thus we obtain the **Breit–Rabi formula**[37]

$$E_{|J=1/2\,m_J;I\,m_I\rangle} = -\frac{\Delta E_{\rm hfs}}{2(2I+1)} + g_I\mu_{\rm B}mB \pm \frac{\Delta E_{\rm hfs}}{2}\left(1 + \frac{4mx}{2I+1} + x^2\right)^{1/2}. \quad \text{(Breit–Rabi formula)} \tag{7.336}$$

Again, this formula applies to states with $J = 1/2$ and arbitrary I (but can of course be adapted to states with $I = 1/2$ and arbitrary J). In order to avoid a sign ambiguity in evaluating (7.336), the more direct formula

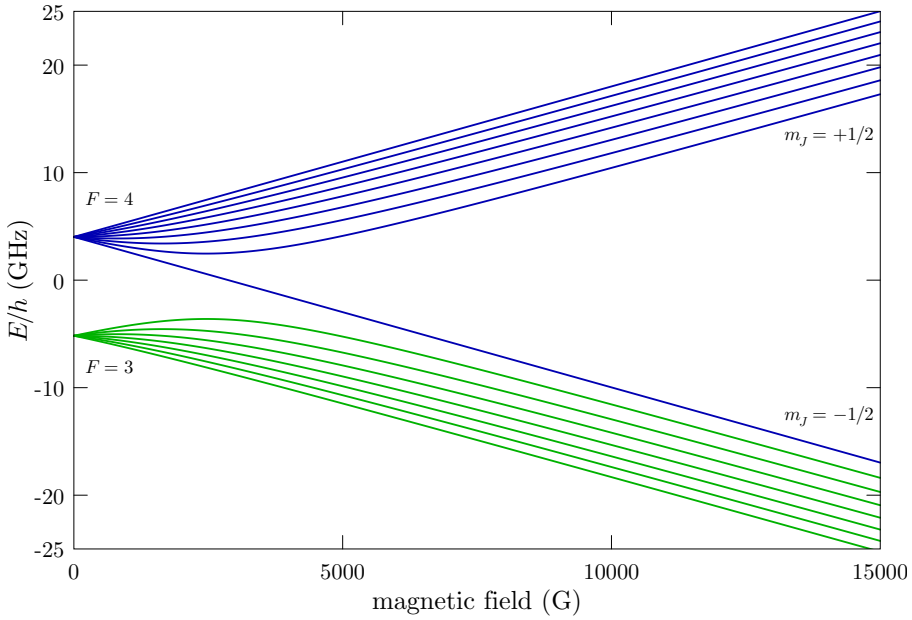
$$E_{|J=1/2\,m_J;I\,m_I\rangle} = \Delta E_{\rm hfs}\frac{I}{2I+1} \pm \frac{1}{2}(g_J + 2Ig_I)\mu_{\rm B}B \tag{7.337}$$

can be used for the two states $m = \pm(I+1/2)$. The Breit–Rabi formula is useful, for example, in computing the shifts of the ground states of the alkali atoms, which are of the form $n^2{\rm S}_{1/2}$. For example, recall above that the ground-state hyperfine $F = 2 \longrightarrow F' = 3$ splitting of ^{133}Cs defines the second, but of course the hyperfine transitions can depend on the local magnetic field. For this reason the $m_F = 0 \longrightarrow m'_F = 0$ transition is used, because both states have no Zeeman shift to lowest order. However, they *do* shift to higher order, and the small frequency shift is

$$\Delta\omega_{\rm clock} = \frac{(g_J - g_I)^2\mu_{\rm B}^2}{2\hbar\Delta E_{\rm hfs}}B^2 \tag{7.338}$$

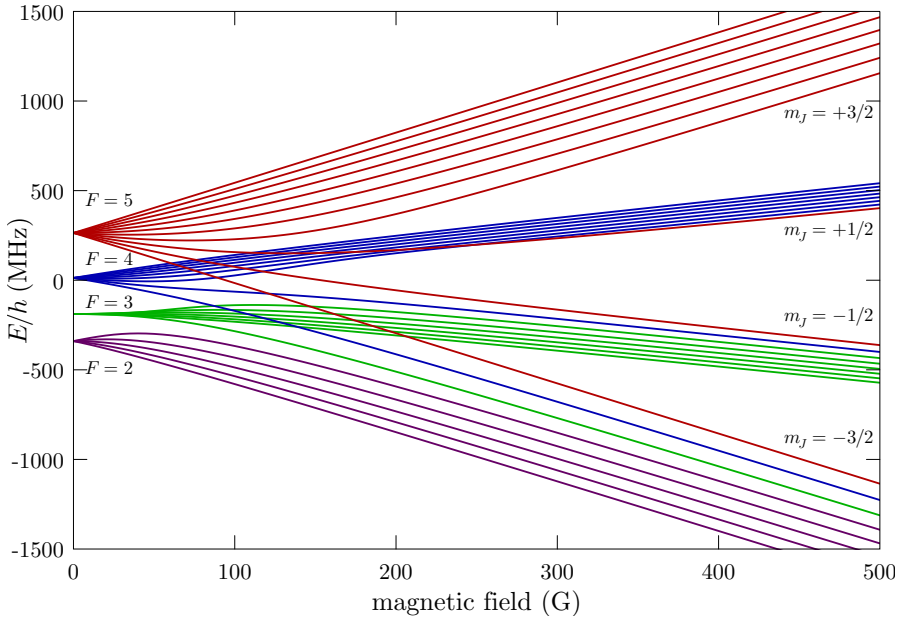
to second order in the field strength, an important systematic effect to keep in mind when designing an atomic clock.

As an example, shown below is the magnetic-field-dependent hyperfine structure of the ground ($6^2{\rm S}_{1/2}$) state of ^{133}Cs, ranging from the weak-field (Zeeman) regime through the hyperfine Paschen-Back regime.



For the ground state, the Breit–Rabi formula applies and can be used to compute the energy levels. However, it does not apply to the D_2 excited ($6^2{\rm P}_{3/2}$) state, where the level structure, shown below, is more complicated.

[37] G. Breit and I. I. Rabi, "Measurement of Nuclear Spin," *Physical Review* **38**, 2082 (1931) (doi: 10.1103/PhysRev.38.2082.2).



In both cases, the features we have discussed are visible. For small magnetic fields, the levels cluster around the hyperfine energies and have shifts proportional to B. For large magnetic fields, the states instead cluster according to their m_J value, with small splittings induced according to the value of m_I. In the incomplete Paschen–Back regime, there is a smooth crossover between the two types of level spacings.

7.4.2 Static Electric Fields: Stark Effect

Like static magnetic fields in the Zeeman effect, static *electric* fields also shift the fine- and hyperfine-structure sublevels. However, the details of the electric-field shifts, or **dc Stark shifts**[38] turn out to be quite different from the case of the Zeeman effect. We will take the atom–field interaction Hamiltonian to be the usual electric-dipole interaction,

$$H_{\text{AF}} = -\mathbf{d}\cdot\mathbf{E}, \tag{7.339}$$

where $\mathbf{E}$ is a static electric-field amplitude, and $\mathbf{d}$ is the atomic dipole operator as usual. The atomic energy-level shifts in level $|\alpha\rangle$ is given up to second order in perturbation theory by

$$\Delta E_\alpha = \langle\alpha|H_{\text{AF}}|\alpha\rangle + \sum_j \frac{|\langle\alpha|H_{\text{AF}}|\beta_j\rangle|^2}{E_\alpha - E_{\beta_j}}, \tag{7.340}$$

where the $|\beta_j\rangle$ label all the other atomic states, and E_α and E_{β_j} are the atomic level energies. Recalling that the dipole operator only couples states of opposite parity (Section 5.1.1), the first-order shift vanishes and we are left only with the second-order term. Thus, the remaining effect is second order in $\mathbf{E}$, and is thus called the **quadratic Stark effect**. (A notable exception to this occurs in hydrogen, where degeneracy of opposite-parity states leads to a first-order shift and thus a **linear Stark effect**.)

[38] J. Stark, "Beobachtungen über den Effekt des elektrischen Feldes auf Spektrallinien I. Quereffekt," *Annalen der Physik* **43**, 965 (1914).

7.4.2.1 Effective, First-Order Interaction

It is conventional to define an *effective* Stark interaction Hamiltonian by[39]

$$H_{\text{Stark}} := \sum_j \frac{H_{\text{AF}}|\beta_j\rangle\langle\beta_j|H_{\text{AF}}}{E_\alpha - E_{\beta_j}} = \sum_j \frac{d_\mu|\beta_j\rangle\langle\beta_j|d_\nu}{E_\alpha - E_{\beta_j}} E_\mu E_\nu, \tag{7.341}$$

where E_μ and E_ν are the electric-field components (make sure to keep track of which E's are energies and which are fields!). With this effective Hamiltonian, we get the same shift, but now it looks like a *first-order* shift:

$$\Delta E_\alpha = \langle\alpha|H_{\text{Stark}}|\alpha\rangle. \tag{7.342}$$

However, it's important to realize that this result is really *second* order in perturbation theory. Now we note that the Stark Hamiltonian has the form of a rank-2 tensor operator, contracted twice with the electric-field vector:

$$H_{\text{Stark}} = S_{\mu\nu}E_\mu E_\nu, \tag{7.343}$$

where the tensor operator is

$$S_{\mu\nu} = \sum_j \frac{d_\mu|\beta_j\rangle\langle\beta_j|d_\nu}{E_\alpha - E_{\beta_j}}. \tag{7.344}$$

This is a symmetric Cartesian tensor of rank 2. Recall from Section (7.3.3.3) that a rank-2 Cartesian tensor may be decomposed into irreducible parts of rank 0, 1, and 2, where the rank-0 part is related to the trace, the rank-1 part is related to the antisymmetric part of the tensor, and the rank-2 part is effectively what is left. Writing the decomposition as in Eq. (7.207),

$$S_{\mu\nu} = \frac{1}{3}S^{(0)}\delta_{\mu\nu} + S^{(2)}_{\mu\nu}, \tag{7.345}$$

where the vector term vanishes since (7.344) is obviously a symmetric tensor, the scalar part is

$$S^{(0)} = S_{\mu\mu}, \tag{7.346}$$

as in Eq. (7.199), and the irreducible tensor part is

$$S^{(2)}_{\mu\nu} = S_{\mu\nu} - \frac{1}{3}S_{\sigma\sigma}\delta_{\mu\nu}, \tag{7.347}$$

as in Eq. (7.206). Thus, the Stark shift from Eq. (7.342) becomes

$$\Delta E_\alpha = \langle\alpha|S_{\mu\nu}|\alpha\rangle E_\mu E_\nu = \frac{1}{3}\langle\alpha|S^{(0)}|\alpha\rangle E^2 + \langle\alpha|S^{(2)}_{\mu\nu}|\alpha\rangle E_\mu E_\nu, \tag{7.348}$$

which is now separated into scalar and tensor parts: the first is the *orientation-independent* part of the shift, while the second is the anisotropic part.

7.4.2.2 Scalar Shift: Fine Structure

The scalar part of the shift is given by the first term of Eq. (7.348). For a fine-structure state $|J\ m_J\rangle$, we may write the shift as

$$\Delta E^{(0)}_{|J\ m_J\rangle} = \sum_{J'\ m'_J} \frac{\langle J\ m_J|d_\mu|J'\ m'_J\rangle\langle J'\ m'_J|d_\mu|J'\ m'_J\rangle}{3(E_J - E_{J'})} E^2. \tag{7.349}$$

[39] For other treatments and more details, see J. R. P. Angel and P. G. H. Sandars, "The Hyperfine Structure Stark Effect. I. Theory," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **305**, 125 (1968); Abbas Khadjavi, Allen Lurio, and W. Happer, "Stark Effect in the Excited States of Rb, Cs, Cd, and Hg," *Physical Review* **167** 128 (1968) (doi: 10.1103/PhysRev.167.128); and Robert W. Schmieder, "Matrix Elements of the Quadratic Stark Effect on Atoms with Hyperfine Structure," *American Journal of Physics* **40**, 297 (1972) (doi: 10.1119/1.1986513).

Using the Wigner–Eckart theorem (7.237) for the matrix elements, we find

$$\Delta E_J^{(0)} = \sum_{J'\,m'_J\,\mu} \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{3(E_J - E_{J'})} \langle J\; m_J | J'\; m'_J; 1\; \mu\rangle^2 E^2. \tag{7.350}$$

Note that we transposed the second matrix element so that we could apply the Wigner–Eckart theorem in exactly the same form on each matrix element, ending up with one form of the reduced matrix element and the Clebsch–Gordan coefficient. From the orthogonality relation (7.46), we can evaluate the double sum

$$\sum_{m'_J\,\mu} \langle J\; m_J | J'\; m'_J; 1\; \mu\rangle^2 = 1. \tag{7.351}$$

Then, by analogy with the classical polarizability, as in Eq. (1.60), we then define a **scalar polarizability** for the fine-structure level J by

$$\alpha^{(0)}(J) := -\frac{2}{3} \sum_{J'} \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_J - E_{J'}}, \tag{7.352}$$
(scalar polarizability)

so that the scalar shift becomes

$$\Delta E_J^{(0)} = -\frac{1}{2} \alpha^{(0)}(J)\, E^2. \tag{7.353}$$
(scalar Stark shift)

The sum in the polarizability here extends over the other (i.e., radial) quantum numbers as well as J'.

7.4.2.3 Tensor Shift: Fine Structure

The remaining (tensor) part of the Stark shift is given by the second term of Eq. (7.348). Again, for a fine-structure state $|J\; m_J\rangle$, we may write the shift as

$$\Delta E^{(2)}_{|J\; m_J\rangle} = \sum_{q=-2}^{2} (-1)^q \langle J\; m_J | S^{(2)}_q | J\; m_J\rangle\, [\mathbf{EE}]^{(2)}_{-q}, \tag{7.354}$$

where we have switched to the scalar product (7.222) of spherical-tensors from the scalar product of the irreducible Cartesian tensors $S^{(2)}_{\mu\nu}$ and the traceless part of $E_\mu E_\nu$. (Note that the scalar part of $E_\mu E_\nu$ has a vanishing interaction with $S^{(2)}_{\mu\nu}$, since its trace has been removed.) Recall that the spherical-tensor components for the rank-2 tensors are given in Eq. (7.219). Again using the Wigner–Eckart theorem (7.237) for the matrix element, we can write

$$\Delta E^{(2)}_{|J\; m_J\rangle} = \sum_{q=-2}^{2} (-1)^q \langle J\|\mathbf{S}^{(2)}\|J\rangle\, [\mathbf{EE}]^{(2)}_{-q}\, \langle J\; m_J | J\; m_J; 2\; q\rangle, \tag{7.355}$$

where the only nonvanishing Clebsch–Gordan coefficient is

$$\langle J\; m_J | J\; m_J; 2\; 0\rangle = \frac{3m_J^2 - J(J+1)}{\sqrt{J(2J-1)(J+1)(2J+3)}} \qquad (J \geq 1), \tag{7.356}$$

where the constraint on J ensures that the triangularity constraint of the Clebsch–Gordan coefficient is satisfied ($|J-2| \leq J \leq J+2$ here). That is, the tensor shift vanishes if $J = 0$ or $J = 1/2$. From Eq. (7.219), the relevant component of the field tensor is

$$[\mathbf{EE}]^{(2)}_0 = \frac{1}{\sqrt{6}}(E_1 E_{-1} + 2E_0^2 + E_{-1}E_1) = \frac{1}{\sqrt{6}}(3E_z^2 - E^2). \tag{7.357}$$

Putting these pieces together, we can define a **tensor polarizability**

$$\alpha^{(2)}(J) := -\langle J\|S^{(2)}_q\|J\rangle \sqrt{\frac{8J(2J-1)}{3(J+1)(2J+3)}}, \tag{7.358}$$
(tensor polarizability)

so that the tensor shift is

$$\Delta E^{(2)}_{|J\,m_J\rangle} = -\frac{1}{4}\alpha^{(2)}(J)\,(3E_z^2 - E^2)\left(\frac{3m_J^2 - J(J+1)}{J(2J-1)}\right). \tag{7.359}$$
(tensor Stark shift)

Combining this with the scalar shift from Eq. (7.353), we can write the total shift as

$$\Delta E_{|J\,m_J\rangle} = -\frac{1}{2}\alpha^{(0)}(J)\,E^2 - \frac{1}{4}\alpha^{(2)}(J)\,(3E_z^2 - E^2)\left(\frac{3m_J^2 - J(J+1)}{J(2J-1)}\right). \tag{7.360}$$
(fine-structure Stark shift)

The shift is largest when $\mathbf{E} = E_z\hat{z}$ (e.g., when we take the quantization axis to be along the $\mathbf{E}$ field), when the Stark shift simplifies to

$$\Delta E_{|J\,m_J\rangle} = -\frac{1}{2}\alpha^{(0)}(J)\,E_z^2 - \frac{1}{2}\alpha^{(2)}(J)\,E_z^2\left(\frac{3m_J^2 - J(J+1)}{J(2J-1)}\right). \tag{7.361}$$
(fine-structure Stark shift)

This latter form explains the normalization of the tensor term: for $m_J = \pm J$, the shift becomes

$$\Delta E_{|J\,\pm J\rangle} = -\frac{1}{2}\alpha^{(0)}(J)\,E_z^2 - \frac{1}{2}\alpha^{(2)}(J)\,E_z^2, \tag{7.362}$$

where the tensor term has the same form as the scalar term. Because of these forms for the Stark shift, we can write down yet another effective Stark Hamiltonian,

$$H_{\text{Stark}}(J) = -\frac{1}{2}\alpha^{(0)}(J)\,E_z^2 - \frac{1}{4}\alpha^{(2)}(J)\,(3E_z^2 - E^2)\left(\frac{3J_z^2/\hbar^2 - J(J+1)}{J(2J-1)}\right), \tag{7.363}$$
(effective Stark Hamiltonian)

which has the same eigenvalues as the effective Hamiltonian (7.341) for fine-structure states $|J\,m_J\rangle$. Again, we have taken the quantization axis to coincide with the electric-field direction. This is because the matrix $\langle J\,m_J|H_{\text{Stark}}|J\,m_J'\rangle$ for the original Hamiltonian is only diagonal for $\mathbf{E} = E_z\hat{z}$. Only for this choice of $\mathbf{E}$ is $[\mathbf{EE}]_0^{(2)}$ the only nonvanishing tensor component, as is evident from Eqs. (7.219), and the off-diagonal analogue of Eq. (7.355) has a Clebsch–Gordan coefficient that vanishes except when $m_J = m_J'$, since $q = 0$.

Obviously, this effective applies so long as the shift is weak enough that states of different J do not mix. Again, we reiterate that the tensor polarizability vanishes if $J = 0$ or $J = 1/2$. More intuitively, we can understand this by noting that if $J = 0$, then there is only one m_J level, and so there can be no orientation dependence; the entire Stark shift is accounted for by the scalar term. Similarly, if $J = 1/2$, the tensor Stark shift as we have seen only depends on $|m_J|$, so the two sublevels are degenerate. Again there is no orientation dependence, and thus no tensor shift.

A more explicit expression for the tensor polarizability comes about if we factorize the reduced matrix element $\langle J\|\mathbf{S}^{(2)}\|J\rangle$ into reduced matrix elements of dipole operators. We wish to apply Eq. (7.273) to perform the factorizaton, regarding $S_q^{(2)}$ to be the rank-2 product of the vector d_μ and the vector

$$\sum_j \frac{|\beta_j\rangle\langle\beta_j|}{E_\alpha - E_{\beta_j}} d_\nu \tag{7.364}$$

from the definition in Eq. (7.344). The factorized matrix element then becomes

$$\begin{aligned}
\langle J\|\mathbf{S}^{(2)}\|J\rangle &= (-1)^{2J}\sum_{J'}\sqrt{5(2J'+1)}\begin{Bmatrix} 1 & 1 & 2 \\ J & J & J' \end{Bmatrix}\langle J\|\mathbf{d}\|J'\rangle\,\langle J'\|\sum_j \frac{|\beta_j\rangle\langle\beta_j|}{E_J - E_{\beta_j}}\mathbf{d}\|J\rangle \\
&= (-1)^{2J}\sum_{J'}\sqrt{5(2J'+1)}\begin{Bmatrix} 1 & 1 & 2 \\ J & J & J' \end{Bmatrix}\frac{\langle J\|\mathbf{d}\|J'\rangle\,\langle J'\|\mathbf{d}\|J\rangle}{E_J - E_{J'}} \\
&= \sum_{J'}(-1)^{J+J'}\sqrt{5(2J+1)}\begin{Bmatrix} 1 & 1 & 2 \\ J & J & J' \end{Bmatrix}\frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_J - E_{J'}},
\end{aligned} \tag{7.365}$$

where in the last step we used Eq. (7.259) to conjugate the second matrix element:

$$\langle J' \| \mathbf{d} \| J \rangle = (-1)^{J-J'} \sqrt{\frac{2J+1}{2J'+1}} \, \langle J \| \mathbf{d} \| J' \rangle^* . \tag{7.366}$$

Putting this matrix element into the expression (7.358) for the tensor polarizability, we arrive at the expression[40]

$$\alpha^{(2)}(J) = \sum_{J'} (-1)^{J+J'+1} \sqrt{\frac{40J(2J+1)(2J-1)}{3(J+1)(2J+3)}} \begin{Bmatrix} 1 & 1 & 2 \\ J & J & J' \end{Bmatrix} \frac{|\langle J \| \mathbf{d} \| J' \rangle|^2}{E_J - E_{J'}},$$

(tensor polarizability) (7.367)

thus giving the tensor polarizability as a direct sum over dipole matrix elements. Again, E_J is the energy of level J, and in the sum over J' we also implicitly sum over any other necessary quantum numbers to enumerate all possible states of the same J'.

7.4.2.4 Hyperfine Structure: Weak Fields

In the case of hyperfine structure, we can make exactly the same arguments as we did for fine structure, but using the hyperfine quantum number F instead of J. Thus, the effective hyperfine Stark Hamiltonian as in Eq. (7.363) is

$$H_{\text{Stark}}(F) = -\frac{1}{2}\alpha^{(0)}(F)\, E_z^2 - \frac{1}{2}\alpha^{(2)}(F)\, (3E_z^2 - E^2) \left(\frac{3F_z^2/\hbar^2 - F(F+1)}{F(2F-1)} \right),$$

(effective hyperfine Stark Hamiltonian) (7.368)

such that the quadratic Stark shift as in Eq. (7.360) is

$$\Delta E_{|F\ m_F\rangle} = -\frac{1}{2}\alpha^{(0)}(F)\, E_z^2 - \frac{1}{4}\alpha^{(2)}(F)\, (3E_z^2 - E^2) \left(\frac{3m_F^2 - F(F+1)}{F(2F-1)} \right),$$

(hyperfine-structure Stark shift) (7.369)

the hyperfine scalar polarizability defined as in Eq. (7.352) by

$$\alpha^{(0)}(F) := -\frac{2}{3} \sum_{F'} \frac{|\langle F \| \mathbf{d} \| F' \rangle|^2}{E_F - E_{F'}},$$

(hyperfine scalar polarizability) (7.370)

and the hyperfine tensor polarizability is given as in Eq. (7.367) by

$$\alpha^{(2)}(F) = \sum_{F'} (-1)^{F+F'+1} \sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}} \begin{Bmatrix} 1 & 1 & 2 \\ F & F & F' \end{Bmatrix} \frac{|\langle F \| \mathbf{d} \| F' \rangle|^2}{E_F - E_{F'}}.$$

(tensor polarizability) (7.371)

These expressions suffice to compute the Stark shift due to static electric fields in the case of hyperfine structure.

However, recalling that the electric field coupled only to the electron and not to the nucleus, the effective Stark Hamiltonian (7.363) in terms of the electron angular momentum J must also be a perfectly good Hamiltonian for the hyperfine Stark shift. For example, we may illustrate this by relating the fine-structure polarizabilities to the hyperfine versions. Starting with the scalar hyperfine polarizability, we can start with Eq. (7.282) to reduce the hyperfine matrix element to a fine-structure matrix element in

[40] cf. Khadjavi *et al.* and Angel *et al.*, noting the difference in convention for the reduced dipole matrix element.

Eq. (7.370),

$$\begin{aligned}\alpha^{(0)}(F) &= -\frac{2}{3}\sum_{F'J'}(2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I\end{array}\right\}^2 \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_F - E_{F'}} \\ &\approx -\frac{2}{3}\sum_{F'J'}(2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I\end{array}\right\}^2 \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_J - E_{J'}} \\ &= -\frac{2}{3}\sum_{J'}\frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_J - E_{J'}},\end{aligned} \tag{7.372}$$

where we used the orthogonality relation (7.82) for the 6-j symbols in the last step, and we assumed that the hyperfine splittings were small enough that $E_F \approx E_J$. Then comparing to Eq. (7.352), we see that the scalar polarizability is the same in either case,

$$\alpha^{(0)}(F) \approx \alpha^{(0)}(J),$$

(hyperfine and fine-structure scalar polarizabilities) (7.373)

at least to the extent that the hyperfine splittings lend a negligible contribution to the polarizability (which is generally true to within modern experimental error in precision measurements of polarizabilities). Similarly, for the tensor polarizability,

$$\begin{aligned}\alpha^{(2)}(F) = \sum_{F'J'}(-1)^{F+F'+1}\sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}}\left\{\begin{array}{ccc} 1 & 1 & 2 \\ F & F & F'\end{array}\right\} \\ \times (2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I\end{array}\right\}^2 \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_F - E_{F'}}.\end{aligned} \tag{7.374}$$

Again, making the approximation $E_F \approx E_J$, we can then carry out the sum over F' via the Biedenharn–Elliott sum rule (7.102), which gives

$$\sum_{F'}(-1)^{F'}(2F'+1)\left\{\begin{array}{ccc} 1 & 1 & 2 \\ F & F & F'\end{array}\right\}\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I\end{array}\right\}^2 = (-1)^{-(2J+J'+2F+I)}\left\{\begin{array}{ccc} 1 & 1 & 2 \\ J & J & J'\end{array}\right\}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I\end{array}\right\}. \tag{7.375}$$

Putting this into the above expression for the tensor polarizability, we have

$$\alpha^{(2)}(F) \approx \sum_{J'}(-1)^{-2J-J'-F-I+1}(2J+1)\sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}}\left\{\begin{array}{ccc} 1 & 1 & 2 \\ J & J & J'\end{array}\right\}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I\end{array}\right\}\frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{E_J - E_{J'}}, \tag{7.376}$$

and on comparison to Eq. (7.367), we can relate the hyperfine and fine-structure polarizabilities by

$$\alpha^{(2)}(F) \approx (-1)^{-3J-2J'-F-I}\sqrt{\frac{(J+1)(2J+1)(2J+3)F(2F+1)(2F-1)}{J(2J-1)(F+1)(2F+3)}}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I\end{array}\right\}\alpha^{(2)}(J). \tag{7.377}$$

From the 6-j symbol, the combination $J+F+I \in \mathbb{Z}$, and since J and J' differ by 0 or 1, we conclude that $J'+F+I \in \mathbb{Z}$. Thus, we can multiply by $(-1)^{2(J'+F+I)}(-1)^{4J} = 1$ to obtain the result

$$\alpha^{(2)}(F) \approx (-1)^{J+I+F}\sqrt{\frac{(J+1)(2J+1)(2J+3)F(2F+1)(2F-1)}{J(2J-1)(F+1)(2F+3)}}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I\end{array}\right\}\alpha^{(2)}(J). \tag{7.378}$$

We can evaluate the 6-j coefficient here, with the result

$$\alpha^{(2)}(F) \approx \frac{3X(X-1) - 4F(F+1)J(J+1)}{(2F+3)(2F+2)J(2J-1)}\alpha^{(2)}(J),$$

(hyperfine and fine-structure tensor polarizabilities) (7.379)

where

$$X := F(F+1) - I(I+1) + J(J+1). \tag{7.380}$$

Then the total hyperfine Stark shift is

$$\Delta E_{|F\ m_F\rangle} \approx -\frac{1}{2}\alpha^{(0)}(J)\,E_z^2 - \alpha^{(2)}(J)\,(3E_z^2 - E^2)\,\frac{[3m_F^2 - F(F+1)][3X(X-1) - 4F(F+1)J(J+1)]}{(2F+3)(2F+2)(2F)(2F-1)(2J)(2J-1)},$$

(hyperfine-structure Stark shift) (7.381)

Again, these expression makes the (generally good) approximation of neglecting hyperfine shifts in computing the relevant transition energies. Also, note that we have assumed in writing down these expressions that F is a good quantum number, and thus we assume the tensor shifts to be much smaller than the hyperfine splittings.

Recalling that the tensor polarizability vanished in the fine-structure cases of $J = 0$ and $J = 1/2$, we can see by the coefficient in this expression that the tensor *hyperfine* polarizability also vanishes for $F = 0$ or $F = 1/2$ (and also still vanishes if $J = 0$ or $J = 1/2$). In particular, for the ground state of alkali atoms, where $J = 1/2$, there is no tensor Stark shift. This statement predicts that for the cesium-clock hyperfine transition, for example, the transition frequency is independent of a dc electric field, because the levels shift together. Actually, this statement is only true in second-order perturbation theory, as we have used here; in *third*-order perturbation theory, it turns out that, with hyperfine structure, there is *still* a small tensor Stark shift.[41]

7.4.2.5 Hyperfine Structure: Stronger Fields

In the case of stronger fields, when F is no longer a good quantum number, the formulae of the previous section no longer apply. However, the effective Stark interaction Hamiltonian (7.363) for the interaction of the electron with the static electric field is still valid. In the limit of a very strong electric field, the interaction $H_{\mathbf{Stark}}(J)$ will dominate the hyperfine Hamiltonian $H_{\rm hfs}$, and as in the Paschen–Back (strong-field) interaction for magnetic fields, the appropriate basis is $|J\ m_J; I\ m_I\rangle$, where $H_{\mathbf{Stark}}(J)$ is diagonal. In this case, ignoring the hyperfine splittings, the energies are given by the fine-structure expression (7.360). Also, in the same way as in Eq. (7.326) for the Paschen–Back effect, we can keep the lowest-order contribution of $H_{\rm hfs}$ by taking its expectation value in the strong-field basis, with the result

$$E_{|J\ m_J;I\ m_I\rangle} \approx A_{\rm hfs} m_I m_J + B_{\rm hfs}\frac{9(m_I m_J)^2 - 3J(J+1)m_I^2 - 3I(I+1)m_J^2 + I(I+1)J(J+1)}{4J(2J-1)I(2I-1)} - \frac{1}{2}\alpha^{(0)}(J)\,E^2 - \frac{1}{4}\alpha^{(2)}(J)\,(3E_z^2 - E^2)\left(\frac{3m_J^2 - J(J+1)}{J(2J-1)}\right).$$

(hyperfine Stark shift, strong field) (7.382)

Thus, in this "electric Paschen–Back" regime, we expect the hyperfine sublevels to split into major groups according to the value of $|m_J|$, with smaller splittings according to m_I. Obviously this only works if the shifts are not so large that they mix state of different J.

However, for general (intermediate) electric fields, we must in general diagonalize $H_{\rm hfs} + H_{\rm Stark}(J)$, as is the case in general for magnetic fields. For example, using

$$\begin{aligned}\langle F\ m_F|J_z^2|F'\ m_F'\rangle &= \sum_{m_J m_I m_J' m_I'} \langle F\ m_F|J\ m_J; I\ m_I\rangle\,\langle J\ m_J; I\ m_I|J_z^2|J\ m_J'; I\ m_I'\rangle\,\langle J\ m_J'; I\ m_I'|F'\ m_F'\rangle \\ &= \sum_{m_J m_I} \hbar^2 m_J^2 \langle J\ m_J; I\ m_I|F\ m_F\rangle\,\langle J\ m_J; I\ m_I|F'\ m_F'\rangle,\end{aligned} \tag{7.383}$$

[41] A. Weis and S. Ulzega "The Stark effect of the hyperfine structure of cesium," *Proceedings of SPIE* **6604**, 660408 (2007) (doi: 10.1117/12.726805).

we can write the matrix elements of the Stark interaction in the hyperfine basis as

$$\begin{aligned}\langle F\ m_F|H_{\rm Stark}(J)|F'\ m'_F\rangle = &-\frac{1}{2}\alpha^{(0)}(J)\,E_z^2\delta_{FF'}\delta_{m_F m'_F}\\ &-\frac{1}{4}\alpha^{(2)}(J)\,\frac{(3E_z^2-E^2)}{J(2J-1)}\Bigg(3\sum_{m_J m_I} m_J^2\langle J\ m_J; I\ m_I|F\ m_F\rangle\,\langle J\ m_J; I\ m_I|F'\ m'_F\rangle\\ &\qquad - J(J+1)\delta_{FF'}\delta_{m_F m'_F}\Bigg),\end{aligned} \tag{7.384}$$

and then diagonalize the resulting matrix numerically. A somewhat nicer expression comes from using the fact that $(3J_z^2-\mathbf{J}^2)$ is the $q=0$ component of a rank-2 spherical tensor operator. Applying the Wigner–Eckart theorem (7.237),

$$\begin{aligned}\langle F\ m_F|(3J_z^2-\mathbf{J}^2)|F'\ m'_F\rangle &= \langle F\|(3J_z^2-\mathbf{J}^2)\|F'\rangle\,\langle F\ m_F|F'\ m'_F; 2\ 0\rangle\\ &= \delta_{m_F m'_F}\langle F\|(3J_z^2-\mathbf{J}^2)\|F'\rangle\,\langle F\ m_F|F'\ m_F; 2\ 0\rangle\\ &= \delta_{m_F m'_F}(-1)^{F'+J+I}\sqrt{(2F'+1)(2J+1)}\begin{Bmatrix} J & J' & 2\\ F' & F & I\end{Bmatrix}\\ &\qquad\times\langle J\|(3J_z^2-\mathbf{J}^2)\|J'\rangle\,\langle F\ m_F|F'\ m_F; 2\ 0\rangle\\ &= \delta_{m_F m'_F}(-1)^{F'+J+I}\sqrt{(2F'+1)(2J+1)}\begin{Bmatrix} J & J & 2\\ F' & F & I\end{Bmatrix}\\ &\qquad\times\delta_{JJ'}\hbar^2\sqrt{J(J+1)(2J-1)(2J+3)}\langle F\ m_F|F'\ m_F; 2\ 0\rangle\end{aligned} \tag{7.385}$$

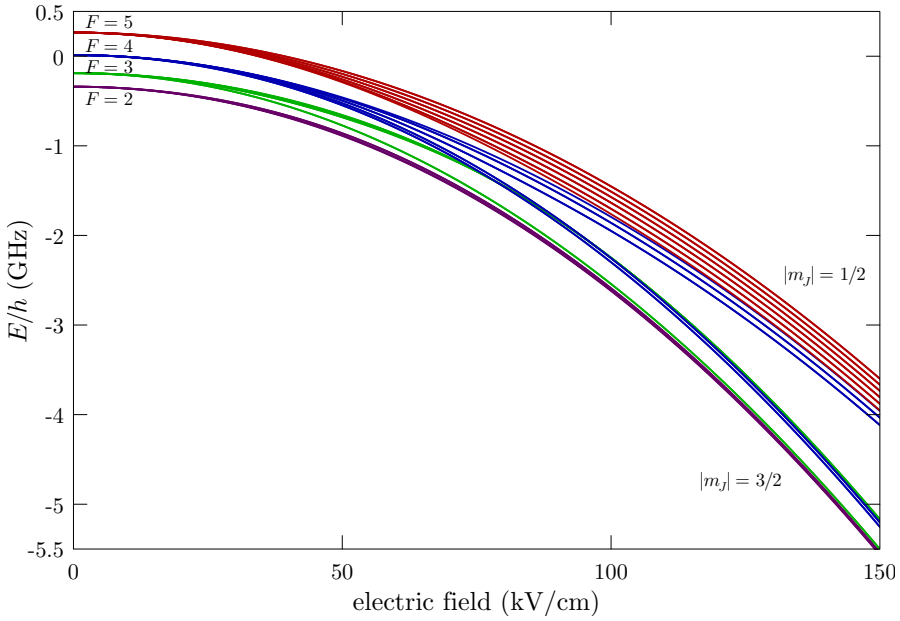
where we used Eq. (7.261) to change to the fine-structure reduced matrix element, and we also used the Wigner–Eckart theorem (7.237) to evaluate the reduced matrix element:

$$\begin{aligned}\langle J\|(3J_z^2-\mathbf{J}^2)\|J'\rangle\,\langle J\ m_J|J'\ m_J; k\ 0\rangle &= \langle J\ m_J|(3J_z^2-\mathbf{J}^2)|J'\ m_J\rangle\\ &= \delta_{JJ'}\hbar^2[3m_J^2-J(J+1)]\\ &= \delta_{JJ'}\hbar^2\langle J\ m_J|J\ m_J; 2\ 0\rangle\sqrt{J(J+1)(2J-1)(2J+3)}.\end{aligned} \tag{7.386}$$

Thus, the hyperfine matrix elements of the Stark Hamiltonian become

$$\begin{aligned}\langle F\ m_F|H_{\rm Stark}(J)|F'\ m'_F\rangle = &-\frac{1}{2}\alpha^{(0)}(J)\,E_z^2-\frac{1}{4}\alpha^{(2)}(J)\,(3E_z^2-E^2)\delta_{JJ'}\delta_{m_F m'_F}\\ &\times(-1)^{F'+J+I}\sqrt{\frac{(2F'+1)(2J+1)(2J+2)(2J+3)}{2J(2J-1)}}\\ &\times\begin{Bmatrix} J & J & 2\\ F' & F & I\end{Bmatrix}\langle F\ m_F|F'\ m_F; 2\ 0\rangle\end{aligned} \tag{7.387}$$

After adding the diagonal matrix $H_{\rm hfs}$, the result can be diagonalized to obtain the energies. Alternately, of course, the hyperfine Hamiltonian can be written in the strong-field basis $|J\ m_J; I\ m_I\rangle$ and then added to the Stark Hamiltonian in the same way; this method also conveniently carries over to the magnetic-field case. Shown below are the energies of the $6\,^2\mathrm{P}_{3/2}$ hyperfine manifold of ^{133}Cs, as in the magnetic-field example above, as obtained via numerical diagonalization. The overall scalar shift is visible as a quadratic downward trend of all the energy levels. The tensor part of the Stark shift is visible as a splitting of the initially degenerate hyperfine sublevels, which break up at the largest fields according to $|m_J|$ in the ''electric Paschen–Back'' regime that we mentioned above.



7.5 Interactions with Optical Fields

7.5.1 Atomic Fine-Structure Hamiltonian

We will now consider the interaction of a monochromatic laser field with an atomic Zeeman-degenerate fine-structure transition $J_g \longrightarrow J_e$, making the simplifying assumption of a closed transition (i.e., no other ground or excited levels are involved), and of course recalling that the selection rule $|J_e - J_g| \leq 1$ applies. The atomic Hamiltonian is

$$H_A = \hbar\omega_0 \sum_{m_e} |J_e\ m_e\rangle\langle J_e\ m_e|, \tag{7.388}$$

if we assume degenerate magnetic sublevels and a transition frequency of ω_0, while choosing the ground level to have zero energy. Since we will further consider an interaction with a field of frequency ω, we can follow the example from the two-level atom and transform into the rotating frame of the laser field, which we recall amounts to shifting the excited states down in energy by $\hbar\omega$. The rotating-frame Hamiltonian is thus

$$\tilde{H}_A = -\hbar\Delta \sum_{m_e} |J_e\ m_e\rangle\langle J_e\ m_e|, \tag{7.389}$$

where the field detuning from the atomic resonance is $\Delta := \omega - \omega_0$ as usual. Effects that break the degeneracy of the sublevels are easily accounted for here by including extra m-dependent shifts, such as to include Zeeman or dc Stark shifts.

7.5.2 Dipole and Atomic Lowering Operators

We turn now to the dipole interaction between atom and field. Just as in the case of the two-level atom, if we assume the ground and excited levels to be of opposite parity, we can now decompose the q component of the dipole operator in the spherical basis into positive- and negative-rotating parts. Recall that we did this equivalently in terms of the time dependence of expectation values or the lowering/raising character of the operators. Denoting the projection operators for the excited and ground levels by

$$P_e := \sum_{m_e} |J_e\ m_e\rangle\langle J_e\ m_e|, \qquad P_g := \sum_{m_g} |J_g\ m_g\rangle\langle J_g\ m_g|, \tag{7.390}$$

respectively, we can conveniently separate the two parts by writing

$$\begin{aligned} d_q &= (P_\mathrm{e} + P_\mathrm{g}) d_q (P_\mathrm{e} + P_\mathrm{g}) \\ &= P_\mathrm{g} d_q P_\mathrm{e} + P_\mathrm{e} d_q P_\mathrm{g} \\ &= d_q^{(+)} + d_q^{(-)}, \end{aligned} \tag{7.391}$$

since the dipole operator does not couple states within the same level as a consequence of its parity properties. Then the positive-rotating part becomes

$$\begin{aligned} d_q^{(+)} &= P_\mathrm{g} d_q P_\mathrm{e} \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} \langle J_\mathrm{g}\ m_\mathrm{g} | d_q | J_\mathrm{e}\ m_\mathrm{e} \rangle\ |J_\mathrm{g}\ m_\mathrm{g}\rangle\langle J_\mathrm{e}\ m_\mathrm{e}| \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} \langle J_\mathrm{g} \| \mathbf{d} \| J_\mathrm{e} \rangle\ \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ q \rangle\ |J_\mathrm{g}\ m_\mathrm{g}\rangle\langle J_\mathrm{e}\ m_\mathrm{e}|, \end{aligned} \tag{7.392}$$

where we have used the Wigner–Eckart theorem (7.242) in the second step. In each term of the sum, the projection quantum numbers are of course subject to the constraints $m_\mathrm{g} = m_\mathrm{e} + q$ for the term to be nonvanishing. Introducing the notation

$$\sigma(m_\mathrm{g}, m_\mathrm{e}) := |J_\mathrm{g}\ m_\mathrm{g}\rangle\langle J_\mathrm{e}\ m_\mathrm{e}| \tag{7.393}$$

for individual lowering operators, we have

$$d_q^{(+)} = \sum_{m_\mathrm{e} m_\mathrm{g}} \langle J_\mathrm{g} \| \mathbf{d} \| J_\mathrm{e} \rangle\ \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ q \rangle\ \sigma(m_\mathrm{g}, m_\mathrm{e}). \tag{7.394}$$

Recall from Section 7.3.5 that d_q is only a Hermitian operator in the case $q = 0$. Thus, $d_q^{(+)}$ and $d_q^{(-)}$ are not Hermitian conjugates unless $q = 0$, but rather

$$\begin{aligned} d_q^{(-)} &= P_\mathrm{e} d_q P_\mathrm{g} \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} \langle J_\mathrm{e}\ m_\mathrm{e} | d_q | J_\mathrm{g}\ m_\mathrm{g} \rangle\ |J_\mathrm{e}\ m_\mathrm{e}\rangle\langle J_\mathrm{g}\ m_\mathrm{g}| \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} \langle J_\mathrm{e} \| \mathbf{d} \| J_\mathrm{g} \rangle\ \langle J_\mathrm{e}\ m_\mathrm{e} | J_\mathrm{g}\ m_\mathrm{g}; 1\ q \rangle\ |J_\mathrm{e}\ m_\mathrm{e}\rangle\langle J_\mathrm{g}\ m_\mathrm{g}| \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} (-1)^q \langle J_\mathrm{g} \| \mathbf{d} \| J_\mathrm{e} \rangle\ \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ -q \rangle\ |J_\mathrm{e}\ m_\mathrm{e}\rangle\langle J_\mathrm{g}\ m_\mathrm{g}| \\ &= \sum_{m_\mathrm{e} m_\mathrm{g}} (-1)^q \langle J_\mathrm{g} \| \mathbf{d} \| J_\mathrm{e} \rangle\ \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ -q \rangle\ \sigma^\dagger(m_\mathrm{g}, m_\mathrm{e}), \end{aligned} \tag{7.395}$$

where we switched the forms of the Clebsch–Gordan coefficient as in Eq. (7.242) along with the reduced-matrix-element fomula (7.259). This operator thus lowers m_e to m_g, but the constraint is now $m_\mathrm{g} = m_\mathrm{e} - q$, so the transition is *not* the reverse of that in $d_q^{(+)}$ unless $q = 0$. Evidently, then, $d_q^{(-)}$ is the Hermitian conjugate of $(-1)^q d_{-q}^{(+)}$, as we expected from Section 7.3.5. Introducing the weighted lowering operators for the entire $J_\mathrm{g} \longrightarrow J_\mathrm{e}$ transition by

$$\begin{aligned} \Sigma_q &:= \sum_{m_\mathrm{g} m_\mathrm{e}} \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ q \rangle\ |J_\mathrm{g}\ m_\mathrm{g}\rangle\langle J_\mathrm{e}\ m_\mathrm{e}| \\ &= \sum_{m_\mathrm{g} m_\mathrm{e}} \langle J_\mathrm{g}\ m_\mathrm{g} | J_\mathrm{e}\ m_\mathrm{e}; 1\ q \rangle\ \sigma(m_\mathrm{g}, m_\mathrm{e}), \end{aligned} \tag{7.396}$$

we can additionally implement the constraint $m_g = m_e + q$ on the sublevels to write

$$\begin{aligned}\Sigma_q &= \sum_{m_e} \langle J_g\ m_e + q | J_e\ m_e; 1\ q\rangle\ |J_g\ m_e + q\rangle\langle J_e\ m_e| \\ &= \sum_{m_g} \langle J_g\ m_g | J_e\ m_g - q; 1\ q\rangle\ |J_g\ m_g\rangle\langle J_e\ m_g - q|.\end{aligned}$$

(atomic lowering operators, spherical basis) (7.397)

Thus, we see that when Σ_q lowers an excited sublevel m_e, it "returns" the sublevel labeled by $m_e + q$. Note, however, that Σ_q is not a proper tensor operator, in the sense that $\Sigma_q^\dagger \neq (-1)^q \Sigma_{-q}$.

Then using Eq. (7.396), we may rewrite the dipole-operator part of Eqs. (7.394) and (7.395) as

$$\begin{aligned}d_q &= d_q^{(+)} + d_q^{(-)} \\ &= \langle J_g \| \mathbf{d} \| J_e \rangle \left(\Sigma_q + (-1)^q \Sigma_{-q}^\dagger \right).\end{aligned}$$

(dipole operator in terms of lowering operators) (7.398)

This form is the generalization of the two-level-atom expression for the dipole operator [cf. Eq. (5.12)] to a fine-structure transition in a physical atom.

7.5.3 Dipole Interaction

Now considering the usual interaction Hamiltonian

$$H_{AF} = -\mathbf{d} \cdot \mathbf{E} = -\sum_q (-1)^q d_q E_{-q} \tag{7.399}$$

for the atom with a monochromatic optical field of frequency ω, we can make the rotating-wave approximation and implement the spherical-basis dot product as in Eq. (7.192) to obtain

$$H_{AF} = -\mathbf{d}^{(+)} \cdot \mathbf{E}^{(-)} - \mathbf{d}^{(-)} \cdot \mathbf{E}^{(+)} = -\sum_q (-1)^q \left(d_q^{(+)} E_{-q}^{(-)} + d_q^{(-)} E_{-q}^{(+)} \right). \tag{7.400}$$

Then using Eq. (7.398) for the dipole operator, the interaction Hamiltonian becomes

$$H_{AF} = -\sum_q \langle J_g \| \mathbf{d} \| J_e \rangle \left[(-1)^q E_{-q}^{(-)}(t)\, \Sigma_q + E_{-q}^{(+)}(t)\, \Sigma_{-q}^\dagger \right]. \tag{7.401}$$

Defining the vector Rabi frequency

$$\Omega_q := -\frac{2\langle J_g \| \mathbf{d} \| J_e \rangle E_q^{(+)}(0)}{\hbar}, \tag{7.402}$$

(vector Rabi frequency)

and noting that this implies

$$(\Omega_q)^* = -(-1)^q \frac{2\langle J_g \| \mathbf{d} \| J_e \rangle E_{-q}^{(-)}(0)}{\hbar}, \tag{7.403}$$

we can write the atom–field interaction as

$$H_{AF} = \frac{\hbar}{2} \sum_q \left[\Omega_q^* \Sigma_q e^{i\omega t} + \Omega_{-q} \Sigma_{-q}^\dagger e^{-i\omega t} \right], \tag{7.404}$$

or letting $q \longrightarrow -q$ in the second term,

$$H_{AF} = \frac{\hbar}{2} \sum_q \left[\Omega_q^* \Sigma_q e^{i\omega t} + \Omega_q \Sigma_q^\dagger e^{-i\omega t} \right].$$

(atom–field interaction, fine-structure transition) (7.405)

As usual, we transform into the rotating frame of the laser field, say, by moving the excited states down in energy by $\hbar\omega$, the time dependence of the interaction goes away:

$$\tilde{H}_{\rm AF} = \frac{\hbar}{2}\sum_q \left[\Omega_q^* \Sigma_q + \Omega_q \Sigma_q^\dagger\right].$$

(atom–field interaction, rotating frame) (7.406)

We can similarly write out the interaction Hamiltonian in terms of all pairs of coupled sublevels as

$$\tilde{H}_{\rm AF} = \frac{\hbar}{2}\sum_{m_{\rm g}m_{\rm e}} \left[\Omega^*(m_{\rm g},m_{\rm e})\,\sigma(m_{\rm g},m_{\rm e}) + \Omega(m_{\rm g},m_{\rm e})\,\sigma^\dagger(m_{\rm g},m_{\rm e})\right],$$

(atom–field interaction, rotating frame) (7.407)

where

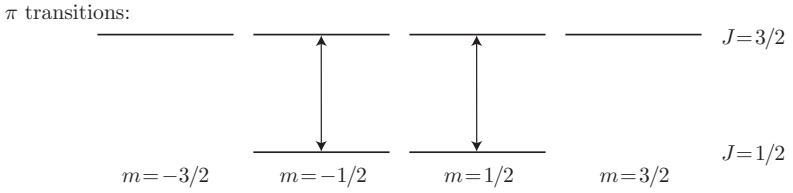
$$\Omega(m_{\rm g},m_{\rm e}) := \langle J_{\rm g}\ m_{\rm g}|J_{\rm e}\ m_{\rm e};1\ -(m_{\rm e}-m_{\rm g})\rangle\ \Omega_{-(m_{\rm e}-m_{\rm g})}.$$

(sublevel Rabi frequencies) (7.408)

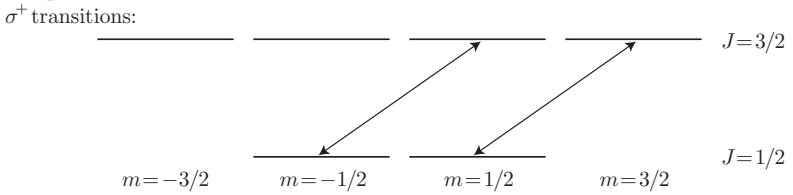
is the Rabi frequency for the $|J_{\rm g}\ m_{\rm g}\rangle \longrightarrow |J_{\rm e}\ m_{\rm e}\rangle$ sublevel transition in terms of the vector Rabi frequency.

7.5.3.1 Magnetic-Sublevel Transitions: Notation

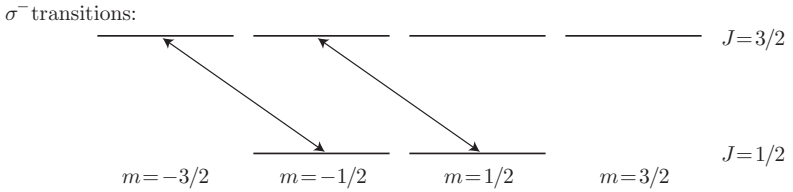
To illustrate the couplings in the above dipole Hamiltonian, we will consider as an example the possible transitions occurring in a $J_{\rm g} = 1/2 \longrightarrow J_{\rm e} = 3/2$ transition. Recall that the selection rules in this case dictate that m can change by at most 1 in a dipole-allowed transition. The transitions where $\Delta m = 0$ are referred to as **π transitions**, and since $q = -(m_{\rm e} - m_{\rm g})$, these transitions correspond to $q = 0$ and are thus coupled by the $q = 0$ polarization of the electric field (i.e., by linearly polarized light along the $\hat{z}$-direction). Correspondingly these transitions are coupled by the Σ_0 operator. There are two possible π transitions in this example atom, shown here.



The transitions where $m_{\rm e} = m_{\rm g} + 1$ are referred to as **σ^+ transitions**, and they correspond to $q = -1$ and are thus coupled by the $q = -1$ polarization of the electric field (i.e., by circularly polarized light in the x-y plane). Correspondingly these transitions are coupled by the Σ_{-1} operator. There are two possible σ^+ transitions in this example atom, shown here.



Finally, the transitions where $m_{\rm e} = m_{\rm g} - 1$ are referred to as **σ^- transitions**, and they correspond to $q = +1$ and are thus coupled by the $q = +1$ polarization of the electric field (i.e., by circularly polarized light in the x-y plane, but rotating in the opposite sense to σ^- light). Correspondingly these transitions are coupled by the Σ_{+1} operator. There are of course two possible σ^- transitions in this example atom, shown here.



To reiterate, because the notation is a bit strange: π transitions are coupled by the $E_0 = E_z$ component of the field, or linear polarization along $\hat{z}$. (Of course, z here defines the ''quantization axis'' for the angular-momentum states, since m is the quantum number for J_z.) The $\sigma^{\pm}$ transitions are coupled by the $E_{\mp 1}$ components, respectively, of the electric field, corresponding to the two circular polarizations orthogonal to $\hat{z}$.

7.5.4 Dipole Interaction: Hyperfine Structure

7.5.4.1 Atomic Hyperfine Hamiltonian

The treatment of hyperfine structure is much the same as for fine structure, but there is some extra complication in typically having more hyperfine levels to deal with and in further decomposing the dipole matrix elements. We will consider transitions from a manifold of ground hyperfine levels to a manifold of excited hyperfine levels. The relevant hyperfine levels are again essentially determined by the quantum numbers of the fine-structure transition $J_\mathrm{g} \longrightarrow J_\mathrm{e}$ (where dipole constraint $|J_\mathrm{e} - J_\mathrm{g}| \leq 1$ is satisfied) as well as the nuclear quantum number I. We still assume the fine-structure transition to be closed, but we include all possible hyperfine ground states $|J_\mathrm{g} - I| \leq F_\mathrm{g} \leq J_\mathrm{g} + I$ and excited states $|J_\mathrm{e} - I| \leq F_\mathrm{e} \leq J_\mathrm{e} + I$, noting that only transitions satisfying $|F_\mathrm{e} - F_\mathrm{g}| \leq 1$ will occur.

The atomic Hamiltonian is

$$H_\mathrm{A} = \hbar \sum_{F_\mathrm{g} m_\mathrm{g}} \delta\omega_{F_\mathrm{g}} |F_\mathrm{g}\ m_\mathrm{g}\rangle\langle F_\mathrm{g}\ m_\mathrm{g}| + \hbar \sum_{m_\mathrm{e}} (\omega_0 + \delta\omega_{F_\mathrm{e}}) |F_\mathrm{e}\ m_\mathrm{e}\rangle\langle F_\mathrm{e}\ m_\mathrm{e}|, \tag{7.409}$$

if we assume degenerate magnetic sublevels within each hyperfine level. Here ω_0 is the transition frequency, which we can choose, for example, to correspond to the frequency difference between the centers of gravity of the hyperfine manifolds. Then $\delta\omega_{F_\mathrm{g}}$ and $\delta\omega_{F_\mathrm{e}}$ are the hyperfine shifts from each respective center of gravity. (Alternately, ω_0 could be chosen to correspond to a particular hyperfine transition of interest such as a laser-cooling transition, with the $\delta\omega_{F_\mathrm{g}}$ and $\delta\omega_{F_\mathrm{e}}$ corresponding to shifts from these levels within each manifold.) Again, with a coupling to a field of frequency ω, it is convenient to work within the rotating frame of the field, where the Hamiltonian is

$$\tilde{H}_\mathrm{A} = \hbar \sum_{F_\mathrm{g} m_\mathrm{g}} \delta\omega_{F_\mathrm{g}} |F_\mathrm{g}\ m_\mathrm{g}\rangle\langle F_\mathrm{g}\ m_\mathrm{g}| + \hbar \sum_{m_\mathrm{e}} (\delta\omega_{F_\mathrm{e}} - \Delta) |F_\mathrm{e}\ m_\mathrm{e}\rangle\langle F_\mathrm{e}\ m_\mathrm{e}|, \tag{7.410}$$

where $\Delta := \omega - \omega_0$ is the usual field detuning from the atomic resonance.

7.5.4.2 Atom–Field Interaction

Then the appropriate projection operators require summations over the hyperfine quantum numbers F_e and F_g as well as the sublevel indices m_e and m_g:

$$P_\mathrm{e} := \sum_{F_\mathrm{e} m_\mathrm{e}} |F_\mathrm{e}\ m_\mathrm{e}\rangle\langle F_\mathrm{e}\ m_\mathrm{e}|, \qquad P_\mathrm{g} := \sum_{F_\mathrm{g} m_\mathrm{g}} |F_\mathrm{g}\ m_\mathrm{g}\rangle\langle F_\mathrm{g}\ m_\mathrm{g}|. \tag{7.411}$$

Then again in our restricted Hilbert space, $P_\mathrm{e} + P_\mathrm{g}$ is the identity, and we can again write the dipole operator as

$$\begin{aligned} d_q &= (P_\mathrm{e} + P_\mathrm{g}) d_q (P_\mathrm{e} + P_\mathrm{g}) \\ &= P_\mathrm{g} d_q P_\mathrm{e} + P_\mathrm{e} d_q P_\mathrm{g} \\ &= d_q^{(+)} + d_q^{(-)}, \end{aligned} \tag{7.412}$$

since the dipole operator does not couple states within the same hyperfine manifold. Then the positive-rotating part becomes

$$\begin{aligned} d_q^{(+)} &= P_{\rm g} d_q P_{\rm e} \\ &= \sum_{F_{\rm e} m_{\rm e} F_{\rm g} m_{\rm g}} \langle F_{\rm g}\ m_{\rm g}|d_q|F_{\rm e}\ m_{\rm e}\rangle\ |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}| \\ &= \sum_{F_{\rm e} m_{\rm e} F_{\rm g} m_{\rm g}} \langle F_{\rm g}\|\mathbf{d}\|F_{\rm e}\rangle\langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle\ |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}| \\ &= \sum_{F_{\rm e} m_{\rm e} F_{\rm g} m_{\rm g}} \langle J_{\rm g}\|\mathbf{d}\|J_{\rm e}\rangle(-1)^{F_{\rm e}+J_{\rm g}+1+I}\sqrt{(2F_{\rm e}+1)(2J_{\rm g}+1)}\left\{\begin{array}{ccc} J_{\rm g} & J_{\rm e} & 1 \\ F_{\rm e} & F_{\rm g} & I \end{array}\right\} \\ &\qquad \times\langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle\ |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}|, \end{aligned} \tag{7.413}$$

where we have used the Wigner–Eckart theorem (7.242) and the decomposition rule (7.282). Clearly, $m_{\rm g} = m_{\rm e} + q$ must still hold for the nonvanishing terms in the sum. Similarly, we may verify directly using Eqs. (7.91), (7.259), and (7.281), that

$$\begin{aligned} d_q^{(-)} &= P_{\rm e} d_q P_{\rm g} \\ &= \sum_{F_{\rm e} m_{\rm e} F_{\rm g} m_{\rm g}} \langle J_{\rm e}\|\mathbf{d}\|J_{\rm g}\rangle(-1)^{F_{\rm g}+J_{\rm e}+1+I}\sqrt{(2F_{\rm g}+1)(2J_{\rm e}+1)}\left\{\begin{array}{ccc} J_{\rm e} & J_{\rm g} & 1 \\ F_{\rm g} & F_{\rm e} & I \end{array}\right\} \\ &\qquad \times\langle F_{\rm e}\ m_{\rm e}|F_{\rm g}\ m_{\rm g};1\ q\rangle\ |F_{\rm e}\ m_{\rm e}\rangle\langle F_{\rm g}\ m_{\rm g}| \\ &= \sum_{F_{\rm e} m_{\rm e} F_{\rm g} m_{\rm g}} \langle J_{\rm g}\|\mathbf{d}\|J_{\rm e}\rangle(-1)^{F_{\rm e}+J_{\rm g}+1+I}(-1)^{q}\sqrt{(2F_{\rm e}+1)(2J_{\rm g}+1)}\left\{\begin{array}{ccc} J_{\rm g} & J_{\rm e} & 1 \\ F_{\rm e} & F_{\rm g} & I \end{array}\right\} \\ &\qquad \times\langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ -q\rangle\ |F_{\rm e}\ m_{\rm e}\rangle\langle F_{\rm g}\ m_{\rm g}| \\ &= (-1)^q\left(d_q^{(+)}\right)^\dagger. \end{aligned} \tag{7.414}$$

We can then indicate the rather complicated dependence in the dipole operator here by defining as before the weighted lowering operator

$$\begin{aligned} \Sigma_q &:= \sum_{F_{\rm g} m_{\rm g} F_{\rm e} m_{\rm e}} (-1)^{F_{\rm e}+J_{\rm g}+1+I}\sqrt{(2F_{\rm e}+1)(2J_{\rm g}+1)}\ \langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle\left\{\begin{array}{ccc} J_{\rm e} & J_{\rm g} & 1 \\ F_{\rm g} & F_{\rm e} & I \end{array}\right\}\ |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}| \\ &= \sum_{F_{\rm g} m_{\rm g} F_{\rm e} m_{\rm e}} (-1)^{F_{\rm e}+J_{\rm g}+1+I}\sqrt{(2F_{\rm e}+1)(2J_{\rm g}+1)}\ \langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle\left\{\begin{array}{ccc} J_{\rm e} & J_{\rm g} & 1 \\ F_{\rm g} & F_{\rm e} & I \end{array}\right\}\ \sigma(F_{\rm g}, m_{\rm g}; F_{\rm e}, m_{\rm e}), \end{aligned}$$

(hyperfine lowering operator) (7.415)

where the individual lowering operators are

$$\sigma(F_{\rm g}, m_{\rm g}; F_{\rm e}, m_{\rm e}) := |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}|. \tag{7.416}$$

We can then write the dipole operator in the same form as for the fine-structure transition as

$$\begin{aligned} d_q &= d_q^{(+)} + d_q^{(-)} \\ &= \langle J_{\rm g}\|d\|J_{\rm e}\rangle\left(\Sigma_q + (-1)^q\Sigma_{-q}^\dagger\right). \end{aligned}$$

(hyperfine dipole operator) (7.417)

Then the dipole-interaction Hamiltonian becomes

$$\begin{aligned} H_{\text{AF}} &= -\mathbf{d}^{(+)} \cdot \mathbf{E}^{(-)} - \mathbf{d}^{(-)} \cdot \mathbf{E}^{(+)} \\ &= -\sum_q (-1)^q \left(d_q^{(+)} E_{-q}^{(-)} + d_q^{(-)} E_{-q}^{(+)} \right) \\ &= -\sum_q \langle J_{\text{g}} \| \mathbf{d} \| J_{\text{e}} \rangle \left[(-1)^q E_{-q}^{(-)}(t)\, \Sigma_q + E_{-q}^{(+)}(t)\, \Sigma_{-q}^{\dagger} \right], \end{aligned} \tag{7.418}$$

and defining the same vector Rabi frequency as before,

$$\Omega_q := -\frac{2\langle J_{\text{g}} \| \mathbf{d} \| J_{\text{e}} \rangle E_q^{(+)}(0)}{\hbar}, \tag{7.419}$$

(vector hyperfine Rabi frequency)

we switch to the rotating frame of the laser field and write the atom–field interaction in the same form as before:

$$\tilde{H}_{\text{AF}} = \frac{\hbar}{2} \sum_q \left[\Omega_q^* \Sigma_q + \Omega_q \Sigma_q^{\dagger} \right]. \tag{7.420}$$

(atom–field interaction, hyperfine transition)

The same form, that is, except that the raising and lowering operators are considerably more complicated. Putting these in explicitly yields the expression

$$\tilde{H}_{\text{AF}} = \frac{\hbar}{2} \sum_{F_{\text{g}} m_{\text{g}} F_{\text{e}} m_{\text{e}}} \left[\Omega^*(F_{\text{g}}, m_{\text{g}}; F_{\text{e}}, m_{\text{e}})\, \sigma(F_{\text{g}}, m_{\text{g}}; F_{\text{e}}, m_{\text{e}}) + \Omega(F_{\text{g}}, m_{\text{g}}; F_{\text{e}}, m_{\text{e}})\, \sigma^{\dagger}(F_{\text{g}}, m_{\text{g}}; F_{\text{e}}, m_{\text{e}}) \right], \tag{7.421}$$

(hyperfine atom–field interaction, rotating frame)

where

$$\begin{aligned} &\Omega(F_{\text{g}}, m_{\text{g}}; F_{\text{e}}, m_{\text{e}}) \\ &\quad := (-1)^{F_{\text{e}}+J_{\text{g}}+1+I} \sqrt{(2F_{\text{e}}+1)(2J_{\text{g}}+1)}\, \langle F_{\text{g}}\, m_{\text{g}} | F_{\text{e}}\, m_{\text{e}}; 1\, -(m_{\text{e}} - m_{\text{g}}) \rangle \begin{Bmatrix} J_{\text{e}} & J_{\text{g}} & 1 \\ F_{\text{g}} & F_{\text{e}} & I \end{Bmatrix} \Omega_{-(m_{\text{e}}-m_{\text{g}})}. \end{aligned} \tag{7.422}$$

(sublevel Rabi frequencies)

is the Rabi frequency for the $|F_{\text{g}}\, m_{\text{g}}\rangle \longrightarrow |F_{\text{e}}\, m_{\text{e}}\rangle$ hyperfine sublevel transition in terms of the vector Rabi frequency.

7.6 Angular Distribution of Dipolar Resonance Fluorescence

Recall from Eq. (5.256) that the scattered intensity can be represented in terms of the scattered field as

$$I(\mathbf{r}, \omega) = \frac{1}{\pi\eta} \int_{-\infty}^{\infty} \left\langle \mathbf{E}^{(-)}(\mathbf{r}, t) \cdot \mathbf{E}^{(+)}(\mathbf{r}, t+\tau) \right\rangle e^{i\omega\tau}\, d\tau, \tag{7.423}$$

when written in terms of the vector electric field. Recall that the classical field due to an oscillating dipole in the radiation zone is

$$\mathbf{E}^{(+)}(\mathbf{r}, t) = \frac{1}{4\pi\epsilon_0 c^2} \left[(\hat{\varepsilon} \cdot \hat{r})\hat{r} - \hat{\varepsilon} \right] \frac{\ddot{d}^{(+)}(t_{\text{r}})}{r}, \tag{7.424}$$

for a dipole orientation $\hat{\varepsilon}$. This carries over to the quantum atom here, as we are still treating the field classically. However, we need to be careful in the spherical basis. Specifically, in the above notation, we mean

$$\left[(\hat{\varepsilon} \cdot \hat{r})\, \hat{r} - \hat{\varepsilon} \right] d^{(+)} = \left[\left(\mathbf{d}^{(+)} \cdot \hat{r} \right) \hat{r} - \mathbf{d}^{(+)} \right], \tag{7.425}$$

and in the spherical basis, the dipole operator may be resolved into components as

$$\mathbf{d}^{(+)} = \sum_q \hat{\text{e}}_q^* d_q = \sum_q \hat{\varepsilon}_q^* d_q, \tag{7.426}$$

if we choose to represent the polarization-vector components as in terms of the usual basis vectors

$$\hat{\varepsilon}_q = \hat{\mathrm{e}}_q. \tag{7.427}$$

Thus, the polarization vector becomes

$$\left[\left(\mathbf{d}^{(+)}\cdot\hat{r}\right)\hat{r} - \mathbf{d}^{(+)}\right] = \left[(\hat{\varepsilon}_q^*\cdot\hat{r})\hat{r} - \hat{\varepsilon}_q^*\right] d^{(+)}, \tag{7.428}$$

and so the radiated electric field becomes

$$\mathbf{E}^{(+)}(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0 c^2}\sum_q \left[(\hat{\varepsilon}_q^*\cdot\hat{r})\hat{r} - \hat{\varepsilon}_q^*\right]\frac{\ddot{d}_q^{(+)}(t_\mathrm{r})}{r} \tag{7.429}$$

in terms of the dipole-vector components. We can then label the radiated field due to each dipole-vector component as

$$\mathbf{E}_q^{(+)}(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0 c^2}\left[(\hat{\varepsilon}_q^*\cdot\hat{r})\hat{r} - \hat{\varepsilon}_q^*\right]\frac{\ddot{d}_q^{(+)}(t_\mathrm{r})}{r}, \tag{7.430}$$

where the polarization vectors are again

$$\hat{\varepsilon}_0 = \hat{\mathrm{e}}_0 = \hat{z}, \qquad \hat{\varepsilon}_{\pm1} = \hat{\mathrm{e}}_{\pm1} = \mp(\hat{x}\pm i\hat{y})/\sqrt{2} \tag{7.431}$$

for linear and circular polarizations, respectively. That is, $\hat{\varepsilon}_q$ is just the unit spherical basis vector $\hat{\mathrm{e}}_q$. Note that $\mathbf{E}_q$ is not a spherical vector, but d_q is, so we should be careful to observe that

$$\mathbf{E}_q^{(-)}(\mathbf{r},t) = \frac{1}{4\pi\epsilon_0 c^2}\left[(\hat{\varepsilon}_q\cdot\hat{r})\hat{r} - \hat{\varepsilon}_q\right]\frac{(-1)^{-q}\ddot{d}_{-q}^{(-)}(t_\mathrm{r})}{r}, \tag{7.432}$$

to ensure that we appropriately conjugate the dipole operator.

The fields that appear in (7.423) are, as we defined them, sums of all three component fields, and so we may write the sum explicitly as

$$I(\mathbf{r},\omega) = \frac{1}{\pi\eta}\sum_{qq'}\int_{-\infty}^{\infty}\left\langle \mathbf{E}_q^{(-)}(\mathbf{r},t)\cdot\mathbf{E}_{q'}^{(+)}(\mathbf{r},t+\tau)\right\rangle e^{i\omega\tau}\,d\tau. \tag{7.433}$$

Paralleling our previous treatment in Section 5.7, we may now write the scattered intensity as

$$I_\mathrm{sc}(\mathbf{r},\omega) = \frac{\omega_0^4}{6\pi^2\epsilon_0 c^3 r^2}\sum_{qq'} f_{qq'}(\theta,\phi)\,(-1)^q\int_{-\infty}^{\infty}\left\langle d_{-q}^{(-)}(\mathbf{r},t)\,d_{q'}^{(+)}(\mathbf{r},t+\tau)\right\rangle e^{i\omega\tau}\,d\tau, \tag{7.434}$$

where we have defined the angular-distribution tensor

$$\begin{aligned} f_{qq'}(\theta,\phi) &:= \frac{3}{8\pi}\left[(\hat{\varepsilon}_q\cdot\hat{r})\hat{r} - \hat{\varepsilon}_q\right]\cdot\left[(\hat{\varepsilon}_{q'}^*\cdot\hat{r})\hat{r} - \hat{\varepsilon}_{q'}^*\right] \\ &= \frac{3}{8\pi}\left[\hat{\varepsilon}_q\cdot\hat{\varepsilon}_{q'}^* - (\hat{\varepsilon}_q\cdot\hat{r})(\hat{\varepsilon}_{q'}^*\cdot\hat{r})\right] \end{aligned} \tag{7.435}$$

We will now reduce this to a more explicit form in terms of the angle variables θ and ϕ.

7.6.1 Angular-Distribution Tensor

We can first use the orthogonality relation $\hat{\varepsilon}_q\cdot\hat{\varepsilon}_{q'}^* = \delta_{qq'}$ to evaluate the first term in Eq. (7.435). For the second, we can use the relation

$$\hat{\varepsilon}_q\cdot\hat{r} = \sqrt{\frac{4\pi}{3}}\,Y_1^q(\theta,\phi) \tag{7.436}$$

for the projection of $\hat{r}$ into basis vectors in terms of the spherical harmonics, which can be verified directly (but certainly makes sense since Y_1^q is the right tensor for representing three-vectors). Along with the conjugation rule

$$[Y_\ell^m(\theta,\phi)]^* = (-1)^m Y_\ell^{-m}(\theta,\phi), \tag{7.437}$$

we can write the angular scattering tensor as

$$f_{qq'}(\theta,\phi) = \frac{3}{8\pi}\left[\delta_{qq'} - \frac{4\pi}{3}(-1)^{q'}\, Y_1^q(\theta,\phi)\, Y_1^{-q'}(\theta,\phi)\right]. \tag{7.438}$$

To reduce this yet more, we may use the recoupling relation (7.187), which in this special case reads

$$Y_1^q(\theta,\phi)\, Y_1^{-q'}(\theta,\phi) = \sum_{\ell=0}^{2}(-1)^{q'-q}\sqrt{\frac{9(2\ell+1)}{4\pi}}\begin{pmatrix} 1 & 1 & \ell \\ q & -q' & q'-q \end{pmatrix}\begin{pmatrix} 1 & 1 & \ell \\ 0 & 0 & 0 \end{pmatrix} Y_\ell^{q-q'}(\theta,\phi), \tag{7.439}$$

Writing the terms out explicitly and using the 3-j symbols

$$\begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \sqrt{\frac{2}{15}}; \quad \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = 0; \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \sqrt{\frac{1}{3}}; \quad \begin{pmatrix} 1 & 1 & 0 \\ q & -q' & 0 \end{pmatrix} = -\frac{(-1)^q}{\sqrt{3}}\delta_{qq'}, \tag{7.440}$$

as well as the symmetry rule

$$\begin{pmatrix} J_1 & J_2 & J \\ m_1 & m_2 & m \end{pmatrix} = (-1)^{J_1+J_2+J}\begin{pmatrix} J_1 & J_2 & J \\ -m_1 & -m_2 & -m \end{pmatrix}, \tag{7.441}$$

we find

$$Y_1^q(\theta,\phi)\, Y_1^{-q'}(\theta,\phi) = \frac{(-1)^q}{4\pi}\delta_{qq'} + \sqrt{\frac{3}{2\pi}}(-1)^{q-q'}\begin{pmatrix} 1 & 1 & 2 \\ -q & q' & q-q' \end{pmatrix} Y_\ell^{q-q'}(\theta,\phi). \tag{7.442}$$

Putting these together, we find the form

$$f_{qq'}(\theta,\phi) = \frac{1}{4\pi}\left[\delta_{qq'} - \sqrt{6\pi}\,(-1)^q\, Y_2^{q-q'}(\theta,\phi)\begin{pmatrix} 1 & 1 & 2 \\ -q & q' & q-q' \end{pmatrix}\right], \tag{7.443}$$

(angular-scattering tensor)

for the scattering tensor. Written out explicitly, the components are shown in the following table.

$f_{qq'}(\theta,\phi)$		q'		
		-1	0	1
q	-1	$\frac{3}{16\pi}(1+\cos^2\theta)$	$-\frac{3}{\sqrt{2}\,16\pi}\sin 2\theta e^{-i\phi}$	$\frac{3}{16\pi}\sin^2\theta e^{-i2\phi}$
	0	$-\frac{3}{\sqrt{2}\,16\pi}\sin 2\theta e^{i\phi}$	$\frac{3}{8\pi}\sin^2\theta$	$\frac{3}{\sqrt{2}\,16\pi}\sin 2\theta e^{-i\phi}$
	1	$\frac{3}{16\pi}\sin^2\theta e^{i2\phi}$	$\frac{3}{\sqrt{2}\,16\pi}\sin 2\theta e^{i\phi}$	$\frac{3}{16\pi}(1+\cos^2\theta)$

(7.444)

7.6.2 Spectral Tensor and Total Scattered Power

We now continue with the intensity spectrum by writing the dipole-operator components using Eq. (7.398), assuming we are dealing with a single $J \longrightarrow J'$ transition. Then defining the spectral tensor

$$S_{qq'}(\omega_s) := \frac{1}{2\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\omega_s\tau}\left\langle \Sigma_q^\dagger(t)\,\Sigma_{q'}(t+\tau)\right\rangle, \tag{7.445}$$

we can write the intensity spectrum (7.434) as

$$I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \frac{\hbar\omega_0\Gamma}{r^2}\left(\frac{2J_{\rm e}+1}{2J_{\rm g}+1}\right)\sum_{qq'} f_{qq'}(\theta,\phi)S_{qq'}(\omega_{\rm s}). \tag{7.446}$$

Here again, from Eq. (7.396), Σ_q is the lowering operator for all transitions coupled by polarization q, corresponding to the radiative decay $|J'\,m'\rangle \longrightarrow |J\,m'+q\rangle$.

The expression (7.446) for the intensity spectrum is valid in the resonant approximation for a single $J \longrightarrow J'$ Zeeman-degenerate transition. In principle, it should then be summed over all possible transitions (accounting for the various detunings in each case).

Physically, the diagonal terms with $q = q'$ in Eq. (7.446) correspond to the radiation of a purely oriented dipole, that is, with an orientation $\hat{\varepsilon} = \hat{\rm e}_q^*$. In general, the atom will oscillate in some mixture of the various components, and the off-diagonal terms with $q \neq q'$ in the sum represent *interference* due to coherence between the different dipole components. These interference terms do not change the total radiated power, but rather they change the angular *distribution* of the radiated power, as we can see by integrating the scattered intensity over all angles:

$$P_{\rm sc}(\omega_{\rm s}) = \int d\Omega\, r^2 I_{\rm sc}(\mathbf{r},\omega_{\rm s}) = \hbar\omega_0\Gamma\left(\frac{2J_{\rm e}+1}{2J_{\rm g}+1}\right)\sum_q S_{qq}(\omega_{\rm s}). \tag{7.447}$$

Here, we have used

$$\int d\Omega\, f_{qq'}(\theta,\phi) = \delta_{qq'}, \tag{7.448}$$

which follows from Eq. (7.443), where the second term always vanished under the angular integration in view of the orthogonality of the spherical harmonics.

Of course, integrating the power spectrum over all frequencies gives

$$P_{\rm sc} = \int_0^\infty d\omega_{\rm s}\, P_{\rm sc}(\omega_{\rm s}) = \hbar\omega_0\Gamma\left(\frac{2J_{\rm e}+1}{2J_{\rm g}+1}\right)\sum_q \left\langle \Sigma_q^\dagger\,\Sigma_q\right\rangle. \tag{7.449}$$

We can compute the operator sum as

$$\begin{aligned}
\sum_q \Sigma_q^\dagger\Sigma_q &= \sum_{q m_{\rm g} m_{\rm e} m'_{\rm e}} \langle J_{\rm g}\; m_{\rm g}|J_{\rm e}\; m'_{\rm e};1\; q\rangle\; \langle J_{\rm g}\; m_{\rm g}|J_{\rm e}\; m_{\rm e};1\; q\rangle\; |J_{\rm e}\; m'_{\rm e}\rangle\langle J_{\rm e}\; m_{\rm e}| \\
&= \sum_{q m_{\rm g} m_{\rm e}} \langle J_{\rm g}\; m_{\rm g}|J_{\rm e}\; m_{\rm e};1\; q\rangle^2\; |J_{\rm e}\; m_{\rm e}\rangle\langle J_{\rm e}\; m_{\rm e}| \\
&= \left(\frac{2J_{\rm g}+1}{2J_{\rm e}+1}\right)\sum_{q m_{\rm g} m_{\rm e}} \langle J_{\rm e}\; m_{\rm e}|J_{\rm g}\; m_{\rm g};1\; -q\rangle^2\; |J_{\rm e}\; m_{\rm e}\rangle\langle J_{\rm e}\; m_{\rm e}| \\
&= \left(\frac{2J_{\rm g}+1}{2J_{\rm e}+1}\right)\sum_{m_{\rm e}} |J_{\rm e}\; m_{\rm e}\rangle\langle J_{\rm e}\; m_{\rm e}|,
\end{aligned} \tag{7.450}$$

where we used the conservation constraint $m_{\rm g} = m'_{\rm e} + q = m_{\rm e} + q$ to get to the second expression, and we used completeness of the angular-momentum states to get to the last expression. Thus, we may rewrite Eq. (7.449) as

$$P_{\rm sc} = \hbar\omega_0\Gamma\langle P_{\rm e}\rangle, \tag{7.451}$$

where recall that $P_{\rm e}$, defined in Eq. (7.390), is the sum over projection operators for all excited sublevels, so that $\langle P_{\rm e}\rangle$ is the total excited-level population. Thus, the total photon scattering rate is sensibly given by Γ times the total excited-state population.

7.6.2.1 Hyperfine Structure and Interference

Of course, this assumes that only one excited level J_e is substantially populated. To see this, note that in the case of a hyperfine transition, the formulae in the last section are valid, except that the lowering operators Σ_q are given by the hyperfine expression (7.415), and thus we have

$$\begin{aligned}
\sum_q \Sigma_q^\dagger \Sigma_q &= \sum_{q F_\mathrm{g} m_\mathrm{g} F_\mathrm{e} m_\mathrm{e} F_\mathrm{e}'} (-1)^{F_\mathrm{e}-F_\mathrm{e}'} \sqrt{(2F_\mathrm{e}'+1)(2F_\mathrm{e}+1)}\,(2J_\mathrm{g}+1) \\
&\qquad \times \langle F_\mathrm{g}\, m_\mathrm{g} | F_\mathrm{e}'\, m_\mathrm{e}; 1\, q\rangle \, \langle F_\mathrm{g}\, m_\mathrm{g} | F_\mathrm{e}\, m_\mathrm{e}; 1\, q\rangle \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e}' & I \end{array} \right\} \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e} & I \end{array} \right\} |F_\mathrm{e}'\, m_\mathrm{e}\rangle\langle F_\mathrm{e}\, m_\mathrm{e}| \\
&= \sum_{q F_\mathrm{g} m_\mathrm{g} F_\mathrm{e} m_\mathrm{e} F_\mathrm{e}'} (-1)^{2F_\mathrm{e}-2F_\mathrm{g}+2m_\mathrm{e}-2m_\mathrm{g}} (2F_\mathrm{g}+1)\,(2J_\mathrm{g}+1) \\
&\qquad \times \langle F_\mathrm{e}'\, m_\mathrm{e} | F_\mathrm{g}\, m_\mathrm{g}; 1\, -q\rangle \, \langle F_\mathrm{e}\, m_\mathrm{e} | F_\mathrm{g}\, m_\mathrm{g}; 1\, -q\rangle \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e}' & I \end{array} \right\} \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e} & I \end{array} \right\} |F_\mathrm{e}'\, m_\mathrm{e}\rangle\langle F_\mathrm{e}\, m_\mathrm{e}| \\
&= \sum_{F_\mathrm{g} F_\mathrm{e} m_\mathrm{e} F_\mathrm{e}'} (2F_\mathrm{g}+1)\,(2J_\mathrm{g}+1)\,\delta_{F_\mathrm{e}F_\mathrm{e}'} \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e}' & I \end{array} \right\} \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e} & I \end{array} \right\} |F_\mathrm{e}'\, m_\mathrm{e}\rangle\langle F_\mathrm{e}\, m_\mathrm{e}| \\
&= \sum_{F_\mathrm{g} F_\mathrm{e} m_\mathrm{e}} (2F_\mathrm{g}+1)\,(2J_\mathrm{g}+1) \left\{ \begin{array}{ccc} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e} & I \end{array} \right\}^2 |F_\mathrm{e}\, m_\mathrm{e}\rangle\langle F_\mathrm{e}\, m_\mathrm{e}|.
\end{aligned} \tag{7.452}$$

In the second step we used the Clebsch–Gordan symmetry relations [see the Wigner–Eckart theorem in the forms of Eq. (7.281)], and then we used the orthogonality relation (7.46). Then using the orthogonality relation (7.82), we find

$$\sum_q \Sigma_q^\dagger \Sigma_q = \left(\frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1} \right) \sum_{F_\mathrm{e} m_\mathrm{e}} |F_\mathrm{e}\, m_\mathrm{e}\rangle\langle F_\mathrm{e}\, m_\mathrm{e}|, \tag{7.453}$$

which is essentially the same result as for the fine-structure case above: this operator is diagonal in the hyperfine basis, and every state decays at the same rate Γ. Note that it is important to work this out to see if the fluorescence operator (7.453) contains *off-diagonal* terms, representing interference between two states with the same m_e, but different F_e. This is possible in principle because such decays are *indistinguishable*: if the atom starts in a superposition of these states, we can't tell by the decay which state the atom "came from." This is true even if the states are not degenerate, since, for example, with a steady-state drive, the dipoles corresponding to $|F_\mathrm{g}\, m_\mathrm{g}\rangle \longrightarrow |F_\mathrm{e}\, m_\mathrm{e}\rangle, |F_\mathrm{e}'\, m_\mathrm{e}'\rangle$ oscillate at the same frequency. Of course, decay from states with *different* m_e don't interfere: we can infer "which-state" information by analyzing the polarization of the fluorescence and the final state of the atom to determine what the initial atomic state was. These interferences correspond to quantum beats between particular excited states, as we discussed before in Section (6.2.4). Evidently, however, while such interferences may influence the decay rates to *individual* ground states, and hence the angular distribution of light, they do not affect the *total* decay rate from any excited state.

7.7 Optical Stark Shifts

Now we would like to handle the Stark shifts due to laser light. In terms of hyperfine states, we may write the Kramers–Heisenberg formula for the polarizability tensor (assuming linear polarization) as see Eq. (14.147)

$$\alpha_{\mu\nu}(F, m_F; \omega) = \sum_{F' m_F'} \frac{2\omega_{F'F} \langle F\, m_F | d_\nu | F'\, m_F'\rangle \langle F'\, m_F' | d_\mu | F\, m_F\rangle}{\hbar(\omega_{F'F}^2 - \omega^2)}. \tag{7.454}$$

(hyperfine polarizability tensor, linear polarization)

For circular polarization, the formula for the polarizability tensor is

$$\alpha^{(1)}_{\mu\nu}(F, m_F; \omega) = \sum_{F' m'_F} \frac{2\omega \langle F\ m_F|d_\nu|F'\ m'_F\rangle\langle F'\ m'_F|d_\mu|F\ m_F\rangle}{\hbar(\omega^2_{F'F} - \omega^2)}.$$

(hyperfine polarizability tensor, circular polarization) (7.455)

The tensor polarizability is defined such that to lowest order (i.e., for weak field intensities), the mean induced dipole moment vector is

$$\left\langle d^{(+)}_\mu(\omega)\right\rangle = \alpha_{\mu\nu}(\omega)(E^{(+)}_0)_\nu, \tag{7.456}$$

and thus according to the electric-dipole interaction, the energy shift (ac Stark shift) due to the optical field is [cf. Eq. (1.65)]

$$\begin{aligned}\Delta E(F, m_F; \omega) &= -\frac{1}{2}\left\langle \mathbf{d}^{(+)}(\omega)\right\rangle \cdot \mathbf{E}^{(-)} - \frac{1}{2}\left\langle \mathbf{d}^{(-)}(\omega)\right\rangle \cdot \mathbf{E}^{(+)} \\ &= -\mathrm{Re}[\alpha_{\mu\nu}(F, m_F; \omega)](E^{(-)}_0)_\mu (E^{(+)}_0)_\nu.\end{aligned} \tag{7.457}$$

(ac Stark shift)

In principle, this is the expression we're after, but now we will break this expression down into parts according to its symmetry and express the result explicitly in terms of the quantum numbers of the states.

7.7.1 Irreducible Parts

Given that the polarizability is a rank-2 tensor, it is convenient to decompose it into its irreducible parts. To simplify notation, we will write the polarizability as

$$\alpha_{\mu\nu}(F, m_F; \omega) = \sum_{F'} \frac{2\omega_{F'F} T_{\mu\nu}}{\hbar(\omega^2_{F'F} - \omega^2)} \qquad \alpha^{(1)}_{\mu\nu}(F, m_F; \omega) = \sum_{F'} \frac{2\omega T_{\mu\nu}}{\hbar(\omega^2_{F'F} - \omega^2)}, \tag{7.458}$$

where we have defined the dipole-product tensor

$$T_{\mu\nu} := \sum_{m'_F} \langle F\ m_F|d_\mu|F'\ m'_F\rangle\langle F'\ m'_F|d_\nu|F\ m_F\rangle. \tag{7.459}$$

Note that we include the sum over m'_F here to avoid any orientation dependence in this tensor, since this is what we will decompose into its irreducible components. Recall from Eqs. (7.199), (7.200), (7.206), and (7.207) that we may write $T_{\mu\nu}$ in terms of its scalar, vector, and tensor parts as

$$\begin{aligned} T_{\mu\nu} &= \frac{1}{3}T^{(0)}\delta_{\mu\nu} + \frac{1}{4}T^{(1)}_\sigma \epsilon_{\sigma\mu\nu} + T^{(2)}_{\mu\nu} \\ T^{(0)} &= T_{\mu\mu} \\ T^{(1)}_\sigma &= \epsilon_{\sigma\mu\nu}(T_{\mu\nu} - T_{\nu\mu}) \\ T^{(2)}_{\mu\nu} &= T_{(\mu\nu)} - \frac{1}{3}T_{\sigma\sigma}\delta_{\mu\nu}. \end{aligned} \tag{7.460}$$

We will handle each irreducible component here separately.[42]

7.7.1.1 Scalar Part

The scalar part is simply the trace,

$$\begin{aligned} T^{(0)} &= T_{\mu\mu} \\ &= \sum_{m'_F} \langle F\ m_F|d_\mu|F'\ m'_F\rangle\langle F'\ m'_F|d_\mu|F\ m_F\rangle \\ &= |\langle F\|\mathbf{d}\|F\rangle|^2. \end{aligned} \tag{7.461}$$

[42] The treatment here more or less follows Ivan H. Deutsch and Poul S. Jessen, "Quantum measurement and dynamics of atomic spins in polarization spectroscopy," (2007, to be published). However, the notation here is somewhat different, as it is designed to be close to that of the dc case.

where we used the normalization formula (7.241) for the reduced matrix element. That was simple enough, but let's redo this a more complicated way to see a couple details that will make the other components easier to calculate. First, recall that when two rank-1 spherical tensors (vectors) $\mathbf{A}$ and $\mathbf{B}$ are multiplied to form a rank-0 (scalar) tensor, the result from Eq. (7.214) is

$$T_0^{(0)} = -\frac{1}{\sqrt{3}}\mathbf{A}\cdot\mathbf{B}. \tag{7.462}$$

It is important to keep the extra overall factors to make use of factorization formulae for matrix elements. Thus, we can rewrite the dipole-vector dot product as a rank-0 spherical tensor and use the Wigner–Eckart theorem (7.237) to find

$$\begin{aligned}
T^{(0)} &= \sum_{m'_F}\langle F\ m_F|d_\mu|F'\ m'_F\rangle\langle F'\ m'_F|d_\mu|F\ m_F\rangle \\
&= -\sqrt{3}\,\langle F\ m_F|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\langle F'\ m'_F|\mathbf{d}\right)^{(0)}|F\ m_F\rangle \\
&= -\sqrt{3}\,\langle F\|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\langle F'\ m'_F|\mathbf{d}\right)^{(0)}\|F\rangle\,\langle F\ m_F|F\ m_F;0\ 0\rangle \\
&= -\sqrt{3}\,(-1)^{2F}\sqrt{2F'+1}\begin{Bmatrix} 1 & 1 & 0 \\ F & F & F'\end{Bmatrix}\langle F\|\mathbf{d}\|F'\rangle\,\langle F'\|\mathbf{d}\|F\rangle \\
&= -\sqrt{3}\,(-1)^{F+F'}\sqrt{2F+1}\begin{Bmatrix} 1 & 1 & 0 \\ F & F & F'\end{Bmatrix}|\langle F\|\mathbf{d}\|F'\rangle|^2 \\
&= |\langle F\|\mathbf{d}\|F'\rangle|^2,
\end{aligned} \tag{7.463}$$

where we used the factorization formula (7.273) for the first matrix element, and then the conjugation formula (7.259) for the last reduced matrix element. We also used $\langle F\ m_F|F\ m_F;0\ 0\rangle = 1$, and the value of the 6-j symbol is $(-1)^{-F-F'-1}/\sqrt{3(2F+1)}$.

7.7.1.2 Vector Part

The vector part is related to the cross product of the dipole vectors.

$$T_\sigma^{(1)} = \epsilon_{\sigma\mu\nu}(T_{\mu\nu} - T_{\nu\mu}) = 2\epsilon_{\sigma\mu\nu}T_{\mu\nu}. \tag{7.464}$$

In the case of the dc Stark shift, this component vanished due to the symmetry of the tensor Hamiltonian; however, here this component does not necessarily vanish, because the electric fields and dipole vectors may be complex (corresponding to circular polarization, which is meaningless in the dc limit). Again, to use the Wigner–Eckart theorem and the subsequent decomposition formula, we must express the vector product of the dipole operators as a rank-1 spherical tensor. From Eq. (7.217), we can express this in terms of the vector cross product as

$$T_q^{(1)} = \frac{i}{\sqrt{2}}(\mathbf{A}\times\mathbf{B})_q, \tag{7.465}$$

where Roman indices to indicate spherical components and Greek indices to indicate Cartesian components. The procedure is otherwise as outlined for the scalar case, with the result

$$\begin{aligned} T_q^{(1)} &= i2\sqrt{2}\,\langle F\ m_F|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\times\langle F'\ m'_F|\mathbf{d}\right)^{(1)}_q|F\ m_F\rangle \\ &= i2\sqrt{2}\,\langle F\|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\times\langle F'\ m'_F|\mathbf{d}\right)^{(1)}\|F\rangle\,\langle F\ m_F|F\ m_F;1\ q\rangle \\ &= -i(-1)^{2F}\frac{2\sqrt{2}}{\sqrt{F(F+1)}}\sqrt{3(2F'+1)}\left\{\begin{array}{ccc}1&1&1\\F&F&F'\end{array}\right\}\langle F\|\mathbf{d}\|F'\rangle\,\langle F'\|\mathbf{d}\|F\rangle\,m_F\,\delta_{q0} \\ &= (-1)^{F+F'}(-i)\sqrt{\frac{24(2F+1)}{F(F+1)}}\left\{\begin{array}{ccc}1&1&1\\F&F&F'\end{array}\right\}|\langle F\|\mathbf{d}\|F'\rangle|^2\,m_F\,\delta_{q0}, \end{aligned} \tag{7.466}$$

where we used $\langle F\ m_F|F\ m_F;1\ q\rangle = \delta_{q0}\,m_F/\sqrt{F(F+1)}$.

7.7.1.3 Tensor Part

Finally, the tensor part is

$$T^{(2)}_{\mu\nu} = T_{(\mu\nu)} - \frac{1}{3}T_{\sigma\sigma}\delta_{\mu\nu}. \tag{7.467}$$

Converted to the spherical basis, there is no scaling factor in this case, and thus, with the usual procedure,

$$\begin{aligned} T_q^{(2)} &= \langle F\ m_F|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\langle F'\ m'_F|\mathbf{d}\right)^{(2)}|F\ m_F\rangle \\ &= \langle F\|\left(\sum_{m'_F}\mathbf{d}|F'\ m'_F\rangle\langle F'\|\mathbf{d}\right)^{(2)}\|F\rangle\,\langle F\ m_F|F\ m_F;2\ 0\rangle \\ &= (-1)^{F+F'}\sqrt{\frac{5(2F+1)}{F(F+1)(2F-1)(2F+3)}}\left\{\begin{array}{ccc}1&1&2\\F&F&F'\end{array}\right\}|\langle F\|\mathbf{d}\|F'\rangle|^2\,[m_F^2 - F(F+1)]\,\delta_{q0}, \end{aligned} \tag{7.468}$$

where we used $\langle F\ m_F|F\ m_F;2\ 0\rangle = \delta_{q0}\,[m_F^2 - F(F+1)]/\sqrt{F(F+1)(2F-1)(2F+3)}$.

7.7.2 Total Shift

Now we can write the total ac Stark shift (7.457) using the polarizability tensor (7.458), along with Eq. (7.460), we can write

$$\begin{aligned} \Delta E(F,m_F;\omega) = -\sum_{F'}&\frac{2}{\hbar(\omega_{F'F}^2-\omega^2)} \\ &\times\left\{\frac{\omega_{F'F}}{3}T^{(0)}\left[\mathbf{E}_0^{(-)}\mathbf{E}_0^{(+)}\right]^{(0)} + \frac{\omega}{4}T_0^{(1)}(-i\sqrt{2})\left[\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)}\right]^{(1)}_0 + \omega_{F'F}T_0^{(2)}\left[\mathbf{E}_0^{(-)}\mathbf{E}_0^{(+)}\right]^{(2)}_0\right\}. \end{aligned} \tag{7.469}$$

Note that the tensor products in spherical form are particularly simple because only the $q=0$ components are involved. Writing out the relevant field components,

$$\begin{aligned} \Delta E(F,m_F;\omega) = -\sum_{F'}&\frac{2}{\hbar(\omega_{F'F}^2-\omega^2)} \\ &\times\left\{\frac{\omega_{F'F}}{3}T^{(0)}|E_0^{(+)}|^2 + \frac{\omega}{4}T_0^{(1)}(\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)})_z + \frac{\omega_{F'F}}{\sqrt{6}}T_0^{(2)}\left(3|E_{0z}^{(+)}|^2 - |E_0^{(+)}|^2\right)\right\}. \end{aligned} \tag{7.470}$$

Now using Eqs. (7.463), (7.466), and (7.468) for the irreducible tensors, we can write the shift as

$$\begin{aligned}\Delta E(F,m_F;\omega) = &-\alpha^{(0)}(F;\omega)\,|E_0^{(+)}|^2 - \alpha^{(1)}(F;\omega)\,(i\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)})_z\frac{m_F}{F}\\ &-\alpha^{(2)}(F;\omega)\,\frac{\left(3|E_{0z}^{(+)}|^2-|E_0^{(+)}|^2\right)}{2}\left(\frac{3m_F^2-F(F+1)}{F(2F-1)}\right),\end{aligned} \tag{7.471}$$

(ac Stark shift)

where we have defined the scalar, vector, and tensor polarizabilities as

$$\begin{aligned}\alpha^{(0)}(F;\omega) &= \sum_{F'}\frac{2\omega_{F'F}|\langle F\|\mathbf{d}\|F'\rangle|^2}{3\hbar(\omega_{F'F}^2-\omega^2)}\\ \alpha^{(1)}(F;\omega) &= \sum_{F'}(-1)^{F+F'+1}\sqrt{\frac{6F(2F+1)}{F+1}}\left\{\begin{array}{ccc}1&1&1\\F&F&F'\end{array}\right\}\frac{\omega|\langle F\|\mathbf{d}\|F'\rangle|^2}{\hbar(\omega_{F'F}^2-\omega^2)}\\ \alpha^{(2)}(F;\omega) &= \sum_{F'}(-1)^{F+F'}\sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}}\left\{\begin{array}{ccc}1&1&2\\F&F&F'\end{array}\right\}\frac{\omega_{F'F}|\langle F\|\mathbf{d}\|F'\rangle|^2}{\hbar(\omega_{F'F}^2-\omega^2)},\end{aligned}$$

(scalar, vector, and tensor polarizabilities) (7.472)

respectively. As in the dc case, the scalar shift causes a level-independent shift, while the tensor shift has the same quadratic dependence on m_F. The vector field causes a shift linear in m_F, which has the same form as a weak-field Zeeman shift. We will see below that the vector shift is "activated" by circular polarizations, and thus in terms of the level shift, circularly polarized light acts as an effective magnetic field. Linearly polarized light drives the tensor shift, and thus acts as an effective dc electric field.

Also as in the dc case, we have chosen the normalization of the tensor polarizability such that the maximum shift (where $\mathbf{E}_0^{(+)} = E_0^{(+)}\hat{z}$) for $m_F = F$ has the form $-\alpha^{(2)}(F;\omega)|E_0^{(+)}|^2$. Indeed, both the scalar and tensor polarizabilities here reduce to the respective dc polarizabilities (7.370) and (7.371) in the dc limit $\omega = 0$. The vector polarizability is similarly normalized such that the maximum shift for $m_F = F$ and σ^+ polarization, $E_0^{(+)} = (E_0^{(+)})_{-1}$, has the same form $-\alpha^{(1)}(F;\omega)|E_0^{(+)}|^2$. To see this, we can write out the field-vector cross product as

$$\begin{aligned}(i\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)})_z &= i\mathbf{E}_{0x}^{(-)}\times\mathbf{E}_{0y}^{(+)} - i\mathbf{E}_{0y}^{(-)}\times\mathbf{E}_{0x}^{(+)}\\ &= \frac{1}{2}\left(\mathbf{E}_{0,1}^{(-)}-\mathbf{E}_{0,-1}^{(-)}\right)\left(\mathbf{E}_{0,1}^{(-)}+\mathbf{E}_{0,-1}^{(-)}\right) - \frac{1}{2}\left(\mathbf{E}_{0,1}^{(-)}+\mathbf{E}_{0,-1}^{(-)}\right)\left(\mathbf{E}_{0,1}^{(-)}-\mathbf{E}_{0,-1}^{(-)}\right)\\ &= \mathbf{E}_{0,1}^{(-)}\mathbf{E}_{0,-1}^{(+)} - \mathbf{E}_{0,-1}^{(-)}\mathbf{E}_{0,1}^{(+)}\\ &= |\mathbf{E}_{0,-1}^{(+)}|^2 - |\mathbf{E}_{0,1}^{(+)}|^2,\end{aligned} \tag{7.473}$$

so that σ^+ ($E_{0,-1}^{(+)}$) and σ_- ($E_{0,1}^{(+)}$) light lead to contributions of opposite sign.

7.7.2.1 Excited States

Note that these formulae also apply to the ac Stark shifts of *excited* states, so long as $\omega_{F'F} = \omega_{F'} - \omega_F$ is interpreted with the proper sign: the excited state of a two-level atom has opposite Stark shifts for the two levels, and the sign of $\omega_{F'F}$ keeps appropriate track of this in the sum over all levels.

7.7.3 Example: Stark Shifts of the $F = 1 \longrightarrow F' = 0$ Transition

As a simple example of the formalism we have presented so far, we will consider an $F = 1 \longrightarrow F' = 0$ transition, such that the ground level has three sublevels, but there is only one excited level. We will assume the field to be sufficiently close to resonance with this transitions that other terms in the polarizability sum are negligible in comparison. In this case, we may drop the summations over F' in Eqs. (7.472) and write

the polarizabilities as

$$\begin{aligned}
\alpha^{(0)}(F,F';\omega) &= \frac{2\omega_{F'F}|\langle F\|\mathbf{d}\|F'\rangle|^2}{3\hbar(\omega_{F'F}^2-\omega^2)} \approx -\frac{|\langle F\|\mathbf{d}\|F'\rangle|^2}{3\hbar\Delta_{F'F}} \\
\alpha^{(1)}(F,F';\omega) &= (-1)^{F+F'+1}\sqrt{\frac{27F(2F+1)}{2(F+1)}}\begin{Bmatrix} 1 & 1 & 1 \\ F & F & F' \end{Bmatrix}\alpha^{(0)}(F,F';\omega) \\
\alpha^{(2)}(F,F';\omega) &= (-1)^{F+F'}\sqrt{\frac{30F(2F+1)(2F-1)}{(F+1)(2F+3)}}\begin{Bmatrix} 1 & 1 & 2 \\ F & F & F' \end{Bmatrix}\alpha^{(0)}(F,F';\omega),
\end{aligned}$$

(polarizabilities for single hyperfine transition) (7.474)

where $\Delta_{F'F} := \omega - \omega_{F'F}$, and we have neglected the counterrotating term in the scalar polarizability (i.e., making the rotating-wave approximation), as is consistent with the two-level approximation. Plugging in $F=1$, $F'=0$, we can write the vector and tensor polarizabilities for this example as

$$\begin{aligned}
\alpha^{(1)}(F,F';\omega) &= -\frac{3}{2}\alpha^{(0)}(F,F';\omega) \\
\alpha^{(2)}(F,F';\omega) &= -\alpha^{(0)}(F,F';\omega),
\end{aligned} \tag{7.475}$$

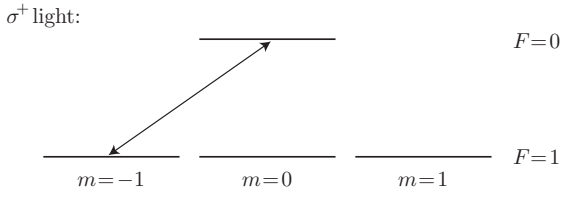
and thus the ac Stark shift (7.474) becomes

$$\begin{aligned}
\Delta E(F=1,m_F;\omega) &= -\alpha^{(0)}(F=1,F'=0;\omega)\Big\{|E_0^{(+)}|^2 - \frac{3}{2}(i\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)})_z m_F \\
&\qquad - \left(3|E_{0z}^{(+)}|^2 - |E_0^{(+)}|^2\right)\left(\frac{3}{2}m_F^2 - 1\right)\Big\} \\
&= -\alpha^{(0)}(F=1,F'=0;\omega)\Big\{|E_0^{(+)}|^2 - \frac{3}{2}\left(|E_{0,-1}^{(+)}|^2 - |E_{0,1}^{(+)}|^2\right)m_F \\
&\qquad - \left(3|E_{0z}^{(+)}|^2 - |E_0^{(+)}|^2\right)\left(\frac{3}{2}m_F^2 - 1\right)\Big\},
\end{aligned} \tag{7.476}$$

where we have explicitly written the dependence on the three spherical components in the last expression. For circularly polarized light, say σ^+ light, with $E_{0,-1}^{(+)} = E_0^{(+)}$, this expression reduces to

$$\Delta E(F=1,m_F;\omega) = -3\alpha^{(0)}(F=1,F'=0;\omega)|E_0^{(+)}|^2\left\{\frac{m_F(m_F-1)}{2}\right\}. \tag{7.477}$$

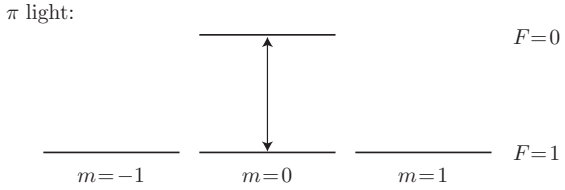
Note that this shift is zero unless $m_F=-1$, where the bracketed expression is unity, which is consistent with our expectation from noting that the $m_F=-1$ state is the only state coupled to the excited state by σ^+-polarized light.



For linearly (π) polarized light, where $E_{0z}^{(+)} = E_0^{(+)}$, the ac Stark shift (7.476) becomes

$$\Delta E(F=1,m_F;\omega) = -3\alpha^{(0)}(F=1,F'=0;\omega)|E_0^{(+)}|^2\{1-m_F^2\}, \tag{7.478}$$

which vanishes unless $m_F=0$, when the bracketed factor is again unity. This is again in accordance with our expectation that the only level coupled to the excited level by the π-polarized light is Stark shifted.



Note that the *magnitudes* of the Stark shifts in both cases are the same, which is a consequence of the simple level structure of this transition.

7.7.4 Polarizability Tensor Revisited

As in the dc case, the expression (7.471) for the ac Stark shift allows us to write down an effective Hamiltonian for the *ac* Stark shift:

$$H_{\text{Stark}}(F;\omega) = -\alpha^{(0)}(F;\omega)\,|E_0^{(+)}|^2 - \alpha^{(1)}(F;\omega)\,(i\mathbf{E}_0^{(-)}\times\mathbf{E}_0^{(+)})_z\frac{F_z}{F} - \alpha^{(2)}(F;\omega)\,\frac{\left(3|E_{0z}^{(+)}|^2 - |E_0^{(+)}|^2\right)}{2}\left(\frac{3F_z^2-\mathbf{F}^2}{F(2F-1)}\right).$$

(effective ac Stark Hamiltonian) (7.479)

Here, F_z and $\mathbf{F}^2$ are the operators, an obviously this Hamiltonian applies to a single hyperfine level, where we have defined the polarizabilities as in (7.472), and the energy shift for the level $|F\ m_F\rangle$ is given by the expectation value of the effective Hamiltonian,

$$\Delta E_{|F\ m_F\rangle}(\omega) = \langle F\ m_F|H_{\text{Stark}}(F;\omega)|F\ m_F\rangle.$$

(shift in terms of effective Hamiltonian) (7.480)

Another, basis-independent way to write the Stark shift comes about if we define the **tensor polarizability operator**[43]

$$\alpha_{\mu\nu}(F;\omega) = \alpha^{(0)}(F;\omega)\,\delta_{\mu\nu} + \alpha^{(1)}(F;\omega)\,i\epsilon_{\sigma\mu\nu}\frac{F_\sigma}{F} + \alpha^{(2)}(F;\omega)\,\frac{3}{F(2F-1)}\left[\frac{1}{2}(F_\mu F_\nu + F_\nu F_\mu) - \frac{1}{3}\mathbf{F}^2\,\delta_{\mu\nu}\right],$$

(tensor polarizability operator) (7.481)

such that the effective Hamiltonian (7.479) becomes

$$H_{\text{Stark}}(F;\omega) = -\alpha_{\mu\nu}(F;\omega)\,E_{0\mu}^{(-)}E_{0\nu}^{(+)}.$$

(effective ac Stark Hamiltonian) (7.482)

Note that the operator parts of the polarizability in terms of $\mathbf{F}$ are the scalar, vector, and tensor reductions of $\mathbf{FF}$; when computing the expectation value with respect to the $|F\ m_F\rangle$ state, the Wigner–Eckart theorem guarantees that only the $q = 0$ spherical-tensor components are projected out, in agreement with the basis-dependent expression (7.479).

7.7.5 Large Detuning

A crucial aspect of the ac polarizabilities lies in the relative detunings of the optical field with respect to the relevant atomic transitions. In particular, depending on the detunings involved, cancellations may occur. To see examples of this, we can take the polarizabilities (7.472) and factor the reduced matrix elements based

[43] Deutsch and Jessen, *op. cit.*

on the decoupling rule (7.282):

$$\begin{aligned}
\alpha^{(0)}(F;\omega) &= \sum_{F'} \frac{2\omega_{F'F}|\langle J\|\mathbf{d}\|J'\rangle|^2}{3\hbar(\omega_{F'F}^2-\omega^2)}(2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I \end{array}\right\}^2 \\
\alpha^{(1)}(F;\omega) &= \sum_{F'}(-1)^{F+F'+1}\sqrt{\frac{6F(2F+1)}{F+1}}\left\{\begin{array}{ccc} 1 & 1 & 1 \\ F & F & F' \end{array}\right\}\frac{\omega|\langle J\|\mathbf{d}\|J'\rangle|^2}{\hbar(\omega_{F'F}^2-\omega^2)} \\
&\qquad \times (2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I \end{array}\right\}^2 \\
\alpha^{(2)}(F;\omega) &= \sum_{F'}(-1)^{F+F'}\sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}}\left\{\begin{array}{ccc} 1 & 1 & 2 \\ F & F & F' \end{array}\right\}\frac{\omega_{F'F}|\langle J\|\mathbf{d}\|J'\rangle|^2}{\hbar(\omega_{F'F}^2-\omega^2)} \\
&\qquad \times (2F'+1)(2J+1)\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I \end{array}\right\}^2 .
\end{aligned} \tag{7.483}$$

In the limit of large detunings compared to the hyperfine splittings, we can use $\omega_{F'F} \approx \omega_{J'J}$, so that the only dependence on F' in the sums is in the sign and the 6-j symbols. We can then use the orthogonality relation (7.82) in the scalar case and the Biedenharn–Elliott sum rule (7.102) in the forms [the second expression has the form of (7.375)]

$$\begin{aligned}
\sum_{F'}(-1)^{F'}(2F'+1)\left\{\begin{array}{ccc} 1 & 1 & 1 \\ F & F & F' \end{array}\right\}\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I \end{array}\right\}^2 &= (-1)^{-(2J+J'+2F+I+1)}\left\{\begin{array}{ccc} 1 & 1 & 1 \\ J & J & J' \end{array}\right\}\left\{\begin{array}{ccc} J & J & 1 \\ F & F & I \end{array}\right\} \\
\sum_{F'}(-1)^{F'}(2F'+1)\left\{\begin{array}{ccc} 1 & 1 & 2 \\ F & F & F' \end{array}\right\}\left\{\begin{array}{ccc} J & J' & 1 \\ F' & F & I \end{array}\right\}^2 &= (-1)^{-(2J+J'+2F+I)}\left\{\begin{array}{ccc} 1 & 1 & 2 \\ J & J & J' \end{array}\right\}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I \end{array}\right\}
\end{aligned} \tag{7.484}$$

for the vector and tensor cases, with the somewhat simpler result

$$\begin{aligned}
\alpha^{(0)}(F;\omega) &\approx \sum_{J'}\frac{2\omega_{J'J}|\langle J\|\mathbf{d}\|J'\rangle|^2}{3\hbar(\omega_{J'J}^2-\omega^2)} \\
\alpha^{(1)}(F;\omega) &\approx \sum_{J'}(-1)^{-2J-J'-F-I}\sqrt{\frac{6F(2F+1)}{F+1}}(2J+1)\frac{\omega|\langle J\|\mathbf{d}\|J'\rangle|^2}{\hbar(\omega_{J'J}^2-\omega^2)}\left\{\begin{array}{ccc} 1 & 1 & 1 \\ J & J & J' \end{array}\right\}\left\{\begin{array}{ccc} J & J & 1 \\ F & F & I \end{array}\right\} \\
\alpha^{(2)}(F;\omega) &\approx \sum_{J'}(-1)^{-2J-J'-F-I}\sqrt{\frac{40F(2F+1)(2F-1)}{3(F+1)(2F+3)}}(2J+1)\frac{\omega_{J'J}|\langle J\|\mathbf{d}\|J'\rangle|^2}{\hbar(\omega_{J'J}^2-\omega^2)} \\
&\qquad \times\left\{\begin{array}{ccc} 1 & 1 & 2 \\ J & J & J' \end{array}\right\}\left\{\begin{array}{ccc} J & J & 2 \\ F & F & I \end{array}\right\}.
\end{aligned}$$

(far-detuned polarizabilities) (7.485)

Thus, as in the dc case, the hyperfine polarizabilities may be expressed for large detunings directly in terms of the fine-structure dipole matrix elements.

7.7.5.1 Effective Dipole Moment

Returning to the ac Stark shift (7.471), the scalar part is

$$\Delta E^{(0)}(F,m_F;\omega) = -\alpha^{(0)}(F;\omega)\,|E_0^{(+)}|^2. \tag{7.486}$$

For detunings large compared to the hyperfine splitting, we can use Eq. (7.485) for the scalar shift so that

$$\Delta E^{(0)}(F,m_F;\omega) = -\sum_{J'}\frac{2\omega_{J'J}|\langle J\|\mathbf{d}\|J'\rangle|^2}{3\hbar(\omega_{J'J}^2-\omega^2)}\,|E_0^{(+)}|^2. \tag{7.487}$$

Suppose that although the detuning is large compared to the hyperfine splitting, only one fine-structure level J' is dominantly resonant. Then we can keep only one term in the sum and make the rotating-wave

approximation via $\omega_{J'J}^2 - \omega^2 \approx 2\omega_{J'J}(\omega_{J'J} - \omega)$:

$$\Delta E^{(0)}(F, m_F; \omega) = \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{3\hbar\Delta_{J'J}} |E_0^{(+)}|^2. \tag{7.488}$$

Again, $\Delta_{J'J} := \omega - \omega_{J'J}$ is the detuning of the optical field from the atomic transition. Comparing (7.487) to the ground-state Stark shift from the two-level atom from Section 5.8, Eq. (5.415),

$$\Delta E = \frac{\hbar|\Omega|^2}{4\Delta}, \tag{7.489}$$

where Ω is the local Rabi frequency

$$\Omega := -\frac{2\langle \mathrm{g}|\hat{\varepsilon}\cdot\mathbf{d}|\mathrm{e}\rangle E_0^{(+)}}{\hbar}, \tag{7.490}$$

we see that the expressions are equivalent provided we identify

$$|\langle \mathrm{g}|\hat{\varepsilon}\cdot\mathbf{d}|\mathrm{e}\rangle|^2 = \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{3}, \tag{7.491}$$

or simply

$$\langle \mathrm{g}|\hat{\varepsilon}\cdot\mathbf{d}|\mathrm{e}\rangle = \frac{\langle J\|\mathbf{d}\|J'\rangle}{\sqrt{3}}. \tag{7.492}$$

(effective dipole moment, large detuning)

The interpretation here is that the field interacts directly with the *fine*-structure transition if the hyperfine shifts are negligible. The factor of $1/3$ simply comes from representing the dipole as $\mathbf{d}^2 = d_x^2 + d_y^2 + d_z^2 = 3d_z^2$ if the atom is spherically symmetric. Since the polarization vector $\hat{\varepsilon}$ picks out a particular direction, the field interacts with only one of three possible components of the dipole operator, and thus the factor $1/3$. Of course, the scalar polarizability only represents the *average* behavior for the transition, ignoring any vector or tensor (and thus m_F-dependent) shifts. However, as we will see below, the vector and tensor shifts disappear anyway in some important cases for large detunings. In any case, it is best to regard this effective matrix element as being for *linearly* polarized light, where there is no approximation in neglecting the vector shift.

In this case, the saturation intensity is defined as usual by

$$\frac{I}{I_{\mathrm{sat}}} = \frac{2\Omega^2}{\Gamma^2}, \tag{7.493}$$

with $I = 2\epsilon_0 c|E_0^{(+)}|^2$ [Eq. (1.68)], and Ω from Eq. (7.490), so that we have

$$I_{\mathrm{sat}} = \frac{3c\epsilon_0\Gamma^2\hbar^2}{4|\langle J\|\mathbf{d}\|J'\rangle|^2}$$

(saturation intensity, linear polarization, far detuned) (7.494)

for the far-detuned saturation intensity of the fine-structure transition (without a hyperfine-resolved excited level).

For the shifts where the atom–field interaction is dominated by one hyperfine transition, we can instead compare Eq. (7.483) to the two-level atom, so that the effective dipole moment is

$$|\langle \mathrm{g}|\hat{\varepsilon}\cdot\mathbf{d}|\mathrm{e}\rangle|^2 = \frac{|\langle J\|\mathbf{d}\|J'\rangle|^2}{3} S_{FF'}, \tag{7.495}$$

(hyperfine effective dipole moment)

where the hyperfine transition-strength factor is

$$S_{FF'} := (2F'+1)(2J+1)\begin{Bmatrix} J & J' & 1 \\ F' & F & I \end{Bmatrix}^2.$$

(hyperfine relative transition strength) (7.496)

Or, including the proper sign from Eq. (7.282),

$$\langle \mathrm{g}|\hat{\varepsilon}\cdot\mathbf{d}|\mathrm{e}\rangle = (-1)^{F'+J+1+I}\frac{\langle J\|\mathbf{d}\|J'\rangle}{\sqrt{3}}\sqrt{S_{FF'}},$$

(hyperfine effective dipole moment) (7.497)

From the orthogonality relation (7.82), we have

$$\sum_{F'} S_{FF'} = 1.$$

(hyperfine strength sum rule) (7.498)

Thus, the factor $S_{FF'}$ acts as a ''relative oscillator strength'' for the hyperfine transitions in a particular fine-structure line. Again, except in certain special cases, this effective dipole moment only captures orientation-averaged behavior, as for example happens with excitation by isotropic light.

7.7.5.2 Alkali Ground States

The expressions above are still rather complicated, so we can get a bit more insight by considering the specific case of the ground state of alkali atoms, where $L = 0$ and $J = 1/2$. Using the 6-j symbols

$$\begin{aligned}
\left\{\begin{array}{ccc} 1 & 1 & 2 \\ 1/2 & 1/2 & J' \end{array}\right\} &= 0 \\
\left\{\begin{array}{ccc} 1 & 1 & 2 \\ 1/2 & 1/2 & J' \end{array}\right\} &= -\frac{2}{3(2J+1)} \\
\left\{\begin{array}{ccc} J & J & 1 \\ F & F & I \end{array}\right\} &= (-1)^{-(F+I+J+1)}\frac{F(F+1)-I(I+1)+J(J+1)}{2\sqrt{F(F+1)(2F+1)J(J+1)(2J+1)}},
\end{aligned} \tag{7.499}$$

the ground-state polarizabilities become

$$\begin{aligned}
\alpha^{(0)}(F;\omega) &\approx \sum_{J'}\frac{2\omega_{J'J}|\langle J=1/2\|\mathbf{d}\|J'\rangle|^2}{3\hbar(\omega_{J'J}^2-\omega^2)} \\
\alpha^{(1)}(F;\omega) &\approx \sum_{J'}(-1)^{J'+3/2}\sqrt{\frac{6}{J(J+1)(2J+1)}}F\,g_F\,\frac{\omega}{\omega_{F'F}}\,\alpha^{(0)}(F;\omega) \\
\alpha^{(2)}(F;\omega) &\approx 0.
\end{aligned}$$

(far-detuned polarizabilities) (7.500)

Here, we have used Eq. (7.319) for the Landé g_F factor

$$g_F \approx g_J\frac{F(F+1)-I(I+1)+J(J+1)}{2F(F+1)}, \tag{7.501}$$

ignoring the term proportional to g_I. Thus, in the regime of large detuning compared to the excited-state hyperfine splitting—when the excited states are effectively degenerate—the tensor component of the ground-state polarizability vanishes. We expect this from the dc case (corresponding to $\omega = 0$ here), where the tensor polarizability vanished for $J = 0$ or $J = 1/2$. Note, however, that the vector polarizability remains in this regime.

In the yet-farther-detuned regime, where the two excited *fine*-structure states $J' = 1/2, 3/2$ are effectively degenerate, the two terms in the vector polarizability become $-2Fg_F\alpha^{(0)}(F;\omega)$ for $J' = 1/2$ and $2Fg_F\alpha^{(0)}(F;\omega)$ for $J' = 3/2$. These two contributions cancel, and thus in this regime

$$\alpha^{(1)}(F;\omega) \approx 0.$$

(far-detuned vector polarizability) (7.502)

Again, we also expected this for the dc case, where the vector polarizability vanished for *any* atomic configuration. (This is because the dc field must be real, and thus the $\sigma^{\pm}$ components must always be present with equal weight.) For the alkali-atom ground state, in the regime of far detuning compared to the fine-structure aplitting, the shift is purely scalar.

7.8 Atomic Master Equation

7.8.1 Fine Structure

To consider a fine structure transition $|J_\mathrm{g}\rangle \longrightarrow |J_\mathrm{e}\rangle$, we will of course need to consider all possible transitions between sublevels $|J_\mathrm{g}\ m_\mathrm{g}\rangle \longrightarrow |J_\mathrm{e}\ m_\mathrm{e}\rangle$. We showed in Section (7.3.7.4) that the decay rate (7.305) of a fine-structure transition is

$$\Gamma = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}\frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1}|\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2, \tag{7.503}$$

while the rate for the $|J_\mathrm{e}\ m_\mathrm{e}\rangle \longrightarrow |J_\mathrm{g}\ m_\mathrm{g}\rangle$ decay process is [Eq. (11.38)]

$$\Gamma_{m_\mathrm{g},m_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}|\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2, \tag{7.504}$$

which using the Wigner–Eckart theorem (7.242) gives

$$\Gamma_{m_\mathrm{g},m_\mathrm{e}} = |\langle J_\mathrm{e}\ m_\mathrm{e}|J_\mathrm{g}\ m_\mathrm{g};1\ (m_\mathrm{e}-m_\mathrm{g})\rangle|^2\,\Gamma. \tag{7.505}$$

However, note that *any* decay corresponding to emission into polarization q is indistinguishable as far as measurement of the radiated field is concerned. Thus, *amplitudes* for such decays should be added together, while decay *rates* for decays of different polarization should be added together. We have already concluded this from our above discussion of the spontaneous decay, where in particular from Eq. (7.449) we may conclude that the rate at which photons are scattered is

$$R_\mathrm{sc} = \Gamma\left(\frac{2J_\mathrm{e}+1}{2J_\mathrm{g}+1}\right)\sum_q\left\langle\Sigma_q^\dagger\,\Sigma_q\right\rangle. \tag{7.506}$$

This expression has precisely the form we want. The master equation that properly accomplishes this decay along with the appropriate Hamiltonian evolution is

$$\partial_t\tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_\mathrm{A}+\tilde{H}_\mathrm{AF},\tilde{\rho}\right]+\Gamma\left(\frac{2J_\mathrm{e}+1}{2J_\mathrm{g}+1}\right)\sum_q\mathcal{D}[\Sigma_q]\tilde{\rho},$$

(master equation, fine-structure transition) (7.507)

where $\tilde{H}_\mathrm{A}$ is defined in Eq. (7.389), where $\tilde{H}_\mathrm{AF}$ is defined in Eq. (7.407), Σ_q is defined in Eq. (7.396) as

$$\Sigma_q = \sum_{m_\mathrm{g}m_\mathrm{e}}\langle J_\mathrm{g}\ m_\mathrm{g}|J_\mathrm{e}\ m_\mathrm{e};1\ q\rangle\ |J_\mathrm{g}\ m_\mathrm{g}\rangle\langle J_\mathrm{e}\ m_\mathrm{e}|, \tag{7.508}$$

and Γ is the total decay rate of any excited sublevel. To verify the action of the decay term here, consider the evolution of the matrix elements $\tilde{\rho}_{\alpha\ m_\alpha,\beta\ m_\beta} \equiv \langle J_\alpha\ m_\alpha|\tilde{\rho}|J_\beta\ m_\beta\rangle$ due only to the decay term:

$$\begin{aligned}
&\partial_t\tilde{\rho}_{\alpha\ m_\alpha,\beta\ m_\beta}\\
&\quad=\Gamma\left(\frac{2J_\mathrm{e}+1}{2J_\mathrm{g}+1}\right)\sum_q\langle J_\alpha\ m_\alpha|\left(\Sigma_q\tilde{\rho}\Sigma_q^\dagger-\frac{1}{2}\Sigma_q^\dagger\Sigma_q\tilde{\rho}-\frac{1}{2}\tilde{\rho}\Sigma_q^\dagger\Sigma_q\right)|J_\beta\ m_\beta\rangle\\
&\quad=\Gamma\left(\frac{2J_\mathrm{e}+1}{2J_\mathrm{g}+1}\right)\left[\sum_q\delta_{\alpha\mathrm{g}}\delta_{\mathrm{g}\beta}\langle J_\mathrm{g}\ m_\alpha|J_\mathrm{e}\ (m_\alpha-q);1\ q\rangle\langle J_\mathrm{g}\ m_\beta|J_\mathrm{e}\ (m_\beta-q);1\ q\rangle\tilde{\rho}_{J_\mathrm{e},(m_\alpha-q);J_\mathrm{e},(m_\beta-q)}\right.\\
&\qquad\qquad\left.-\frac{1}{2}\delta_{\alpha\mathrm{e}}\left(\frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1}\right)\tilde{\rho}_{J_\mathrm{e},m_\alpha;J_\beta,m_\beta}-\frac{1}{2}\delta_{\mathrm{e}\beta}\left(\frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1}\right)\tilde{\rho}_{J_\alpha,m_\alpha;J_\mathrm{e},m_\beta}\right]\\
&\quad=\Gamma\left[\sum_q\delta_{\alpha\mathrm{g}}\delta_{\beta\mathrm{g}}\langle J_\mathrm{e}\ (m_\alpha+q)|J_\mathrm{g}\ m_\alpha;1\ q\rangle\langle J_\mathrm{e}\ (m_\beta+q)|J_\mathrm{g}\ m_\beta;1\ q\rangle\tilde{\rho}_{J_\mathrm{e},(m_\alpha+q);J_\mathrm{e},(m_\beta+q)}\right.\\
&\qquad\qquad\left.-\frac{1}{2}(\delta_{\alpha\mathrm{e}}\delta_{\mathrm{g}\beta})\tilde{\rho}_{J_\mathrm{e},m_\alpha;J_\mathrm{g},m_\beta}-\frac{1}{2}(\delta_{\alpha\mathrm{g}}\delta_{\mathrm{e}\beta})\tilde{\rho}_{J_\mathrm{g},m_\alpha;J_\mathrm{e},m_\beta}-(\delta_{\alpha\mathrm{e}}\delta_{\mathrm{e}\beta})\tilde{\rho}_{J_\mathrm{e},m_\alpha;J_\mathrm{e},m_\beta}\right].
\end{aligned} \tag{7.509}$$

Thus, we can see that the excited-state populations (and coherences) decay at rate Γ, excited-ground coherences decay at rate $\Gamma/2$, and ground-state populations (and coherences) increase according to the decay of the excited states and the branching factors (Clebsch–Gordan coefficients) into each state. Working out the similar terms for Hamiltonian evolution is relatively straightforward, and thus we can write out the evolution equations for the density matrix elements according to the master equation (7.507) as

$$\begin{aligned}
\frac{\partial}{\partial t}\tilde{\rho}_{\alpha\, m_\alpha,\, \beta\, m_\beta} = -\frac{i}{2}\Bigg[& \delta_{\alpha \mathrm{e}} \sum_{m_\mathrm{g}} \Omega(m_\mathrm{g}, m_\alpha)\, \tilde{\rho}_{\mathrm{g}\, m_\mathrm{g},\, \beta\, m_\beta} - \delta_{\mathrm{g}\beta} \sum_{m_\mathrm{e}} \Omega(m_\beta, m_\mathrm{e})\, \tilde{\rho}_{\alpha\, m_\alpha,\, \mathrm{e}\, m_\mathrm{e}} \\
& + \delta_{\alpha \mathrm{g}} \sum_{m_\mathrm{e}} \Omega^*(m_\alpha, m_\mathrm{e})\, \tilde{\rho}_{\mathrm{e}\, m_\mathrm{e},\, \beta\, m_\beta} - \delta_{\mathrm{e}\beta} \sum_{m_\mathrm{g}} \Omega^*(m_\mathrm{g}, m_\beta)\, \tilde{\rho}_{\alpha\, m_\alpha,\, \mathrm{g}\, m_\mathrm{g}} \Bigg] \quad \text{(pump field)} \\
& - \delta_{\alpha \mathrm{e}} \delta_{\mathrm{e}\beta}\, \Gamma\, \tilde{\rho}_{\alpha\, m_\alpha,\, \beta\, m_\beta} \\
& - \delta_{\alpha \mathrm{e}} \delta_{\mathrm{g}\beta}\, \frac{\Gamma}{2}\, \tilde{\rho}_{\alpha\, m_\alpha,\, \beta\, m_\beta} \\
& - \delta_{\alpha \mathrm{g}} \delta_{\mathrm{e}\beta}\, \frac{\Gamma}{2}\, \tilde{\rho}_{\alpha\, m_\alpha,\, \beta\, m_\beta} \quad \text{(dissipation)} \\
& + \delta_{\alpha \mathrm{g}} \delta_{\mathrm{g}\beta}\, \Gamma \sum_{q=-1}^{1} \Big[\tilde{\rho}_{\mathrm{e}\, (m_\alpha+q),\, \mathrm{e}\, (m_\beta+q)} \\
& \qquad \langle J_\mathrm{e}\, (m_\alpha+q) | J_\mathrm{g}\, m_\alpha; 1\, q\rangle \langle J_\mathrm{e}\, (m_\beta+q) | J_\mathrm{g}\, m_\beta; 1\, q\rangle \Big] \\
& + \quad i(\delta_{\alpha \mathrm{e}} \delta_{\mathrm{g}\beta} - \delta_{\alpha \mathrm{g}} \delta_{\mathrm{e}\beta})\, \Delta\, \tilde{\rho}_{\alpha\, m_\alpha,\, \beta\, m_\beta}. \quad \text{(free evolution)}
\end{aligned}$$

(master equation, fine-structure transition) (7.510)

The Rabi frequencies $\Omega(m_\mathrm{g}, m_\mathrm{e})$ here were defined before in Eq. (7.408) by

$$\Omega(m_\mathrm{g}, m_\mathrm{e}) := \langle J_\mathrm{g}\, m_\mathrm{g} | J_\mathrm{e}\, m_\mathrm{e}; 1\, -(m_\mathrm{e} - m_\mathrm{g})\rangle\, \Omega_{-(m_\mathrm{e}-m_\mathrm{g})}. \tag{7.511}$$

The form here is rather complicated but is suited for numerical computations.

7.8.2 Hyperfine Structure

For an atom with hyperfine structure, the interaction still occurs between the atom and the atomic dipole, and thus the master equation still has exactly the same form as for fine structure:

$$\partial_t \tilde{\rho} = -\frac{i}{\hbar}\left[\tilde{H}_\mathrm{A} + \tilde{H}_\mathrm{AF}, \tilde{\rho}\right] + \Gamma\left(\frac{2J_\mathrm{e}+1}{2J_\mathrm{g}+1}\right) \sum_q \mathcal{D}[\Sigma_q]\tilde{\rho}.$$

(master equation, fine-structure transition) (7.512)

However, the dipole-related symbols here must be interpreted in terms of hyperfine structure. The atomic Hamiltonian is given in the rotating frame by shifting the center of gravity of the excited state down by $\hbar\omega$, where ω is the laser frequency, combined with the hyperfine energy shifts expressed in Eq. (7.134) in the hyperfine basis,

$$\begin{aligned}
\Delta E_\mathrm{hfs}(J, I, F) = {} & \frac{1}{2} A_\mathrm{hfs} K + B_\mathrm{hfs} \frac{\frac{3}{2}K(K+1) - 2I(I+1)J(J+1)}{4I(2I-1)J(2J-1)} \\
& + C_\mathrm{hfs} \frac{5K^2(K/4+1) + K[I(I+1) + J(J+1) + 3] - 3I(I+1)J(J+1) - 5I(I+1)J(J+1)}{I(I-1)(2I-1)J(J-1)(2J-1)},
\end{aligned} \tag{7.513}$$

so that for a single *fine*-structure transition,

$$\begin{aligned}\tilde{H}_{\rm A} = &\sum_{F_{\rm e}m_{\rm e}} \left[\Delta E_{\rm hfs}(J_{\rm e},I,F_{\rm e}) - \hbar\Delta\right] |J_{\rm e},I;F_{\rm e}\ m_{\rm e}\rangle\langle J_{\rm e},I;F_{\rm e}\ m_{\rm e}| \\ &+ \sum_{F_{\rm g}m_{\rm g}} \Delta E_{\rm hfs}(J_{\rm g},I,F_{\rm g})\, |J_{\rm g},I;F_{\rm g}\ m_{\rm g}\rangle\langle J_{\rm g},I;F_{\rm g}\ m_{\rm g}|,\end{aligned} \tag{7.514}$$

where as usual $\Delta := \omega - \omega_0$, but now ω_0 is the transition frequency for the transition center of gravity (i.e., corresponding to the energy difference in the absence of the hyperfine interaction). The atom–field interaction Hamiltonian in the rotating frame is given in Eq. (7.420) by

$$\tilde{H}_{\rm AF} = \frac{\hbar}{2}\sum_q \left[\Omega_q^*\Sigma_q + \Omega_q\Sigma_q^\dagger\right], \tag{7.515}$$

where the lowering operator from Eq. (7.415) is

$$\begin{aligned}\Sigma_q &= \sum_{F_{\rm g}m_{\rm g}F_{\rm e}m_{\rm e}} (-1)^{F_{\rm e}+J_{\rm g}+1+I}\sqrt{(2F_{\rm e}+1)(2J_{\rm g}+1)}\ \langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle \begin{Bmatrix} J_{\rm e} & J_{\rm g} & 1 \\ F_{\rm g} & F_{\rm e} & I\end{Bmatrix} |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}| \\ &= \sum_{F_{\rm g}m_{\rm g}F_{\rm e}m_{\rm e}} (-1)^{F_{\rm e}+J_{\rm g}+1+I}\sqrt{S_{F_{\rm g}F_{\rm e}}}\ \langle F_{\rm g}\ m_{\rm g}|F_{\rm e}\ m_{\rm e};1\ q\rangle\ |F_{\rm g}\ m_{\rm g}\rangle\langle F_{\rm e}\ m_{\rm e}|,\end{aligned} \tag{7.516}$$

where $S_{F_{\rm g}F_{\rm e}}$ is defined by Eq. (7.496), and the vector Rabi frequency is given in Eq. (7.402) by

$$\Omega_q = -\frac{2\langle J_{\rm g}\|\mathbf{d}\|J_{\rm e}\rangle E_{0q}^{(+)}}{\hbar} \tag{7.517}$$

for a field of the form [cf. Eq. (7.190)]

$$\mathbf{E}(t) = \sum_q (-1)^q \hat{\mathbf{e}}_{-q} E_{0q}^{(+)} e^{i\omega t} + \text{c.c.} \tag{7.518}$$

Often, it is most useful to consider the interaction of *multiple* fields with hyperfine transitions (as in laser cooling of alkali atoms). From Eq. (7.405) we can get the time-dependent form of the interaction Hamiltonian for an additional ''probe'' field of frequency $\omega_{\rm p}$ and transform to the rotating frame by shifting the frequency by ω, obtaining

$$H'_{\rm AF} = \frac{\hbar}{2}\sum_q \left[\Omega_q^*\Sigma_q e^{i\Delta_{\rm p}t} + \Omega_q\Sigma_q^\dagger e^{-i\Delta_{\rm p}t}\right], \tag{7.519}$$

where $\Delta_{\rm p} = \omega_{\rm p} - \omega$ is the detuning of the probe field from the main (''pump'') field. Of course, as many extra fields as necessary may be added in this way.

7.8.3 Rate-Equation Limit

In certain cases, the atom–field master equation (optical Bloch equations) including Zeeman degeneracy may be excessively difficult to solve. This happens, for example, when the detuning from a fine-structure level, or with well-resolved hyperfine structure, where the hyperfine splittings are much larger than the natural linewidths, necessarily leading to large detunings of the field from *some* levels. If we don't care about the fast oscillations due to the large detunings, but choose instead to focus on other, relatively slow dynamics, we can make an adiabatic approximation to obtain a rate-equation formalism to replace the full master equation.

First, we will consider a transition $J_{\rm g} \longrightarrow J_{\rm e}$ between two fine-structure levels, ignoring any hyperfine structure. From Eq. (7.510), the equations of motion for the excited-state populations and Zeeman coherences (coherences between different excited states) are

$$\partial_t \rho_{{\rm e}\ m_{\rm e},{\rm e}\ m'_{\rm e}} = -\frac{i}{2}\sum_{m_{\rm g}} \left[\Omega(m_{\rm g},m_{\rm e})\,\tilde{\rho}_{{\rm g}\ m_{\rm g},{\rm e}\ m'_{\rm e}} - \Omega^*(m_{\rm g},m'_{\rm e})\,\tilde{\rho}_{{\rm e}\ m_{\rm e},{\rm g}\ m_{\rm g}}\right] - \Gamma\rho_{{\rm e}\ m_{\rm e},{\rm e}\ m'_{\rm e}}, \tag{7.520}$$

while the equations of motion for the *ground*-state populations and Zeeman coherences are

$$\begin{aligned}\partial_t \rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} = \frac{i}{2}\sum_{m_\mathrm{e}}\Big[\Omega(m'_\mathrm{g},m_\mathrm{e})\,\tilde\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{e}\ m_\mathrm{e}} - \Omega^*(m_\mathrm{g},m_\mathrm{e})\,\tilde\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{g}\ m'_\mathrm{g}}\Big] \\ + \Gamma\sum_q \rho_{\mathrm{e}\ (m_\mathrm{g}+q),\mathrm{e}\ (m'_\mathrm{g}+q)}\,\langle J_\mathrm{e}\ m_\mathrm{g}+q|J_\mathrm{g}\ m_\mathrm{g};1\ q\rangle\,\langle J_\mathrm{e}\ m'_\mathrm{g}+q|J_\mathrm{g}\ m'_\mathrm{g};1\ q\rangle.\end{aligned} \tag{7.521}$$

Similarly, the *optical* coherences (coherences between ground and excited states) evolve as

$$\begin{aligned}\partial_t\tilde\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{g}\ m_\mathrm{g}} &= -\frac{i}{2}\sum_{m'_\mathrm{g}}\Omega(m'_\mathrm{g},m_\mathrm{e})\,\rho_{\mathrm{g}\ m'_\mathrm{g},\mathrm{g}\ m_\mathrm{g}} + \frac{i}{2}\sum_{m'_\mathrm{e}}\Omega(m_\mathrm{g},m'_\mathrm{e})\,\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}} - \left(\frac{\Gamma}{2}-i\Delta\right)\tilde\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{g}\ m_\mathrm{g}} \\ \partial_t\tilde\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{e}\ m_\mathrm{e}} &= \frac{i}{2}\sum_{m'_\mathrm{g}}\Omega^*(m'_\mathrm{g},m_\mathrm{e})\,\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} - \frac{i}{2}\sum_{m'_\mathrm{e}}\Omega^*(m_\mathrm{g},m'_\mathrm{e})\,\rho_{\mathrm{e}\ m'_\mathrm{e},\mathrm{e}\ m_\mathrm{e}} - \left(\frac{\Gamma}{2}+i\Delta\right)\tilde\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{e}\ m_\mathrm{e}}.\end{aligned} \tag{7.522}$$

Now to make the adiabatic approximation, we set $\partial_t\tilde\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{g}\ m_\mathrm{g}} \approx \partial_t\tilde\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{e}\ m_\mathrm{e}} \approx 0$, so that we assume the coherences to be always in equilibrium with respect to the populations. This is justified, for example, when the detuning is much larger than the damping rate, where the large separation of time scales justifies ignoring the fast rotations caused by the terms proportional to Δ (recall this argument for the two-level atom in Section 5.8.3). It also may be that the optical coherences are damped quickly, as would be the case for strong collisional dephasing (Section 5.6.2). Then solving Eqs. (7.522) for the coherences in this approximation, we find that we can write the optical coherences in terms of populations and Zeeman coherences as

$$\begin{aligned}\tilde\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{g}\ m_\mathrm{g}} &= -\frac{i}{2(\Gamma/2-i\Delta)}\left[\sum_{m'_\mathrm{g}}\Omega(m'_\mathrm{g},m_\mathrm{e})\,\rho_{\mathrm{g}\ m'_\mathrm{g},\mathrm{g}\ m_\mathrm{g}} - \sum_{m'_\mathrm{e}}\Omega(m_\mathrm{g},m'_\mathrm{e})\,\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}}\right] \\ \tilde\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{e}\ m_\mathrm{e}} &= \frac{i}{2(\Gamma/2+i\Delta)}\left[\sum_{m'_\mathrm{g}}\Omega^*(m'_\mathrm{g},m_\mathrm{e})\,\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} - \sum_{m'_\mathrm{e}}\Omega^*(m_\mathrm{g},m'_\mathrm{e})\,\rho_{\mathrm{e}\ m'_\mathrm{e},\mathrm{e}\ m_\mathrm{e}}\right].\end{aligned} \tag{7.523}$$

Putting these expressions into Eqs. (7.520) and (7.521), we obtain the following ''rate equations'' for the populations and Zeeman coherences of a fine-structure transition coupled to a single field:

$$\begin{aligned}\partial_t\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}} &= \sum_{m_\mathrm{g}m'_\mathrm{g}}\left[\frac{\Omega^*(m_\mathrm{g},m'_\mathrm{e})\,\Omega(m'_\mathrm{g},m_\mathrm{e})}{4(\Gamma/2-i\Delta)} + \frac{\Omega(m_\mathrm{g},m_\mathrm{e})\,\Omega^*(m'_\mathrm{g},m'_\mathrm{e})}{4(\Gamma/2+i\Delta)}\right]\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} \\ &\quad - \sum_{m_\mathrm{g}m''_\mathrm{e}}\left[\frac{\Omega^*(m_\mathrm{g},m'_\mathrm{e})\,\Omega(m_\mathrm{g},m''_\mathrm{e})}{4(\Gamma/2-i\Delta)}\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m''_\mathrm{e}} + \frac{\Omega(m_\mathrm{g},m_\mathrm{e})\,\Omega^*(m_\mathrm{g},m''_\mathrm{e})}{4(\Gamma/2+i\Delta)}\rho_{\mathrm{e}\ m''_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}}\right] \\ &\quad - \Gamma\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}} \\ \partial_t\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} &= \sum_{m_\mathrm{e}m'_\mathrm{e}}\left[\frac{\Omega^*(m_\mathrm{g},m_\mathrm{e})\,\Omega(m'_\mathrm{g},m'_\mathrm{e})}{4(\Gamma/2-i\Delta)} + \frac{\Omega(m'_\mathrm{g},m'_\mathrm{e})\,\Omega^*(m_\mathrm{g},m_\mathrm{e})}{4(\Gamma/2+i\Delta)}\right]\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}} \\ &\quad - \sum_{m_\mathrm{e}m''_\mathrm{g}}\left[\frac{\Omega^*(m_\mathrm{g},m_\mathrm{e})\,\Omega(m''_\mathrm{g},m_\mathrm{e})}{4(\Gamma/2-i\Delta)}\rho_{\mathrm{g}\ m''_\mathrm{g},\mathrm{g}\ m'_\mathrm{g}} + \frac{\Omega(m'_\mathrm{g},m_\mathrm{e})\,\Omega^*(m''_\mathrm{g},m_\mathrm{e})}{4(\Gamma/2+i\Delta)}\rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m''_\mathrm{g}}\right] \\ &\quad + \Gamma\sum_q \rho_{\mathrm{e}\ (m_\mathrm{g}+q),\mathrm{e}\ (m'_\mathrm{g}+q)}\,\langle J_\mathrm{e}\ m_\mathrm{g}+q|J_\mathrm{g}\ m_\mathrm{g};1\ q\rangle\,\langle J_\mathrm{e}\ m'_\mathrm{g}+q|J_\mathrm{g}\ m'_\mathrm{g};1\ q\rangle.\end{aligned}$$

(fine-structure rate equations) (7.524)

These equations greatly reduce the complexity of the full master equation, since now the optical coherences are eliminated, and only the populations and Zeeman coherences must be tracked.

7.8.3.1 Single Field Polarization

At first glance, it may seem strange that the rate equations (7.524) should still contain the Zeeman coherences, since rate equations explicitly ignore coherences. In fact, for a *single* polarization of the field in the spherical basis (i.e., not more than one of $\Omega_{-1,0,1}$ is nonzero), the rate equations simplify. Noting that due to the Clebsch–Gordan coefficients involved, products such as $\Omega(m_\mathrm{g}, m_\mathrm{e})\,\Omega^*(m_\mathrm{g}, m'_\mathrm{e})$ are proportional to Kronecker symbols like $\delta_{m_\mathrm{e} m'_\mathrm{e}}$, since as in this example, a single polarization couples a ground state to at most one excited state. Then the equation of motion for the excited-state populations reduces to

$$\partial_t \rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m_\mathrm{e}} = -\sum_{m_\mathrm{g}} \frac{|\Omega(m_\mathrm{g}, m_\mathrm{e})|^2}{\Gamma(1 + 4\Delta^2/\Gamma^2)} (\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m_\mathrm{e}} - \rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m_\mathrm{g}}) - \Gamma \rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m_\mathrm{e}}.$$

(fine-structure rate equations, single polarization) (7.525)

We can perform a similar reduction for the ground-state populations, but to simplify the decay terms we will also assume that the excited-state Zeeman coherences vanish, $\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m'_\mathrm{e}} = 0$ $(m_\mathrm{e} \neq m'_\mathrm{e})$. The resulting equation is

$$\partial_t \rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m_\mathrm{g}} = \sum_{m_\mathrm{e}} \frac{|\Omega(m_\mathrm{g}, m_\mathrm{e})|^2}{\Gamma(1 + 4\Delta^2/\Gamma^2)} (\rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m_\mathrm{e}} - \rho_{\mathrm{g}\ m_\mathrm{g},\mathrm{g}\ m_\mathrm{g}}) + \Gamma \sum_{m_\mathrm{e}} \rho_{\mathrm{e}\ m_\mathrm{e},\mathrm{e}\ m_\mathrm{e}} \langle J_\mathrm{e}\ m_\mathrm{e} | J_\mathrm{g}\ m_\mathrm{g}; 1\ (m_\mathrm{e} - m_\mathrm{g}) \rangle^2.$$

(fine-structure rate equations, single polarization, no Zeeman coherence) (7.526)

Under these conditions, the rate equations for the *populations only* are closed, and thus we need not consider the Zeeman coherences. Note that from the original rate equations (7.524), we can see that the excited-state coherences, if initially zero, will remain so if the ground-state coherences are also zero, and vice versa. This is because the field transfers coherence between the excited and ground levels via Rabi flopping. Thus, for a single polarization, our neglecting Zeeman coherences is justified if they all start out as zero (including the ground-state coherences). This may be due to an unoriented atom, but note that since the Zeeman coherences decay with time but are not otherwise excited by the field (for a single polarization), this assumption is eventually justified anyway.

Thus we see the importance of the Zeeman coherences in the rate equations (7.524): they are necessary to represent an arbitrary orientation of the atom in the ground or excited state. When the atom is in an initial state with no Zeeman coherence, such as a single ground sublevel, light with an arbitrary polarization will in general put the atom in a coherent superposition of excited states, thus inducing Zeeman coherence that represents the field-induced orientation of the atom.

7.8.3.2 Multiple Fields

We already indicated above in Eq. (7.519) that adding a second field introduces an extra Hamiltonian interaction with explicit time dependence if the second field is of a different frequency from the main field. This is because the rotating-frame transformation can only eliminate explicit time dependence at a single frequency; a second frequency must be dealt with directly. This is difficult to handle analytically, as the nonlinear response of the atom will in general generate slowly varying dynamics as well as dynamics at the probe detuning Δ_p, and multiples thereof. We have already studied the interaction of atoms with bichromatic fields, for example, in the probe absorption by a driven two-level atom (Section 5.7.6, Problem 5.20), in the Autler-Townes doublet (Section 5.7.6.1 Problem 5.16), in stimulated Raman scattering (Section 6.1), and in coherent population trapping (Section 6.2). We have seen in these cases that the nonlinear mixing of the two fields can lead to strong, coherent effects such as level splittings and population transfer, and in general these effects can only be fully caputured in a full master-equation treatement. Under conditions where these effects are negligible or unimportant, however, the interaction of an atom with multiple fields can still be treated within a rate-equation formalism. The basic assumption required here is that all fields perturb the atom in the linear-response regime, and thus there are no cooperative effects induced by the multiple fields. This requires that the fields are weak (with either intensities well below the saturation intensities, or far detuned compared to the Rabi frequency from any transition), and that no multiphoton resonance (e.g., Raman resonance) occurs. Also, we should assume that any beat frequencies between the multiple fields are fast on time scales of interest (slow beats can be crudely modeled using a slowly varying Rabi frequency).

At this level of approximation, rate-equation terms involving Ω, for example as in Eqs. (7.524), are simply repeated for each field.

7.8.3.3 Hyperfine Structure

The hyperfine case is somewhat more complicated by the proliferation of states and the fact that not all sublevels are degenerate in view of the hyperfine shifts. From the master equation (7.512), the equation of motion for the excited-state populations and coherences is

$$\begin{aligned}\partial_t \tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F'_\mathrm{e}\, m'_\mathrm{e}} = &-\left(\Gamma + i\omega_{F_\mathrm{e}F'_\mathrm{e}}\right)\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F'_\mathrm{e}\, m'_\mathrm{e}} \\ &- \frac{i}{2}\sum_{F_\mathrm{g} m_\mathrm{g}}\left[\Omega(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F'_\mathrm{e}\, m'_\mathrm{e}} - \Omega^*(F_\mathrm{g}, m_\mathrm{g}; F'_\mathrm{e}, m'_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F_\mathrm{g}\, m_\mathrm{g}}\right],\end{aligned} \tag{7.527}$$

where the hyperfine splittings are given in terms of the hyperfine shifts (7.134) as

$$\omega_{F_\mathrm{e}F'_\mathrm{e}} := \frac{\Delta E_\mathrm{hfs}(J_\mathrm{e}, I, F_\mathrm{e}) - \Delta E_\mathrm{hfs}(J_\mathrm{e}, I, F'_\mathrm{e})}{\hbar}, \tag{7.528}$$

the hyperfine Rabi frequencies are defined by Eq. (7.422), and we used the form (7.453) for the decay operator to work out the decay term. Using essentially the same procedure leading to Eq. (7.453) we can also work out the remaining decay term $\Sigma_q \rho \Sigma_q^\dagger$ to find the equation of motion

$$\begin{aligned}\partial_t \tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F'_\mathrm{g}\, m'_\mathrm{g}} = &-i\omega_{F_\mathrm{g}F'_\mathrm{g}}\,\tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F'_\mathrm{g}\, m'_\mathrm{g}} + \sum_{F_\mathrm{e}F'_\mathrm{e}}\sum_{q=-1}^{1}\Gamma(F_\mathrm{g}, m_\mathrm{g}; F'_\mathrm{g}, m'_\mathrm{g}; F_\mathrm{e}; F'_\mathrm{e}; q)\,\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{g}+q, F'_\mathrm{e}\, m'_\mathrm{g}+q} \\ &+ \frac{i}{2}\sum_{F_\mathrm{e} m_\mathrm{e}}\left[\Omega(F'_\mathrm{g}, m'_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F_\mathrm{e}\, m_\mathrm{e}} - \Omega^*(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F'_\mathrm{g}\, m'_\mathrm{g}}\right],\end{aligned} \tag{7.529}$$

where we have defined

$$\begin{aligned}\Gamma(F_\mathrm{g}, m_\mathrm{g}; F'_\mathrm{g}, m'_\mathrm{g}; F_\mathrm{e}; F'_\mathrm{e}; q) := \Gamma(-1)^{F'_\mathrm{g}-F_\mathrm{g}}\sqrt{(2F'_\mathrm{g}+1)(2F_\mathrm{g}+1)}\,(2J_\mathrm{e}+1)\begin{Bmatrix} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F'_\mathrm{g} & F'_\mathrm{e} & I\end{Bmatrix}\begin{Bmatrix} J_\mathrm{e} & J_\mathrm{g} & 1 \\ F_\mathrm{g} & F_\mathrm{e} & I\end{Bmatrix} \\ \times \langle F'_\mathrm{e}\; m'_\mathrm{g} + q | F'_\mathrm{g}\; m'_\mathrm{g}; 1\; q\rangle\,\langle F_\mathrm{e}\; m_\mathrm{g} + q | F_\mathrm{g}\; m_\mathrm{g}; 1\; q\rangle\end{aligned} \tag{7.530}$$

as the return rate for the ground-state populations and coherences, and the ground-state hyperfine splittings are defined in the same way as for the excited states:

$$\omega_{F_\mathrm{g}F'_\mathrm{g}} := \frac{\Delta E_\mathrm{hfs}(J_\mathrm{g}, I, F_\mathrm{g}) - \Delta E_\mathrm{hfs}(J_\mathrm{g}, I, F'_\mathrm{g})}{\hbar}. \tag{7.531}$$

The equations of motion for the optical coherences are

$$\begin{aligned}\partial_t \tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F_\mathrm{g}\, m_\mathrm{g}} = &-\left(\frac{\Gamma}{2} - i\Delta(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\right)\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F_\mathrm{g}\, m_\mathrm{g}} \\ &- \frac{i}{2}\sum_{F'_\mathrm{g} m'_\mathrm{g}}\Omega(F'_\mathrm{g}, m'_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\,\tilde{\rho}_{F'_\mathrm{g}\, m'_\mathrm{g}, F_\mathrm{g}\, m_\mathrm{g}} + \frac{i}{2}\sum_{F'_\mathrm{e} m'_\mathrm{e}}\Omega(F_\mathrm{g}, m_\mathrm{g}; F'_\mathrm{e}, m'_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{e}\, m_\mathrm{e}, F'_\mathrm{e}\, m'_\mathrm{e}} \\ \partial_t \tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F_\mathrm{e}\, m_\mathrm{e}} = &-\left(\frac{\Gamma}{2} + i\Delta(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\right)\tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F_\mathrm{e}\, m_\mathrm{e}} \\ &+ \frac{i}{2}\sum_{F'_\mathrm{g} m'_\mathrm{g}}\Omega^*(F'_\mathrm{g}, m'_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{g}\, m_\mathrm{g}, F'_\mathrm{g}\, m'_\mathrm{g}} - \frac{i}{2}\sum_{F'_\mathrm{e} m'_\mathrm{e}}\Omega^*(F_\mathrm{g}, m_\mathrm{g}; F'_\mathrm{e}, m'_\mathrm{e})\,\tilde{\rho}_{F'_\mathrm{e}\, m'_\mathrm{e}, F_\mathrm{e}\, m_\mathrm{e}},\end{aligned} \tag{7.532}$$

where $\Delta(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e})$ is the detuning from the $|F_\mathrm{g}\rangle \longrightarrow |F_\mathrm{e}\rangle$ hyperfine transitions,

$$\begin{aligned}\Delta(F_\mathrm{g}, m_\mathrm{g}; F_\mathrm{e}, m_\mathrm{e}) &= \omega - \left(\omega_0 + \frac{\Delta E_\mathrm{hfs}(J_\mathrm{e}, I, F_\mathrm{e}) - \Delta E_\mathrm{hfs}(J_\mathrm{g}, I, F_\mathrm{g})}{\hbar}\right) \\ &= \Delta - \frac{\Delta E_\mathrm{hfs}(J_\mathrm{e}, I, F_\mathrm{e}) - \Delta E_\mathrm{hfs}(J_\mathrm{g}, I, F_\mathrm{g})}{\hbar},\end{aligned} \tag{7.533}$$

where again ω_0 is the transition frequency of the center of gravity of the hyperfine transition, and Δ is the laser detuning with respect to ω_0. In the adiabatic approximation, Eqs. (7.532) become

$$\begin{aligned}
\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F_\mathrm{g}\,m_\mathrm{g}} &= -\frac{i}{2[\Gamma/2 - i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\Bigg[\sum_{F'_\mathrm{g}m'_\mathrm{g}} \Omega(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\tilde{\rho}_{F'_\mathrm{g}\,m'_\mathrm{g},F_\mathrm{g}\,m_\mathrm{g}} \\
&\qquad\qquad - \sum_{F'_\mathrm{e}m'_\mathrm{e}} \Omega(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F'_\mathrm{e}\,m'_\mathrm{e}}\Bigg] \\
\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F_\mathrm{e}\,m_\mathrm{e}} &= \frac{i}{2[\Gamma/2 + i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\Bigg[\sum_{F'_\mathrm{g}m'_\mathrm{g}} \Omega^*(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F'_\mathrm{g}\,m'_\mathrm{g}} \\
&\qquad\qquad - \sum_{F'_\mathrm{e}m'_\mathrm{e}} \Omega^*(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})\,\tilde{\rho}_{F'_\mathrm{e}\,m'_\mathrm{e},F_\mathrm{e}\,m_\mathrm{e}}\Bigg],
\end{aligned} \tag{7.534}$$

and putting these into the equations of motion for the populations and hyperfine coherences, we obtain the rather complicated rate equations

$$\begin{aligned}
\partial_t\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F'_\mathrm{e}\,m'_\mathrm{e}} &= \sum_{F_\mathrm{g}m_\mathrm{g}F'_\mathrm{g}m'_\mathrm{g}}\Bigg[\frac{\Omega^*(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})\,\Omega(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})}{4[\Gamma/2 - i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]} \\
&\qquad\qquad + \frac{\Omega(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\Omega^*(F'_\mathrm{g},m'_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})}{4[\Gamma/2 + i\Delta(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})]}\Bigg]\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F'_\mathrm{g}\,m'_\mathrm{g}} \\
&\quad - \sum_{F_\mathrm{g}m_\mathrm{g}F''_\mathrm{e}m''_\mathrm{e}}\Bigg[\frac{\Omega^*(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})\,\Omega(F_\mathrm{g},m_\mathrm{g};F''_\mathrm{e},m''_\mathrm{e})}{4[\Gamma/2 - i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F''_\mathrm{e}\,m''_\mathrm{e}} \\
&\qquad\qquad + \frac{\Omega(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\Omega^*(F_\mathrm{g},m_\mathrm{g};F''_\mathrm{e},m''_\mathrm{e})}{4[\Gamma/2 + i\Delta(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})]}\tilde{\rho}_{F''_\mathrm{e}\,m''_\mathrm{e},F'_\mathrm{e}\,m'_\mathrm{e}}\Bigg] \\
&\quad - \left(\Gamma + i\omega_{F_\mathrm{e}F'_\mathrm{e}}\right)\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F'_\mathrm{e}\,m'_\mathrm{e}} \\
\partial_t\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F'_\mathrm{g}\,m'_\mathrm{g}} &= \sum_{F_\mathrm{e}m_\mathrm{e}F'_\mathrm{e}m'_\mathrm{e}}\Bigg[\frac{\Omega^*(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\Omega(F'_\mathrm{g},m'_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})}{4[\Gamma/2 - i\Delta(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]} \\
&\qquad\qquad + \frac{\Omega(F'_\mathrm{g},m'_\mathrm{g};F'_\mathrm{e},m'_\mathrm{e})\,\Omega^*(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})}{4[\Gamma/2 + i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\Bigg]\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{e},F'_\mathrm{e}\,m'_\mathrm{e}} \\
&\quad - \sum_{F_\mathrm{e}m_\mathrm{e}F''_\mathrm{g}m''_\mathrm{g}}\Bigg[\frac{\Omega^*(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\Omega(F''_\mathrm{g},m''_\mathrm{g};F_\mathrm{e},m_\mathrm{e})}{4[\Gamma/2 - i\Delta(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\tilde{\rho}_{F''_\mathrm{g}\,m''_\mathrm{g},F'_\mathrm{g}\,m'_\mathrm{g}} \\
&\qquad\qquad + \frac{\Omega(F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e},m_\mathrm{e})\,\Omega^*(F''_\mathrm{g},m''_\mathrm{g};F_\mathrm{e},m_\mathrm{e})}{4[\Gamma/2 + i\Delta(F_\mathrm{g},m_\mathrm{g};F_\mathrm{e},m_\mathrm{e})]}\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F''_\mathrm{g}\,m''_\mathrm{g}}\Bigg] \\
&\quad + \sum_{F_\mathrm{e}F'_\mathrm{e}q}\Gamma(F_\mathrm{g},m_\mathrm{g};F'_\mathrm{g},m'_\mathrm{g};F_\mathrm{e};F'_\mathrm{e};q)\,\tilde{\rho}_{F_\mathrm{e}\,m_\mathrm{g}+q,F'_\mathrm{e}\,m'_\mathrm{g}+q} - i\omega_{F_\mathrm{g}F'_\mathrm{g}}\,\tilde{\rho}_{F_\mathrm{g}\,m_\mathrm{g},F'_\mathrm{g}\,m'_\mathrm{g}}.
\end{aligned}$$

(hyperfine-structure rate equations) (7.535)

Note that we still have coherences between hyperfine levels (say, between levels $F_\mathrm{g} \neq F'_\mathrm{g}$ and $F_\mathrm{e} \neq F'_\mathrm{e}$) that rotate at the hyperfine splittings. If the hyperfine structure is well-resolved, so that the hyperfine splittings are much larger than Γ, then it may be that we are similarly uninterested in these fast oscillations, and we can adiabatically eliminate these hyperfine coherences as well. We can obtain the adiabatic relations for

$F'_{\rm e} \neq F_{\rm e}$ by setting $\partial_t \tilde\rho_{F_{\rm e}\,m_{\rm e},F'_{\rm e}\,m'_{\rm e}} \approx 0$, with the result

$$\begin{aligned}\tilde\rho_{F_{\rm e}\,m_{\rm e},F'_{\rm e}\,m'_{\rm e}} = \sum_{F_{\rm g}m_{\rm g}F'_{\rm g}m'_{\rm g}} \Bigg[&\frac{\Omega^*(F_{\rm g},m_{\rm g};F'_{\rm e},m'_{\rm e})\,\Omega(F'_{\rm g},m'_{\rm g};F_{\rm e},m_{\rm e})}{4(\Gamma+i\omega_{F_{\rm e}F'_{\rm e}})[\Gamma/2-i\Delta(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})]} \\ &+ \frac{\Omega(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})\,\Omega^*(F'_{\rm g},m'_{\rm g};F'_{\rm e},m'_{\rm e})}{4(\Gamma+i\omega_{F_{\rm e}F'_{\rm e}})[\Gamma/2+i\Delta(F_{\rm g},m_{\rm g};F'_{\rm e},m'_{\rm e})]}\Bigg]\tilde\rho_{F_{\rm g}\,m_{\rm g},F'_{\rm g}\,m'_{\rm g}} \\ - \sum_{F_{\rm g}m_{\rm g}F''_{\rm e}m''_{\rm e}} \Bigg[&\frac{\Omega^*(F_{\rm g},m_{\rm g};F'_{\rm e},m'_{\rm e})\,\Omega(F_{\rm g},m_{\rm g};F''_{\rm e},m''_{\rm e})}{4(\Gamma+i\omega_{F_{\rm e}F'_{\rm e}})[\Gamma/2-i\Delta(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})]}\tilde\rho_{F_{\rm e}\,m_{\rm e},F''_{\rm e}\,m''_{\rm e}} \\ &+ \frac{\Omega(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})\,\Omega^*(F_{\rm g},m_{\rm g};F''_{\rm e},m''_{\rm e})}{4(\Gamma+i\omega_{F_{\rm e}F'_{\rm e}})[\Gamma/2+i\Delta(F_{\rm g},m_{\rm g};F'_{\rm e},m'_{\rm e})]}\tilde\rho_{F''_{\rm e}\,m''_{\rm e},F'_{\rm e}\,m'_{\rm e}}\Bigg],\end{aligned} \tag{7.536}$$

and the relations for $F'_{\rm g} \neq F_{\rm g}$ follow by setting $\partial_t \tilde\rho_{F_{\rm g}\,m_{\rm g},F'_{\rm g}\,m'_{\rm g}} \approx 0$, with the result

$$\begin{aligned}\tilde\rho_{F_{\rm g}\,m_{\rm g},F'_{\rm g}\,m'_{\rm g}} = \sum_{F_{\rm e}m_{\rm e}F'_{\rm e}m'_{\rm e}} \Bigg[&\frac{\Omega^*(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})\,\Omega(F'_{\rm g},m'_{\rm g};F'_{\rm e},m'_{\rm e})}{4i\omega_{F_{\rm g}F'_{\rm g}}[\Gamma/2-i\Delta(F'_{\rm g},m'_{\rm g};F_{\rm e},m_{\rm e})]} \\ &+ \frac{\Omega(F'_{\rm g},m'_{\rm g};F'_{\rm e},m'_{\rm e})\,\Omega^*(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})}{4i\omega_{F_{\rm g}F'_{\rm g}}[\Gamma/2+i\Delta(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})]}\Bigg]\tilde\rho_{F_{\rm e}\,m_{\rm e},F'_{\rm e}\,m'_{\rm e}} \\ - \sum_{F_{\rm e}m_{\rm e}F''_{\rm g}m''_{\rm g}} \Bigg[&\frac{\Omega^*(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})\,\Omega(F''_{\rm g},m''_{\rm g};F_{\rm e},m_{\rm e})}{4i\omega_{F_{\rm g}F'_{\rm g}}[\Gamma/2-i\Delta(F'_{\rm g},m'_{\rm g};F_{\rm e},m_{\rm e})]}\tilde\rho_{{\rm g}\,m''_{\rm g},{\rm g}\,m'_{\rm g}} \\ &+ \frac{\Omega(F'_{\rm g},m'_{\rm g};F_{\rm e},m_{\rm e})\,\Omega^*(F''_{\rm g},m''_{\rm g};F_{\rm e},m_{\rm e})}{4i\omega_{F_{\rm g}F'_{\rm g}}[\Gamma/2+i\Delta(F_{\rm g},m_{\rm g};F_{\rm e},m_{\rm e})]}\tilde\rho_{F_{\rm g}\,m_{\rm g},F''_{\rm g}\,m''_{\rm g}}\Bigg] \\ + \frac{1}{i\omega_{F_{\rm g}F'_{\rm g}}} \sum_{F_{\rm e}F'_{\rm e}q} &\Gamma(F_{\rm g},m_{\rm g};F'_{\rm g},m'_{\rm g};F_{\rm e};F'_{\rm e};q)\,\tilde\rho_{F_{\rm e}\,m_{\rm g}+q,F'_{\rm e}\,m'_{\rm g}+q}.\end{aligned} \tag{7.537}$$

Putting these two relations into the rate equations (7.535) leads to a yet more complicated set of rate equations, but that only retain coherences between degenerate levels (which properly accounts for the orientation of the atom in the various levels). Note that the terms generated by this adiabatic elimination are *quartic* in the field Ω, whereas there are also *quadratic* terms in the field. If we work in the weak-field approximation, where all Rabi frequencies are small compared to the hyperfine splittings, Γ, or the detunings from the hyperfine resonances (or even better, small compared to all of these), then we can ignore these higher-order terms. This corresponds to taking

$$\begin{aligned}\tilde\rho_{F_{\rm e}\,m_{\rm e},F'_{\rm e}\,m'_{\rm e}} &\approx 0 \\ \tilde\rho_{F_{\rm g}\,m_{\rm g},F'_{\rm g}\,m'_{\rm g}} &\approx \frac{1}{i\omega_{F_{\rm g}F'_{\rm g}}} \sum_{F_{\rm e}q} \Gamma(F_{\rm g},m_{\rm g};F'_{\rm g},m'_{\rm g};F_{\rm e};F_{\rm e};q)\,\tilde\rho_{F_{\rm e}\,m_{\rm g}+q,F_{\rm e}\,m'_{\rm g}+q},\end{aligned} \tag{7.538}$$

again for $F_g \neq F_g'$ and $F_e \neq F_e'$. Putting these into the rate equations (7.535), we obtain the low-intensity rate equations

$$\begin{aligned}
\partial_t \tilde{\rho}_{F_e\, m_e, F_e\, m_e'} = & \sum_{F_g m_g m_g'} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e')\, \Omega(F_g, m_g'; F_e, m_e)}{4[\Gamma/2 - i\Delta(F_g, m_g; F_e, m_e)]} + \frac{\Omega(F_g, m_g; F_e, m_e)\, \Omega^*(F_g, m_g'; F_e, m_e')}{4[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e')]} \right] \\
& \qquad \times \tilde{\rho}_{F_g\, m_g, F_g\, m_g'} \\
& + \sum_{\substack{F_g m_g F_g' m_g' F_e q \\ (F_g \neq F_g')}} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e')\, \Omega(F_g', m_g'; F_e, m_e)}{4[\Gamma/2 - i\Delta(F_g, m_g; F_e, m_e)]} + \frac{\Omega(F_g, m_g; F_e, m_e)\, \Omega^*(F_g', m_g'; F_e, m_e')}{4[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e')]} \right] \\
& \qquad \times \frac{1}{i\omega_{F_g F_g'}} \Gamma(F_g, m_g; F_g', m_g'; F_e; F_e; q)\, \tilde{\rho}_{F_e\, m_g + q, F_e\, m_g' + q} \\
& - \sum_{F_g m_g m_e''} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e')\, \Omega(F_g, m_g; F_e, m_e'')}{4[\Gamma/2 - i\Delta(F_g, m_g; F_e, m_e)]}\, \tilde{\rho}_{F_e\, m_e, F_e\, m_e''} \right. \\
& \qquad \left. + \frac{\Omega(F_g, m_g; F_e, m_e)\, \Omega^*(F_g, m_g; F_e, m_e'')}{4[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e')]}\, \tilde{\rho}_{F_e\, m_e'', F_e\, m_e'} \right] \\
& - \Gamma \tilde{\rho}_{F_e\, m_e, F_e\, m_e'}
\end{aligned}$$

(hyperfine-structure rate equations, small intensity) (7.539)

for the excited states and

$$\begin{aligned}
\partial_t \tilde{\rho}_{F_g\, m_g, F_g\, m_g'} = & \sum_{F_e m_e m_e'} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e)\, \Omega(F_g, m_g'; F_e, m_e')}{4[\Gamma/2 - i\Delta(F_g, m_g'; F_e, m_e)]} \right. \\
& \qquad \left. + \frac{\Omega(F_g, m_g'; F_e, m_e')\, \Omega^*(F_g, m_g; F_e, m_e)}{4[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e)]} \right] \tilde{\rho}_{F_e\, m_e, F_e\, m_e'} \\
& - \sum_{F_e m_e m_g''} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e)\, \Omega(F_g, m_g''; F_e, m_e)}{4[\Gamma/2 - i\Delta(F_g', m_g'; F_e, m_e)]} \tilde{\rho}_{F_g\, m_g'', F_g\, m_g'} \right. \\
& \qquad \left. + \frac{\Omega(F_g, m_g'; F_e, m_e)\, \Omega^*(F_g, m_g''; F_e, m_e)}{4[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e)]} \tilde{\rho}_{F_g\, m_g, F_g\, m_g''} \right] \\
& - \sum_{\substack{F_e m_e F_g'' m_g'' F_e' q \\ (F_g'' \neq F_g)}} \left[\frac{\Omega^*(F_g, m_g; F_e, m_e)\, \Omega(F_g'', m_g''; F_e, m_e)}{4i\omega_{F_g'' F_g}[\Gamma/2 - i\Delta(F_g, m_g'; F_e, m_e)]} \right. \\
& \qquad \times \Gamma(F_g'', m_g''; F_g, m_g'; F_e'; F_e'; q)\, \tilde{\rho}_{F_e'\, m_g'' + q, F_e'\, m_g' + q} \\
& \qquad + \frac{\Omega(F_g, m_g'; F_e, m_e)\, \Omega^*(F_g'', m_g''; F_e, m_e)}{4i\omega_{F_g F_g''}[\Gamma/2 + i\Delta(F_g, m_g; F_e, m_e)]} \\
& \qquad \left. \times \Gamma(F_g, m_g; F_g'', m_g''; F_e'; F_e'; q)\, \tilde{\rho}_{F_e'\, m_g + q, F_e'\, m_g'' + q} \right] \\
& + \sum_{F_e q} \Gamma(F_g, m_g; F_g, m_g'; F_e; F_e; q)\, \tilde{\rho}_{F_e\, m_g + q, F_e\, m_g' + q}.
\end{aligned}$$

(hyperfine-structure rate equations, small intensity) (7.540)

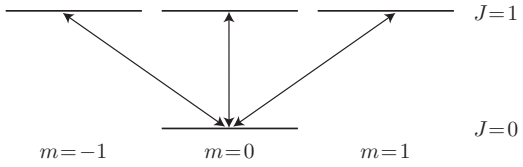
for the ground states.

7.9 Whither has Wandered the Two-Level Atom?

As we have seen, the real situation with atomic angular-momentum structure is considerably more complicated than the idealized model of a two-level atom. So to what extent is the two-level atom a useful model? Actually, there are some important situations under which atoms with angular-momentum degeneracy behave as two-level atoms. One obvious candidate is the a transition of the form $J = 0 \longrightarrow J' = 0$

(equivalently, $F = 0 \longrightarrow F' = 0$, though for simplicity here we will only refer to fine-structure quantum numbers), where each level has only one sublevel. Unfortunately, we have already seen according to the dipole selection rules that this transition is forbidden.

The most direct realization of the two-level atom comes in the form of a $J = 0 \longrightarrow J' = 1$ transition, where depending on the polarization, the ground state can be coupled to one of three excited states.



The important thing to realize is that, given an *arbitrary* polarization of the field, the ground state is coupled to *some* linear combination of the three excited states, while two other linear combinations are not coupled. The transition matrix elements from the Wigner–Eckart theorem (7.237) are

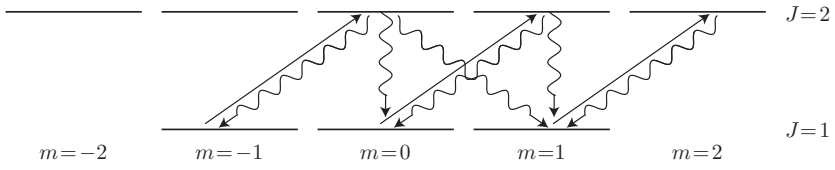
$$\langle J = 0,\ m_J = 0|d_q|J' = 1,\ m'_J\rangle = \langle J = 0\|\mathbf{d}\|J' = 1\rangle\, \langle 0\ 0|1\ m'_J; 1\ q\rangle = \langle J = 0\|\mathbf{d}\|J' = 1\rangle\, \frac{(-1)^{1+q}}{\sqrt{3}}\, \delta_{m'_J,-q} \tag{7.541}$$

Thus, we see that the amplitude of the matrix element for every polarization is identical, and in fact equal to the effective matrix element (7.492) for linear polarization that we wrote down before, up to a minus sign. The point is that an any polarization couples to the atom with the same strength, and so except for the atomic orientation in the excited state (corresponding to the orientation of the induced dipole moment) there is no dependence of any of the physics on the polarization. Further, since only one excited state is coupled, we may regard this system as a two-level atom. Any decay will be back to the ground state, corresponding to dipole radiation with the same orientation as the inducing field.

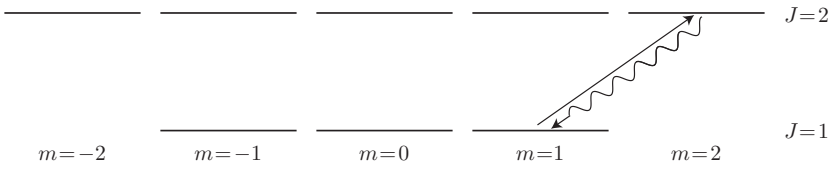
Of course, this argument breaks down if there are *two* fields present with different orientation, such as a second laser with another polarization, or a static magnetic or electric field. Then one field induces an orientation that modifies the interaction with the other field.

7.9.1 Optical Pumping to Stretched States

Another important situation comes in the form of a $J \longrightarrow J' = J + 1$ transition pumped by circularly polarized light. As a concrete example, we can consider a $J = 1 \longrightarrow J' = 2$ transition coupled by σ^+=polarized light. Recall that this light drives sublevel transitions of the form $m_J \longrightarrow m'_J = m_J + 1$. However, spontaneous decay occurs from any possible $m'_J \longrightarrow m_J = m'_J \pm 1, 0$.



The $J = 1, m_J = 1 \longrightarrow J' = 2, m'_J = 2$ transition is thus closed in this scheme. Atoms starting in any other state will eventually become pumped into this cycling transition (on the ‘‘stretched states’’), and thus, at least in steady state, we effectively have a two-level atom.



The dipole matrix element for this transition is given by the Wigner–Eckart theorem (7.237) as

$$\begin{aligned}\langle J,\, m_J = J|d_{-1}|J' = J+1,\, m'_J = J'\rangle &= \langle J\|\mathbf{d}\|J+1\rangle\,\langle J\ J|J+1\ J+1;1\ -1\rangle \\ &= \langle J\|\mathbf{d}\|J+1\rangle\sqrt{\frac{2J+1}{2(J+1)+1}} \\ &= \langle J\|\mathbf{d}\|J'\rangle\sqrt{\frac{2J+1}{2J'+1}}.\end{aligned} \tag{7.542}$$

The matrix element is thus the reduced matrix element for the hyperfine transition multiplied by a degeneracy ratio for the transition. Notice that this is precisely the same degeneracy ratio that appears in the decay-rate formula (7.305). That is, if we define the effective dipole moment

$$d_{\text{stretch}} := \langle J\|\mathbf{d}\|J'\rangle\sqrt{\frac{2J+1}{2J'+1}}, \tag{7.543}$$
(effective stretched-transition dipole)

for the stretched-state transition, then we obtain the two-level-atom formula (7.299) when we write the decay rate in terms of this transition:

$$\Gamma_{J_g J_e} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}|d_{\text{stretch}}|^2. \tag{7.544}$$

Physically, this is because the stretched excited state has only one decay path, which decays at the full rate $\Gamma_{J_g J_e}$. Thus, no summation—as is implied in the reduced matrix element—is necessary to compute the full decay rate $\Gamma_{J_g J_e}$.

The same thing happens in a closed *hyperfine* transition between stretched states, in the case where the excited state also has only one decay path. This happens again for the fine structure transition $J \longrightarrow J' = J+1$, in particular for the σ^+ hyperfine transition $F = J+I, m_F = F \longrightarrow F' = J'+I = F+1, m'_F = F'$. In the hyperfine-structure diagram for ^{133}Cs on p. 321, we are referring to the $F = 4 \longrightarrow F' = 5$ (laser-cooling) transition. The transition here is closed because the F' excited level can only decay to a ground level with $F = F' \pm 1, 0$, and thus only has one decay option. As in the fine-structure case, the stretched excited state only decays to the stretched ground state. The dipole moment for this hyperfine transition is

$$\begin{aligned}\langle F,\, m_F = F|d_{-1}|F' = F+1,\, m'_F = F'\rangle &= \langle F\|\mathbf{d}\|F+1\rangle\,\langle F\ F|F+1\ F+1;1\ -1\rangle \\ &= \langle F\|\mathbf{d}\|F'\rangle\sqrt{\frac{2F+1}{2F'+1}} \\ &= \langle J\|\mathbf{d}\|J'\rangle\,(-1)^{F'+J+1+I} \\ &\qquad \times\sqrt{(2F'+1)(2J+1)}\begin{Bmatrix} J & J' & 1 \\ F' & F & I\end{Bmatrix}\sqrt{\frac{2F+1}{2F'+1}} \\ &= \langle J\|\mathbf{d}\|J'\rangle\,(-1)^{F'-J+1-I}\sqrt{\frac{2J+1}{2J'+1}} \\ &= \langle J\|\mathbf{d}\|J'\rangle\sqrt{\frac{2J+1}{2J'+1}}\end{aligned} \tag{7.545}$$

after using the hyperfine Wigner–Eckart theorem (7.281), the decomposition rule (7.282), the 6-j symbol

$$\begin{Bmatrix} J & J' & 1 \\ F' & F & I\end{Bmatrix} = \begin{Bmatrix} J & J+1 & 1 \\ I+J+1 & I+J & I\end{Bmatrix} = \frac{(-1)^{-2(I+J)}}{\sqrt{(2J'+1)(2F+1)}}, \tag{7.546}$$

and $F' - J + 1 - I = F - J - I + 2 = 2$. Thus, exactly the same effective dipole moment applies to the *hyperfine* stretched-state transition as to the similar fine-structure transition.

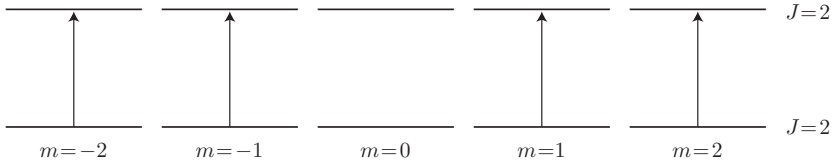
Typically, the effective dipole moment defined here is larger than that for large detunings, as in Eq. (7.492), because the optical pumping of the atom towards the stretched state produces an atomic

orientation that is aligned with the field. For $J = 0$, of course, the stretched-state transition has an effective squared dipole moment of $|\langle J\|\mathbf{d}\|J'\rangle|^2/3$, while in the limit as $J \longrightarrow \infty$, the effective squared dipole approaches $|\langle J\|\mathbf{d}\|J'\rangle|^2$, corresponding to perfect alignment with the field (so that the full weight of the atomic dipole is represented by only one spherical component).

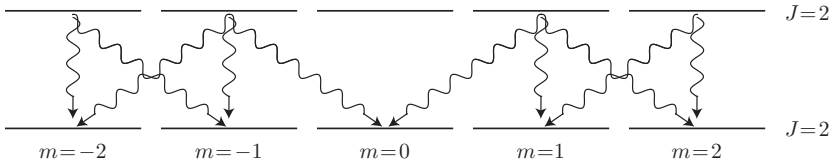
While the effect here is restricted to transitions of certain forms, obviously *some* alignment occurs when an atom is pumped by a circularly polarized field, even for transitions of arbitrary form. However, for $J \longrightarrow J' = J$ transition interacting with circularly polarized light, the atoms still become completely aligned in a stretched state, which is a dark state, in a way that is essentially equivalent to the mechanism described below.

7.9.2 Optical Pumping with Linearly Polarized Light

When pumping atoms with *linearly* polarized light, alignment phenomena occur that are similar to the circular-polarization case. Consider a $J \longrightarrow J' = J$ transition pumped by linearly polarized light, where J is some integer. For concreteness, we can consider a $J = 2 \longrightarrow J' = 2$ transition.



Because the $m_J = 0 \longrightarrow m'_J = 0$ transition is forbidden, the $m_J = 0$ ground sublevel is not excited, but all the other sublevels are. However, other states can decay into the $m_J = 0$ sublevel by $\sigma^{\pm}$ transitions.



Since the $m_J = 0$ sublevel has no excitation path, but there are excitation/decay routes from any other state into $m_J = 0$, in steady state the atom will end up entirely in $m_J = 0$. Again, this state is not coupled to an excited level by the linearly polarized light, so the atoms are in a dark state, no longer interacting with the light. Of course, particularly in the hyperfine case, there are other levels around for which transitions are not dipole-forbidden, so there will be leakage to some extent; furthermore there may be multiple ground hyperfine levels, in which case repumping from the other ground level(s) is necessary to pump the atoms into the dark state.

In a general $J \longrightarrow J'$ fine-structure transition, some degree of alignment towards $m_J = 0$ tends to occur. The exception is the case of $J \longrightarrow J' = J - 1$, where the $m_J = \pm J$ stretched ground states are dark, and thus the atoms tend to accumulate in those states. However, the less trivial cases are $J \longrightarrow J' = J + 1$ for arbitrary J or $J \longrightarrow J' = J$ for half-integer J. In steady state, the transition takes on a well-defined alignment, and behaves as a two level atom, so long as the transition is taken to have an appropriate (geometry-dependent) dipole moment. To find the explicit steady-state solutions in this case,[44] we can start with the rate equations (7.525) and (7.526). In deriving these equations, we assumed a single polarization and that the coherences were in quasi-steady state; thus, for computing steady states of the full master equation (7.507) or (7.510), they will produce *exact* results. Starting with the ground-state rate equation (7.525), the steady-state condition $\partial_t P_{\mathrm{e},m} = 0$ for the population $P_{\mathrm{e},m} := \rho_{\mathrm{e}\,m,\mathrm{e}\,m} = 0$ in state $|J_{\mathrm{e}}\,m\rangle$ gives

$$\frac{|\Omega(m,m)|^2}{\Gamma(1+4\Delta^2/\Gamma^2)}(P_{\mathrm{e},m} - P_{\mathrm{g},m}) = -\Gamma P_{\mathrm{e},m}, \tag{7.547}$$

[44] These solutions were derived with the resolvent method by Bo Gao, "Effects of Zeeman degeneracy on the steady-state properties of an atom interacting with a near-resonant laser field: Analytic results," *Physical Review A* **48**, 2443 (1993) (doi: 10.1103/PhysRevA.48.2443). For further results, see also Bo Gao, "Effects of Zeeman degeneracy on the steady-state properties of an atom interacting with a near-resonant laser field: Probe spectra," *Physical Review A* **49**, 3391 (1994) (doi: 10.1103/PhysRevA.49.3391); Bo Gao, "Effects of Zeeman degeneracy on the steady-state properties of an atom interacting with a near-resonant laser field: Resonance fluorescence," *Physical Review A* **50**, 4139 (1994). (doi: 10.1103/PhysRevA.50.4139).

while the same condition $\partial_t P_{\mathrm{g},m} = 0$ for the population $P_{\mathrm{g},m} := \rho_{\mathrm{g}\,m,\mathrm{g}\,m} = 0$ in the ground state $|J_\mathrm{g}\ m\rangle$ gives

$$\frac{|\Omega(m,m)|^2}{\Gamma(1+4\Delta^2/\Gamma^2)}(P_{\mathrm{e},m} - P_{\mathrm{g},m}) = -\Gamma \sum_{m'} P_{\mathrm{e},m'} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m; 1\ (m'-m)\rangle^2. \tag{7.548}$$

Combining these two relations to eliminate the Rabi frequency, we find a closed equation for the excited-state populations:

$$P_{\mathrm{e},m} = \sum_{m'} P_{\mathrm{e},m'}\ \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m; 1\ (m'-m)\rangle^2. \tag{7.549}$$

This is somewhat surprising: for a linearly polarized pump, the relative populations of the excited states (i.e., the orientation of the atomic excitation) is *completely independent* of the driving intensity. If we define the tridiagonal matrix

$$A_{mm'} := \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m; 1\ (m'-m)\rangle^2, \tag{7.550}$$

then Eq. (7.549) amounts to the homogeneous, tridiagonal linear system

$$(A_{mm'} - \delta_{mm'})\, x_{m'} = 0, \tag{7.551}$$

which we must now solve for the populations x_m. Note that the matrix $A_{mm'}$ has the extra constraints $A_{mm'} = A_{-m,-m'}$ from Eq. (7.67) and $\sum_{m'} A_{mm'} = 1$ from the orthogonality relation (7.46), and by the symmetry of the problem, the populations satisfy $x_m = x_{-m}$. The solution for $m > 0$ is given by the recursion formula (Problem 7.3)

$$x_{m+1} = \frac{A_{m+1,m}}{A_{m,m+1}} x_m, \tag{7.552}$$

which we may explicitly iterate to find

$$x_m = \left(\frac{\displaystyle\prod_{m'=0}^{m-1} A_{m'+1,m'}}{\displaystyle\prod_{m'=0}^{m-1} A_{m',m'+1}} \right) x_0 = \left(\frac{\displaystyle\prod_{m'=0}^{m-1} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m'+1; 1\ (-1)\rangle^2}{\displaystyle\prod_{m'=0}^{m-1} \langle J_\mathrm{e}\ m'+1|J_\mathrm{g}\ m'; 1\ 1\rangle^2} \right) x_0 \tag{7.553}$$

for $m > 0$. Since we will explicitly normalize these populations anyway, we can take a convenient normalization by writing

$$\left(\prod_{m'=0}^{J_\mathrm{g}-1} \langle J_\mathrm{e}\ m'+1|J_\mathrm{g}\ m'; 1\ 1\rangle^2 \right) \frac{x_m}{x_0} = \left(\prod_{m'=0}^{m-1} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m'+1; 1\ (-1)\rangle^2 \right) \left(\prod_{m'=m}^{J_\mathrm{g}-1} \langle J_\mathrm{e}\ m'+1|J_\mathrm{g}\ m'; 1\ 1\rangle^2 \right), \tag{7.554}$$

then using Eq. (7.67) to reverse the m_J quantum numbers in the first factor while letting $m' \longrightarrow m'-1$ in the second,

$$\left(\prod_{m'=0}^{J_\mathrm{g}-1} \langle J_\mathrm{e}\ m'+1|J_\mathrm{g}\ m'; 1\ 1\rangle^2 \right) \frac{x_m}{x_0} = \left(\prod_{m'=0}^{m-1} \langle J_\mathrm{e}\ -m'|J_\mathrm{g}\ (-m'-1); 1\ 1\rangle^2 \right) \left(\prod_{m'=m+1}^{J_\mathrm{g}} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m'-1; 1\ 1\rangle^2 \right), \tag{7.555}$$

then letting $m' \longrightarrow -m'$ in the first factor,

$$\left(\prod_{m'=0}^{J_\mathrm{g}-1} \langle J_\mathrm{e}\ m'+1|J_\mathrm{g}\ m'; 1\ 1\rangle^2 \right) \frac{x_m}{x_0} = \left(\prod_{m'=-m+1}^{0} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m'-1; 1\ 1\rangle^2 \right) \left(\prod_{m'=m+1}^{J_\mathrm{g}} \langle J_\mathrm{e}\ m'|J_\mathrm{g}\ m'-1; 1\ 1\rangle^2 \right), \tag{7.556}$$

and finally multiplying through by the same factor on the left,

$$\left(\prod_{m'=0}^{J_g-1}\langle J_e\ m'+1|J_g\ m';1\ 1\rangle^2\right)^2\frac{x_m}{x_0}=\left(\prod_{m'=-m+1}^{J_g}\langle J_e\ m'|J_g\ m'-1;1\ 1\rangle^2\right)\left(\prod_{m'=m+1}^{J_g}\langle J_e\ m'1|J_g\ m'-1;1\ 1\rangle^2\right). \tag{7.557}$$

This expression is explicitly invariant under $m\longrightarrow -m$, and so we can write out these weights explicitly normalized by defining a new symbol for the left-hand side of the above equation and then normalizing it:

$$\chi_{e,m}:=\left(\prod_{m'=-m+1}^{J_g}\langle J_e\ m'|J_g\ m'-1;1\ 1\rangle^2\right)\left(\prod_{m'=m+1}^{J_g}\langle J_e\ m'|J_g\ m'-1;1\ 1\rangle^2\right)$$

$$w_{e,m}:=\frac{\chi_{e,m}}{\displaystyle\sum_{m'=-J_g}^{J_g}\chi_{e,m'}}. \tag{7.558}$$

(relative excited-state weighting factors)

Then the relative excited-state populations are given by the normalized weights $w_{e,m}$,

$$P_{e,m}=w_{e,m}P_e, \tag{7.559}$$

(excited-state populations)

where the total excited-state population is

$$P_e:=\sum_{m'=-J_g}^{J_g}P_{e,m'}. \tag{7.560}$$

To find the total excited-state population, we can write Eq. (7.547) in the form

$$\left(\frac{|\Omega(m,m)|^2}{\Gamma(1+4\Delta^2/\Gamma^2)}+\Gamma\right)P_{e,m}=\frac{|\Omega(m,m)|^2}{\Gamma(1+4\Delta^2/\Gamma^2)}P_{g,m}, \tag{7.561}$$

or

$$\left(1+\frac{\Gamma^2+4\Delta^2}{|\Omega(m,m)|^2}\right)w_{e,m}P_e=P_{g,m}. \tag{7.562}$$

Summing over m, we find

$$\left[1+(\Gamma^2+4\Delta^2)\left(\sum_m\frac{w_{e,m}}{|\Omega(m,m)|^2}\right)\right]P_e=P_g, \tag{7.563}$$

where the total ground-state population is

$$P_g:=\sum_{m=-J_g}^{J_g}P_{g,m}. \tag{7.564}$$

Now using Eq. (7.408) to factor the Rabi frequencies, we can define the geometric factor

$$g:=\left[\sum_{m=-J_g}^{J_g}\frac{w_{e,m}}{\langle J_g\ m|J_e\ m;1\ 0\rangle^2}\right]^{-1}, \tag{7.565}$$

(geometric factor)

so that

$$\left(1+\frac{\Gamma^2+4\Delta^2}{g\Omega_0^2}\right)P_e=P_g, \tag{7.566}$$

where again Ω_0 is the $q = 0$ component of the Rabi-frequency vector defined by Eq. (7.402). Using $P_\mathrm{g}+P_\mathrm{e} = 1$, we can solve this equation to write the total excited-state population

$$P_\mathrm{e}(t \longrightarrow \infty) = \frac{\dfrac{g\Omega_0^2}{\Gamma^2}}{1 + \left(\dfrac{2\Delta}{\Gamma}\right)^2 + \dfrac{2g\Omega_0^2}{\Gamma^2}}.$$

(steady-state excitation, linearly polarized drive) (7.567)

Note that this expression is exactly the same as the corresponding expression (5.137) for the excited-state population of the two-level atom if we identify the two-level-atom Rabi frequency via $\Omega^2 \longrightarrow g\Omega_0^2$. Alternately, we can write the excitation in the standard form [as in Eq. (5.250) for the two-level atom]

$$P_\mathrm{e}(t \longrightarrow \infty) = \left(\frac{1}{2}\right) \frac{I/I_\mathrm{sat}}{1 + 4\Delta^2/\Gamma^2 + I/I_\mathrm{sat}}, \tag{7.568}$$

so long as we identify

$$\frac{I}{I_\mathrm{sat}} = \frac{2g\Omega_0^2}{\Gamma^2}. \tag{7.569}$$

Using Eq. (7.402), we can solve this to obtain the effective saturation intensity

$$I_\mathrm{sat} = \frac{c\epsilon_0\Gamma^2\hbar^2}{4g|\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2}$$

(effective saturation intensity, linear polarization) (7.570)

in terms of the reduced dipole matrix element for linearly-polarized excitation. Surprisingly, the entire effect of the atomic Zeeman-degenerate structure is wrapped up in the single geometric factor g, at least as far as the total excitation is concerned. For some representative values, for a $J_\mathrm{g} = 0 \longrightarrow J_\mathrm{e} = 1$ transition, $g = 1/3$; for $J_\mathrm{g} = 1 \longrightarrow J_\mathrm{e} = 2$, $g = 6/17 \approx 0.35294$; $J_\mathrm{g} = 2 \longrightarrow J_\mathrm{e} = 3$, $g = 180/461 \approx 0.39046$; $J_\mathrm{g} = 3 \longrightarrow J_\mathrm{e} = 4$, $g = 4004/9651 \approx 0.41488$; $J_\mathrm{g} = 4 \longrightarrow J_\mathrm{e} = 5$, $g = 39780/92377 \approx 0.43063$; $J_\mathrm{g} = 50 \longrightarrow J_\mathrm{e} = 51$, $g \approx 0.49271$; and $J_\mathrm{g} = 100 \longrightarrow J_\mathrm{e} = 101$, $g \approx 0.49630$.[45] Evidently, $g \longrightarrow 1/2$ as $J_\mathrm{g} \longrightarrow \infty$ (at least for the case of integer angular momenta), though this is certainly not obvious from its definition. As in the circular-polarization case, the effective squared dipole moment of the transition is $|\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2/3$ for $J_\mathrm{g} = 0$ and increases with J_g. However, it saturates at a smaller value than in the circular case, indicating that the atomic alignment is not as complete. This is not entirely surprising, as when excited with linearly polarized light the emitted photons may have any polarization, whereas with a circular pump, the emitted photons may have only one polarization.

Explicitly, then, the individual excited-state populations in steady state are

$$P_{\mathrm{e},m}(t \longrightarrow \infty) = \frac{w_{\mathrm{e},m}\dfrac{g\Omega_0^2}{\Gamma^2}}{1 + \left(\dfrac{2\Delta}{\Gamma}\right)^2 + \dfrac{2g\Omega_0^2}{\Gamma^2}},$$

(excited-state populations, linearly polarized drive) (7.571)

and using Eq. (7.562), the ground-state populations are

$$P_{\mathrm{g},m}(t \longrightarrow \infty) = \frac{w_{\mathrm{e},m}\, g\left(\dfrac{\Omega_0^2}{\Gamma^2} + \dfrac{1 + (2\Delta/\Gamma)^2}{\langle J_\mathrm{g}\, m|J_\mathrm{e}\, m; 1\, 0\rangle^2}\right)}{1 + \left(\dfrac{2\Delta}{\Gamma}\right)^2 + \dfrac{2g\Omega_0^2}{\Gamma^2}}.$$

(ground-state populations, linearly polarized drive) (7.572)

Note that unlike the excited-state case, the relative ground-state populations depend on the intensity of the field.

[45] Note that the analogous factor g_s defined by Gao, *op. cit.*, is related to the factor g here by $g = (2J_\mathrm{g} + 1)g_s$, due to the difference in conventions for the Rabi frequency (and hence the reduced dipole matrix element).

7.10 Exercises

Problem 7.1

Show that the frequency shift of the ^{133}Cs $6^2\mathrm{S}_{1/2}(F=3, m_F=0) \longrightarrow 6^2\mathrm{S}_{1/2}(F=4, m_F=0)$ "clock" transition due to a constant applied magnetic field is given by

$$\Delta\omega_{\text{clock}} = \frac{(g_J - g_I)^2 \mu_{\text{B}}^2}{2\hbar \Delta E_{\text{hfs}}} B^2 \tag{7.573}$$

for small fields, to second order the in the field B. Put in numbers appropriate for the transition ($g_J = 2.0025$, $g_I = -0.000\ 398\ 95$, $I = 7/2$, and $\Delta E_{\text{hfs}} = h \cdot 9.192\ 631\ 770$ GHz, and express the shift in Hz/G^2.

Problem 7.2

Prove the identity

$$\epsilon_{\mu\alpha\beta}\epsilon_{\mu\sigma\tau} = \delta_{\alpha\sigma}\delta_{\beta\tau} - \delta_{\alpha\tau}\delta_{\beta\sigma} \tag{7.574}$$

for the Levi-Civita symbol by writing out the "bac-cab" vector identity $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$, in terms of components.

Problem 7.3

Given the homogeneous linear system from (7.551),

$$(A_{mm'} - \delta_{mm'})\, x_{m'} = 0, \tag{7.575}$$

where m, m' are either integers or half-integers, $-J_{\text{g}} \leq m, m' \leq J_{\text{g}}$, and the tridiagonal matrix $A_{mm'}$ satisfies $A_{mm'} = A_{-m,-m'}$ and $\sum_{m'} A_{mm'} = 1$, prove by induction that

$$x_{m+1} = \frac{A_{m+1,m}}{A_{m,m+1}} x_m \tag{7.576}$$

for $m > 0$.

Problem 7.4

(a) Consider the following expression for the ground-state dipole shift of an atom with fine structure in a linearly polarized laser field of frequency ω,

$$V_{\text{dip}} = \frac{\hbar}{4} \sum_{J'} \Omega_{JJ'}^2 |\langle J\ m | J'\ m; 1\ 0\rangle|^2 \left(\frac{1}{\omega - \omega_{J'J}} - \frac{1}{\omega + \omega_{J'J}} \right), \tag{7.577}$$

where J is the angular-momentum quantum number of the ground state, the sum is over the excited-state quantum numbers J' (with an implied sum over other quantum numbers labeling relevant excited states), the frequency of the $J \longrightarrow J'$ transition is

$$\omega_{J'J} := \frac{E_{J'} - E_J}{\hbar}, \tag{7.578}$$

and the Rabi frequency is given in terms of the reduced dipole matrix element by

$$\Omega_{JJ'} := -\frac{2\langle J || d_z || J' \rangle E_0^{(+)}}{\hbar}, \tag{7.579}$$

where $E_0^{(+)}$ is the positive-rotating electric-field amplitude. Argue that this expression is correct to lowest order in the field intensity, and interpret all the factors. Note that the atom is assumed to be in the state $|J\ m\rangle$, but the shift is independent of m.

Note that in the case of a ground state with $J = 0$ or $J = 1/2$, the expression above simplifies, since $|\langle J\ m|J'\ m;1\ 0\rangle|^2 = 1/3$, so that

$$V_{\rm dip} = \frac{\hbar}{4}\sum_{J'}\frac{\Omega_{JJ'}^2}{3}\left(\frac{1}{\omega-\omega_{J'J}}-\frac{1}{\omega+\omega_{J'J}}\right). \tag{7.580}$$

The reduction in complexity is sensible in these cases because for such simple ground states there can be no dependence on the m quantum number.

(b) Consider the following estimate for the scattering rate for the above atom–field system,

$$R_{\rm sc} = \frac{1}{4}\sum_q\left[\sum_{J'}\Omega_{JJ'}\langle J\ m|J'\ m;1\ 0\rangle\sqrt{\Gamma_{J'}}\langle J'\ m|J\ m-q;1\ q\rangle \times\left(\frac{\omega}{\omega_{J'J}}\right)^{3/2}\left(\frac{1}{\omega-\omega_{J'J}}-\frac{1}{\omega+\omega_{J'J}}\right)\right]^2 \tag{7.581}$$

where we sum over the (spherical) polarization index q for the scattered light, and $\Gamma_{J'}$ is the total decay rate of level J'. Note that this expression assumes the atom to be in the particular state $|J\ m\rangle$, and thus the scattering rate should be averaged over all populated ground states, weighted by their steady-state population. This expression assumes that all spontaneous-scattering events return the atom to one of the J levels. Argue that this expression is correct to lowest order in the field intensity, subject to the above assumptions.

(c) Argue that for an atom with *hyperfine* structure, the expression for the ac Stark shift should be modified to read

$$V_{\rm dip} = \frac{\hbar}{4}\sum_{F'}\Omega_{FF'}^2|\langle F\ m_F|F'\ m_F;1\ 0\rangle|^2\left(\frac{1}{\omega-\omega_{F'F}}-\frac{1}{\omega+\omega_{F'F}}\right), \tag{7.582}$$

where the overall hyperfine Rabi frequency is

$$\begin{aligned}\Omega_{FF'} &:= -\frac{2\langle F||d_z||F'\rangle E_0^{(+)}}{\hbar}\\ &= -\frac{2\langle J||d_z||J'\rangle E_0^{(+)}}{\hbar}(-1)^{F'+J+1+I}\sqrt{(2F'+1)(2J+1)}\left\{\begin{array}{ccc} J & J' & 1\\ F' & F & I\end{array}\right\},\end{aligned} \tag{7.583}$$

where I is the nuclear angular momentum, and the hyperfine states F and F' are also labeled by J and J', respectively, while the scattering rate becomes

$$R_{\rm sc} = \frac{1}{4}\sum_{q\,F''}\left[\sum_{F'}\Omega_{FF'}\langle F\ m_F|F'\ m_F;1\ 0\rangle\sqrt{\Gamma_{F'F''}}\langle F'\ m_F|F''\ m_F-q;1\ q\rangle \times\left(\frac{\omega}{\omega_{F'F}}\right)^{3/2}\left(\frac{1}{\omega-\omega_{F'F}}-\frac{1}{\omega+\omega_{F'F}}\right)\right]^2 \tag{7.584}$$

where the decay-rate factor, with the proper emission phase factor, is

$$\sqrt{\Gamma_{F'F''}} = \sqrt{\Gamma_{J'}}(-1)^{F''+J'+1+I}\sqrt{(2F''+1)(2J'+1)}\left\{\begin{array}{ccc} J' & J'' & 1\\ F'' & F' & I\end{array}\right\} \tag{7.585}$$

for small hyperfine splittings. Note that we now sum over all possible final hyperfine levels F'', in the case of hyperfine-changing Raman scattering events. This expression also ignores cascading transitions (i.e., it only accounts for two-photon processes), and assumes that the final states $|F''\ m_F - q\rangle$ are nearly degenerate with $|F\ m_F\rangle$, so that $\omega_{F'F} \approx \omega_{F'F''}$.

Problem 7.5

The D-line fine-structure doublet in hydrogen and hydrogen-like (alkali) atoms is a transition doublet from an $S_{1/2}$ ground state to a pair of excited states, $P_{1/2}$ and a $P_{3/2}$, with the $P_{3/2}$ at higher energy.

(a) Derive an expression for the ac Stark shift for σ^+ light for the $m_J = \pm 1/2$ fine-structure ground states, in terms of the fine-structure reduced matrix element $\langle J \| \mathbf{d} \| J' \rangle$. Exclude the contributions due to states outside the D line.

(b) For ^{87}Rb, estimate the laser wavelength for which the shift of the $|J, m_J = -1/2\rangle$ ground state vanishes. Also estimate the shift of the $|J, m_F = +J\rangle$ state at this wavelength.

(c) Derive an expression for the ac Stark shift for σ^+ light for the *hyperfine* ground states, assuming the hyperfine splittings are negligibly small. Exclude the contributions due to states outside the D line.

(d) For ^{87}Rb, estimate the laser wavelengths for which the shift of the $|F, m_F = \pm F\rangle$ ground states vanish. What is the (intensity-dependent) shift of the other $|F, m_F\rangle$ states at these wavelengths?

Problem 7.6

Work out the reduced matrix element $\langle F = 1 \| \boldsymbol{\mu} \| F' = 2 \rangle$ as well as the transition matrix element $\langle F = 1, m_F = 0 | \mu_z | F' = 2, m'_F = 0 \rangle$ for the 6.8 GHz ground-state hyperfine "clock" transition in ^{87}Rb ($L = L' = 0$, $S = S' = 1/2$, $I = I' = 3/2$). Here, the magnetic-dipole operator is

$$\boldsymbol{\mu} = -\frac{\mu_B}{\hbar}(g_S \mathbf{S} + g_L \mathbf{L} + g_I \mathbf{I}) \tag{7.586}$$

[cf. Eqs. (7.307) and (7.315)].

Problem 7.7

(a) Use the Breit–Rabi formula to show that the splitting between the two states with $m_F = -1$ (i.e., $m = -1$ in the formula) reaches a minimum for some magnetic-field strength. Hence, at this field value, the splitting is insensitive to first order to fluctuations in the field. Derive an expression for the field strength and the minimum splitting.

(b) Put in numbers for the ground-state hyperfine transition, $F = 1, m_F = -1 \longrightarrow F = 2, m_F = -1$ in ^{87}Rb ($I = 3/2$), where $\Delta E_{\text{hfs}} = 6.835$ GHz, $g_I = -0.0009951$, and $g_J = 2.002331$.

Part II

Quantum Light–Matter Interactions

Chapter 8

Quantization of the Electromagnetic Field

Now we will switch to exclusively quantum-mechanical models. Before, we had treated the atom quantum mechanically, and introduced the idea of the photon, but now we will give a proper description of the quantum electromagnetic field.

8.1 Classical Electromagnetic Field

Recall that we can write the source-free Maxwell equations in free space as

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{B} &= \frac{1}{c^2} \partial_t \mathbf{E}. \end{aligned} \tag{8.1}$$

(Maxwell's equations)

In addition to the fields, we can consider the potentials $\mathbf{A}$ and ϕ. For nonrelativistic calculations in vacuum, it is convenient to choose the Coulomb gauge,[1] where $\nabla \cdot \mathbf{A} = 0$. In this gauge, $\phi = 0$ in the absence of charges (we will reintroduce the sources later when we treat the atom–field interaction), so the fields are given by

$$\mathbf{E} = -\partial_t \mathbf{A}, \qquad \mathbf{B} = \nabla \times \mathbf{A}. \tag{8.2}$$

(Coulomb gauge)

The last of the Maxwell equations thus implies the wave equation for the vector potential [using the gauge relations and $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$]:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2} \partial_t{}^2 \mathbf{A} = 0. \tag{8.3}$$

(vector-potential wave equation)

This equation is essentially the entire content of the Maxwell equations in this gauge, since the other three equations are implied by the Coulomb-gauge condition and the relations between the fields and the potentials. The first equation follows from

$$\nabla \cdot \mathbf{E} = -\partial_t \nabla \cdot \mathbf{A} = 0, \tag{8.4}$$

[1] The choice of Coulomb gauge is common in quantum optics, where the calculations are typically nonrelativistic, though relativistic field theorists prefer the **Lorenz gauge**. See L. Lorenz, ''On the Identity of the Vibrations of Light with Electrical Currents,'' *Philosophical Magazine* **34**, 287 (1867). Incidentally, the Lorenz gauge is still commonly misattributed to Hendrik A. Lorentz; see J. van Bladel, ''Lorenz or Lorentz?'' *IEEE Antennas and Propagation Magazine* **33**, No. 2 (April 1991); and Robert Nevels and Chang-Seok Shin, ''Lorenz, Lorentz, and the Gauge,'' *IEEE Antennas and Propagation Magazine* **43**, No. 3 (June 2001) (doi: 10.1109/74.934904).

the second follows from

$$\nabla \cdot \mathbf{B} = \nabla \cdot (\nabla \times \mathbf{A}) = 0, \tag{8.5}$$

and the third follows from

$$\nabla \times \mathbf{E} = -\partial_t \nabla \times \mathbf{A} = -\partial_t \mathbf{B}. \tag{8.6}$$

The vector potential $\mathbf{A}$ thus compactly represents the fields. It also turns out to be fundamentally important in preserving the locality of quantum-mechanical particle-field interactions, so we will often work with it.

8.2 Hamiltonian Structure of the Classical Electromagnetic Field

In order to quantize the field, we need the Lagrangian (and action principle), to identify the canonical coordinates and the Hamiltonian for the field, and then promote the canonical coordinates to operators.[2] Before doing this, though, we will review briefly the ideas that underlie the Hamiltonian.

8.2.1 Variational Calculus

First off, a (real-valued) **functional**[3] is a function $F : \mathbb{F} \longrightarrow \mathbb{R}$, where $\mathbb{F}$ is a space of functions. Usually a functional will involve an integral to reduce functions to scalars, as in the action functional that we consider below. Let $x(t) \in \mathbb{F}$; then $F[x]$ is a real number, and the **first variation** of the functional $F[x]$ is given by

$$\delta F[x;\delta x] := \lim_{\epsilon\to 0} \frac{F[x+\epsilon\delta x] - F[x]}{\epsilon} = \frac{d}{d\epsilon} F[x+\epsilon\delta x]\bigg|_{\epsilon=0}, \tag{8.7}$$
(first variation)

where ϵ and δx are subject to the constraint that $x + \epsilon\delta x \in \mathbb{F}$. The first variation is essentially the ''linear response'' of $F[x]$ to a small perturbation $x(t) \longrightarrow x(t) + \epsilon\delta x(t)$.

We will then define the **functional derivative** $\delta F/\delta x$ such that

$$\left\langle \frac{\delta F}{\delta x}, \delta x \right\rangle := \int_{t_1}^{t_2} \frac{\delta F}{\delta x}\, \delta x\, dt := \delta F[x;\delta x]. \tag{8.8}$$
(functional derivative)

Note that the brackets here denote an inner product of two vectors (here, functions), which is defined by the integral over the two vectors (functions). Thus, the functional derivative is the part of the first variation after dropping the integral and the variation δx.

We can generalize this derivative in a couple of ways. A functional may depend on derivatives of the function; suppose that

$$F[x, x_t, x_{tt}, \ldots; t] = \int_{t_1}^{t_2} f(t, x, x_t, x_{tt}, \ldots)\, dt, \tag{8.9}$$

where $x_t \equiv \partial x/\partial t$. Then the first variation becomes

$$\delta F[x, x_t, x_{tt}, \cdots; t] = \int_{t_1}^{t_2} \left[\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial x_t}\delta x_t + \frac{\partial f}{\partial x_{tt}}\delta x_{tt} + \cdots\right] dt. \tag{8.10}$$

Integrating by parts,

$$\delta F[x, x_t, x_{tt}, \cdots; t] = \int_{t_1}^{t_2} \left[\frac{\partial f}{\partial x} - \frac{d}{dt}\frac{\partial f}{\partial x_t} + \frac{d^2}{dt^2}\frac{\partial f}{\partial x_{tt}} + \cdots\right] \delta x\, dt + \left[\frac{\partial f}{\partial x_t}\delta x + \cdots\right]_{t_1}^{t_2}, \tag{8.11}$$

[2] Dirac discusses this basic ''quantization recipe'' and its advantages compared to alternatives in: Paul A. M. Dirac, *Lectures on Quantum Mechanics* (Belfer Graduate School of Science, 1964), Lecture 1.

[3] For more details on variational calculus and action principles, see P. J. Morrison, ''Hamiltonian description of the ideal fluid,'' *Reviews of Modern Physics* **70**, 467 (1998) (doi: 10.1103/RevModPhys.70.467).

and typically the variation is arranged such that the surface terms vanish. This is usually enforced via **fixed-endpoint variations** $\delta x(t_1) = \delta x(t_2) = 0$. Then, using the definition (8.8), we can read off the functional derivative,

$$\frac{\delta F}{\delta x} = \frac{\partial f}{\partial x} - \frac{d}{dt}\frac{\partial f}{\partial x_t} + \frac{d^2}{dt^2}\frac{\partial f}{\partial x_{tt}} + \cdots, \tag{8.12}$$

being just the remaining bracketed expression in the integrand.

The other generalization to the functional derivative is to many functions and dimensions, which occurs in the natural way,

$$\left\langle \frac{\delta F}{\delta \mathbf{y}}, \delta \mathbf{y} \right\rangle := \int_D \frac{\delta F}{\delta y_\alpha} \delta y_\alpha \, d^n x := \delta F[\mathbf{y}; \delta \mathbf{y}] \tag{8.13}$$

for functions $\mathbf{y}(\mathbf{x})$, where D is the domain of integration, and recall that we are implicitly summing over the repeated index α.

8.2.2 Action Principles

The calculus of variations is important in physics in setting up **action principles**, where equations of motion follow from the stationary points of some functional. For example, in a least-action principle, functions that minimize the functional correspond to physical solutions.

As a concrete example, we can state an ''action principle'' for straight lines, in the sense of constructing an action principle that says that the shortest path between two points p_1 and p_2 is a straight line. That is, consider the length functional ℓ for the curve $y(x)$, with $x_{1,2}$ marking the points $p_{1,2}$:

$$\ell[y] = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx. \tag{8.14}$$

Then the condition $\delta \ell = 0$ (under fixed-endpoint variations) implies that $y(x)$ is a straight line. We can see this from

$$\delta \ell[y] = \int_{x_1}^{x_2} \left[-\frac{d}{dx} \frac{y_x}{\sqrt{1 + y_x^2}} \right] dx = 0. \tag{8.15}$$

Setting the integrand to zero, after a bit of algebra we see that the only way the integrand can vanish is for $y = \alpha x + \beta$, where α and β are constants.

8.2.2.1 Lagrangian

In general, we will define a scalar **Lagrangian function** $L(\mathbf{q}, \dot{\mathbf{q}}; t)$ to describe our system. Naturally, this may be generalized to higher time derivatives. The **action functional** is defined by the integral

$$S[L] := \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}; t)\, dt, \tag{8.16}$$

where the Lagrangian L is typically of the form $L = T(\dot{\mathbf{q}}) - V(\mathbf{q})$ in particle mechanics. Then **Hamilton's principle**

$$\delta S[L] = 0 \tag{8.17}$$
(Hamilton's principle)

implies the **Euler–Lagrange equation**

$$\frac{\partial L}{\partial q^\alpha} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}^\alpha} = 0 \tag{8.18}$$
(Euler–Lagrange equation)

under the condition of fixed-endpoint variation $\delta \mathbf{q}(t_1) = \delta \mathbf{q}(t_2) = 0$, by applying the vector generalization of Eq. (8.11). This proceeds along the lines of Eq. (8.13), noting that each variation $\delta q^\alpha(t)$ is independent.

For a particle Lagrangian of the form

$$L = \frac{1}{2}m\dot{q}^2 - V(q), \tag{8.19}$$

the Euler–Lagrange equation implies

$$m\ddot{q} = -\partial_q V, \tag{8.20}$$

which is Newton's Second Law.

8.2.2.2 Hamiltonian

The Hamiltonian is defined by a Legendre transformation of the Lagrangian via

$$H(\mathbf{q}, \mathbf{p}; t) := \dot{q}^\alpha p_\alpha - L(\mathbf{q}, \dot{\mathbf{q}}; t), \tag{8.21}$$
(Hamiltonian)

where the **conjugate momentum** to the generalized coordinate q^α is

$$p_\alpha := \frac{\partial L}{\partial \dot{q}^\alpha}. \tag{8.22}$$
(conjugate momentum)

The conjugate momentum is used to eliminate dependence on $\dot{\mathbf{q}}$ in the Hamiltonian introduced by the Lagrangian. Then the **phase-space action**

$$S[\mathbf{q}, \mathbf{p}] := \int_{t_1}^{t_2} \left[\dot{q}^\alpha p_\alpha - H(\mathbf{q}, \mathbf{p}; t)\right] dt \tag{8.23}$$
(phase-space action)

(note the bracketed quantity is basically the Lagrangian) along with the action principle

$$\frac{\delta S}{\delta q^\alpha} = 0, \qquad \frac{\delta S}{\delta p_\alpha} = 0, \tag{8.24}$$
(phase-space action principle)

imply **Hamilton's equations**,

$$\partial_t p_\alpha = -\frac{\partial H}{\partial q^\alpha}, \qquad \partial_t q^\alpha = \frac{\partial H}{\partial p_\alpha}, \tag{8.25}$$
(Hamilton's equations)

again under the condition of fixed position endpoints $\delta q(t_1) = \delta q(t_2) = 0$, but now $p(t_1)$ and $p(t_2)$ are allowed to vary. For example, for the particle Lagrangian (8.19), the conjugate momentum is $p = \partial L/\partial \dot{q} = m\dot{q}$, so the Hamiltonian becomes

$$H = \frac{p^2}{2m} + V(q). \tag{8.26}$$

Then Hamilton's equations become

$$\partial_t p = -\partial_q H = -\partial_q V, \qquad \partial_t q = \partial_p H = \frac{p}{m}. \tag{8.27}$$

The first is again Newton's Second Law, while the second is just the definition of the momentum.

8.2.3 Electromagnetic Lagrangian and Hamiltonian

We now identify the Lagrangian for the electromagnetic field, which in terms of the vector potential is

$$L = \frac{\epsilon_0}{2}\int d^3r \left[(\partial_t \mathbf{A})^2 - c^2(\nabla \times \mathbf{A})^2\right]. \tag{8.28}$$
(electromagnetic Lagrangian)

We can see that this is the appropriate Lagrangian by employing the Euler–Lagrange equation (generalized for functional derivatives, since the generalized coordinates are fields)

$$\frac{\delta L}{\delta \mathbf{A}} - \partial_t \frac{\delta L}{\delta(\partial_t \mathbf{A})} = 0, \tag{8.29}$$

where we take the generalized coordinate to be the vector potential $\mathbf{A}$. Computing the functional derivatives (see below), we find

$$-\epsilon_0 c^2(\nabla \times (\nabla \times \mathbf{A})) - \partial_t \epsilon_0(\partial_t \mathbf{A}) = 0, \tag{8.30}$$

which we see is equivalent to the wave equation (8.3) for the vector potential.

Since $\mathbf{A}$ is the ''position'' coordinate, which we can see from the ''kinetic energy'' term $(\partial_t \mathbf{A})^2$ term in the Lagrangian, the conjugate momentum is given by

$$\boxed{\mathbf{\Pi} := \frac{\delta L}{\delta(\partial_t \mathbf{A})} = \epsilon_0 \partial_t \mathbf{A} = -\epsilon_0 \mathbf{E}.} \tag{8.31}$$
(conjugate momentum to $\mathbf{A}$)

Then the Hamiltonian is given by the Legendre transform of the Lagrangian:

$$\begin{aligned} H &:= \int d^3r\, (\mathbf{\Pi} \cdot \partial_t \mathbf{A}) - L \\ &= \epsilon_0 \int d^3r\, (\partial_t \mathbf{A})^2 - L \\ &= \frac{\epsilon_0}{2} \int d^3r \left[(\partial_t \mathbf{A})^2 + c^2(\nabla \times \mathbf{A})^2\right] \\ &= \frac{\epsilon_0}{2} \int d^3r \left[E^2 + c^2(\nabla \times \mathbf{A})^2\right]. \end{aligned} \tag{8.32}$$

In terms of the conjugate variables, the Hamiltonian is

$$\boxed{H = \int d^3r \left[\frac{\Pi^2}{2\epsilon_0} + \frac{1}{2}\epsilon_0 c^2(\nabla \times \mathbf{A})^2\right],} \tag{8.33}$$
(electromagnetic Hamiltonian)

while in terms of the fields,

$$\boxed{H = \frac{\epsilon_0}{2} \int d^3r \left[E^2 + c^2 B^2\right],} \tag{8.34}$$
(electromagnetic Hamiltonian)

it is clear that the Hamiltonian is just the total energy of the electromagnetic field.

Hamilton's equations then recover the Maxwell equations. The first Hamilton equation is

$$\partial_t \mathbf{\Pi} = -\frac{\delta H}{\delta \mathbf{A}}, \tag{8.35}$$

which gives

$$-\epsilon_0 \partial_t \mathbf{E} = -\epsilon_0 c^2 \nabla \times (\nabla \times \mathbf{A}) = -\epsilon_0 c^2 \nabla \times \mathbf{B}, \tag{8.36}$$

and thus yields the last Maxwell equation of Eqs. (8.1). The other Hamilton equation,

$$\partial_t \mathbf{A} = \frac{\delta H}{\delta \mathbf{\Pi}}, \tag{8.37}$$

contains essentially no information, as it implies

$$-\epsilon_0 \mathbf{E} = \frac{\delta H}{-\delta \mathbf{E}}. \tag{8.38}$$

The other three Maxwell equations, as we indicated before, follow simply from the fact that the fields derive from the vector potential. Thus, we see how the Hamiltonian structure of the electromagnetic field arises within the Coulomb gauge, which will now allow us to quantize the field.

8.2.3.1 Electromagnetic Functional Derivatives

Here we compute the functional derivatives that arise in Eqs. (8.29), (8.35), and (8.37) as follows. Recall that a **functional** is a function that maps functions to scalars. The Lagrangian and Hamiltonian here satisfy this definition, due to the spatial integration, in the same way as the action integral above. The first variation of the Lagrangian is

$$\delta L(\mathbf{A}, \partial_t \mathbf{A}) := \frac{d}{d\epsilon} L(\mathbf{A} + \epsilon\delta\mathbf{A}, \partial_t \mathbf{A} + \epsilon\delta(\partial_t \mathbf{A}))\Big|_{\epsilon=0} \tag{8.39}$$

for variations $\delta\mathbf{A}$ and $\delta(\partial_t \mathbf{A})$ of the potential. This is easy to evaluate, as this is just a linearization in the variations, giving

$$\delta L(\mathbf{A}, \partial_t \mathbf{A}) = \epsilon_0 \int d^3r \left[(\partial_t \mathbf{A}) \cdot \delta(\partial_t \mathbf{A}) - c^2 (\nabla \times \mathbf{A}) \cdot (\nabla \times \delta\mathbf{A})\right]. \tag{8.40}$$

Using the vector identity

$$\nabla \cdot (\delta\mathbf{A} \times (\nabla \times \mathbf{A})) = (\nabla \times \mathbf{A}) \cdot (\nabla \times \delta\mathbf{A}) - \delta\mathbf{A} \cdot (\nabla \times (\nabla \times \mathbf{A})), \tag{8.41}$$

we can use the divergence theorem and integrate by parts to obtain

$$\delta L(\mathbf{A}, \partial_t \mathbf{A}) = \epsilon_0 \int d^3r \left[(\partial_t \mathbf{A}) \cdot \delta(\partial_t \mathbf{A}) - c^2 (\nabla \times (\nabla \times \mathbf{A})) \cdot \delta\mathbf{A}\right] + \epsilon_0 c^2 \int_{\text{surface}} (\delta\mathbf{A} \times \mathbf{B}) \cdot d\mathbf{a}. \tag{8.42}$$

We will assume a fixed-boundary variation, so that $\delta\mathbf{A} = 0$ on the surface of the integration volume, so that the surface term vanishes.

We defined the functional derivatives in terms of the first variation via inner products with the variations

$$\delta L(\mathbf{A}, \partial_t \mathbf{A}) =: \left\langle \frac{\delta L}{\delta(\partial_t \mathbf{A})}, \delta(\partial_t \mathbf{A}) \right\rangle + \left\langle \frac{\delta L}{\delta \mathbf{A}}, \delta\mathbf{A} \right\rangle. \tag{8.43}$$

Interpreting the spatial integral with the dot product as the inner product here, we can write down the functional derivatives:

$$\begin{aligned} \frac{\delta L}{\delta(\partial_t \mathbf{A})} &= \epsilon_0 \partial_t \mathbf{A} \\ \frac{\delta L}{\delta \mathbf{A}} &= -\epsilon_0 c^2 \nabla \times (\nabla \times \mathbf{A}). \end{aligned} \tag{8.44}$$

Note that from the form of the Hamiltonian, the same functional derivatives of the Hamiltonian have the same forms, except that the functional derivative with respect to $\mathbf{A}$ changes sign.

8.3 Quantization of a Single Field Mode

Now we can proceed to quantize the field, considering only a single field mode.[4] The idea is to take advantage of the fact that we have a linear field theory (because the QED Hamiltonian is quadratic in $\mathbf{\Pi}$ and $\mathbf{A}$), so we can perform separation of variables and decompose the field operators into noninteracting normal modes. These normal modes are much simpler to deal with than the full fields. Taking an implicit Fourier transform, we can assume a monochromatic solution of frequency ω:

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) &= \alpha(t)\mathbf{f}(\mathbf{r}) + \text{c.c.} \\ &= \alpha(0)e^{-i\omega t}\mathbf{f}(\mathbf{r}) + \text{c.c.} \end{aligned} \tag{8.45}$$

[4]For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Section 2.4, p. 40. The first quantum treatment of the electromagnetic field was M. Born, W. Heisenberg, and P. Jordan, ''Zur Quantenmechanik II,'' *Zeitschrift für Physik* **35**, 557 (1926) (doi: 10.1007/BF01379806). Other important early papers on field quantization include P. A. M. Dirac, ''The Quantum Theory of the Emission and Absorption of Radiation,'' *Proceedings of the Royal Society of London. Series A* **114**, 243 (1927) (doi: 10.1098/rspa.1927.0039); W. Heisenberg and W. Pauli, ''Zur Quantendynamik der Wellenfelder,'' *Zeitschrift für Physik* **56**, 1 (1929) (doi: 10.1007/BF01340129); and Enrico Fermi, ''Sopra l'etettrodinamica quantistica,'' *Atti della Reale Accademia Nazionale dei Lincei*, **12**, 431 (1930).

The space and time dependences are now explicitly separated—the separation of variables is allowed by the form of the wave equation for $\mathbf{A}$. The function $\mathbf{f}(\mathbf{r})$ is the **mode function**, which contains all the spatial dependence of the field. (In general, there are many possible mode functions for a given frequency, so we will simply choose one.) We assume them to be normalized such that

$$\int d^3r\, |\mathbf{f}(\mathbf{r})|^2 = 1. \tag{8.46}$$

The wave equation (8.3) then implies that the mode function satisfies the Helmholtz equation

$$\left(\nabla^2 + k^2\right)\mathbf{f}(\mathbf{r}) = 0, \tag{8.47}$$

(Helmholtz equation for mode function)

where $k = \omega/c$.

We can now simplify the field Hamiltonian in the case of a single mode. We will need the relation

$$\begin{aligned}\int d^3r\, (\nabla \times \mathbf{A})^2 &= \int d^3r\, \mathbf{A} \cdot (\nabla \times (\nabla \times \mathbf{A})) \\ &= -\int d^3r\, \mathbf{A} \cdot (\nabla^2 \mathbf{A}) \\ &= k^2 \int d^3r\, A^2,\end{aligned} \tag{8.48}$$

where we again integrated by parts and discarded the surface term, then used the fact that the vector potential (8.45) satisfies the Helmholtz equation. Then the Hamiltonian becomes

$$H = \int d^3r\, \left[\frac{\Pi^2}{2\epsilon_0} + \frac{1}{2}\epsilon_0\omega^2 A^2\right]. \tag{8.49}$$

This form suggests the Hamiltonian for a harmonic oscillator of frequency ω and mass ϵ_0, again with momentum $\mathbf{\Pi} = -\epsilon_0\mathbf{E}$ and position $\mathbf{A}$.

Of course, the spatial integral does not appear in the usual harmonic-oscillator Hamiltonian. However, the spatial dependence of the mode is fixed, so we can go ahead and carry out the integral to complete the analogy. Noting that the mode electric field is given by

$$\mathbf{E} = -\partial_t \mathbf{A} = -i\omega\alpha(t)\mathbf{f}(\mathbf{r}) + \text{c.c.}, \tag{8.50}$$

we can define a momentum coordinate to be the temporal part of $\mathbf{\Pi}$, but with a different phase choice for $\alpha(0)$,

$$p := -\omega\epsilon_0[\alpha(t) + \text{c.c.}], \tag{8.51}$$

and a position coordinate to be the temporal part of $\mathbf{A}$, with the same phase choice,

$$q := -[i\alpha(t) + \text{c.c.}], \tag{8.52}$$

so that $p = \epsilon_0\partial_t q$. Then we can rewrite the Hamiltonian as

$$\begin{aligned}H &= \frac{\epsilon_0}{2}\int d^3r\, \left[E^2 + \omega^2 A^2\right] \\ &= \frac{\epsilon_0}{2}\int d^3r\, \left[4\omega^2|\alpha(t)|^2|\mathbf{f}(\mathbf{r})|^2\right] \\ &= \frac{\epsilon_0}{2}\left[4\omega^2|\alpha(t)|^2\right] \\ &= \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2,\end{aligned} \tag{8.53}$$

with $m = \epsilon_0$. Here, the connection to the harmonic oscillator is more obvious, and the variables p and q are still clearly canonically conjugate.

The usual quantum relations for the coordinate operators in terms of the creation and annihilation operators are

$$\begin{aligned} q &= \sqrt{\frac{\hbar}{2m\omega}}\left(a + a^\dagger\right) \\ p &= \sqrt{\frac{m\hbar\omega}{2}}\left(\frac{a - a^\dagger}{i}\right), \end{aligned} \tag{8.54}$$

and the usual commutation relation is $[a, a^\dagger] = 1$. Comparing the relations (8.51) and (8.52) to the quantum relations, we can identify

$$\alpha(t) \longrightarrow i\sqrt{\frac{\hbar}{2\omega\epsilon_0}}a(t), \tag{8.55}$$

(quantization replacement, single field mode)

which will be our ''recipe'' for quantization: after a rescaling, we replace the function $\alpha(t)$ with the annihilation operator $a(t)$ (and a scaling factor). We can thus write the quantum fields as

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) &= i\sqrt{\frac{\hbar}{2\omega\epsilon_0}}\,\mathbf{f}(\mathbf{r})\,a(t) + \text{H.c.} \\ \mathbf{E}(\mathbf{r}, t) &= -\sqrt{\frac{\hbar\omega}{2\epsilon_0}}\,\mathbf{f}(\mathbf{r})\,a(t) + \text{H.c.} \\ \mathbf{B}(\mathbf{r}, t) &= i\sqrt{\frac{\hbar}{2\omega\epsilon_0}}\,[\nabla \times \mathbf{f}(\mathbf{r})]\,a(t) + \text{H.c.} \end{aligned} \tag{8.56}$$

(quantized fields)

Note that the *mode functions* in the quantum fields are entirely classical; the *quantum* part of the field modes only enters in the ''time dependence.''

Also, with the relations (8.54), the Hamiltonian (8.53) becomes

$$\begin{aligned} H &= \frac{\hbar\omega}{2}\left(a^\dagger a + aa^\dagger\right) \\ &= \hbar\omega\left(a^\dagger a + \frac{1}{2}\right). \end{aligned} \tag{8.57}$$

(Hamiltonian for single field mode)

Of course, this is the usual quantum Hamiltonian for the harmonic oscillator. Thus, we have explicitly shown that a *single* field mode behaves both classically and quantum-mechanically as an ordinary harmonic oscillator, and we have defined the annihilation operator for this oscillator. Of course, the energy level $|n\rangle$ is colloquially called the ''number of **photons**,'' and $a(t)$ is the annihilation operator that removes a photon from the field.

8.4 Quantization of Many Modes

Classically, there are orthogonal modes corresponding to different wave vectors $\mathbf{k}$: different frequencies correspond to different magnitudes $|\mathbf{k}|$, and for the same frequency, different directions correspond to different spatial mode profiles. Also, there are two distinct polarizations for each possible wave vector (due the the three-vector nature of $\mathbf{A}$ and the constraint $\nabla \cdot \mathbf{A} = 0$), which we will label by ζ ($\zeta = 1$ or 2). In this case, we have the orthonormal mode functions $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$, satisfying

$$\int_V d^3r\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) \cdot \mathbf{f}^*_{\mathbf{k}',\zeta'}(\mathbf{r}) = \delta^3_{\mathbf{k},\mathbf{k}'}\delta_{\zeta,\zeta'}, \tag{8.58}$$

(mode orthnormality)

where V is the volume of the cavity enclosing the mode functions (the **quantization volume**). Then each mode is completely independent of the others, and by extending the above analysis, the Hamiltonian

becomes

$$H = \sum_{\mathbf{k},\zeta} \hbar\omega_{\mathbf{k}} \left(a^\dagger_{\mathbf{k},\zeta} a_{\mathbf{k},\zeta} + \frac{1}{2} \right), \tag{8.59}$$
(many-mode Hamiltonian)

where $\omega_{\mathbf{k}} = c|\mathbf{k}|$. In this case, we also have

$$\left[a_{\mathbf{k},\zeta}, a^\dagger_{\mathbf{k}',\zeta'} \right] = \delta^3_{\mathbf{k},\mathbf{k}'} \delta_{\zeta,\zeta'} \tag{8.60}$$
(bosonic commutation relation)

for the bosonic commutation relations for the field operators. We can now write the field operators as a sum over the field modes, including the operator parts:

$$\begin{aligned}
\mathbf{A}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\, a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\
\mathbf{E}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} -\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\, a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\
\mathbf{B}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\, [\nabla \times \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})]\, a_{\mathbf{k},\zeta}(t) + \text{H.c.}
\end{aligned} \tag{8.61}$$
(quantized fields)

Since we have quantized the classical field theory while preserving its Hamiltonian structure, we have performed **canonical quantization** or **second quantization** of the electromagnetic field.

8.4.1 Example: Quantization in a Perfectly Conducting Box

For a perfectly conducting box of lengths L_x, L_y, and L_z (with one corner at the origin), the transverse components of the electric fields must vanish at the boundaries, and thus the mode functions become

$$\begin{aligned}
\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \sqrt{\frac{8}{V}} \big[&\hat{x}(\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{x}) \cos(k_x x) \sin(k_y y) \sin(k_z z) \\
&+ \hat{y}(\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{y}) \sin(k_x x) \cos(k_y y) \sin(k_z z) \\
&+ \hat{z}(\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{z}) \sin(k_x x) \sin(k_y y) \cos(k_z z) \big],
\end{aligned} \tag{8.62}$$
(mode functions, perfectly conducting box)

where $V := L_x L_y L_z$, $\hat{\varepsilon}_{\mathbf{k},\zeta}$ is the unit polarization vector of the mode, and the $\hat{x}_j$ are the Cartesian unit vectors along the x_j-direction. The wave vectors are given by

$$k_x = \frac{\pi n_x}{L_x}, \qquad k_y = \frac{\pi n_y}{L_y}, \qquad k_z = \frac{\pi n_z}{L_z}, \tag{8.63}$$

where the n_α are nonnegative integers (and not all zero). Since $\nabla \cdot \mathbf{E} = 0$, we have

$$\frac{n_x E_x}{L_x} + \frac{n_y E_y}{L_y} + \frac{n_z E_z}{L_z} = 0, \tag{8.64}$$

which constrains the number of independent polarizations per (n_x, n_y, n_z) triple to at most 2.

8.4.2 Example: Quantization in Free Space

Quantization in free space[5] is similar to the case of the box cavity, and in fact free-space results can be obtained with some care as the limit of a box where $V \longrightarrow \infty$.[6] However, it is aesthetically better to have

[5] For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Section 2.5, p. 43.

[6] Note that in principle, we should *always* quantize in free space, if we treat all matter quantum mechanically. However, it should be a good approximation to treat macroscopic ''boundaries'' of matter in terms of classical boundary conditions, which justifies our quantization inside a cavity and in half space. See P. W. Milonni, ''Casimir forces without the vacuum radiation field,'' *Physical Review A*, **25**, 1315 (1982) (doi: 10.1103/PhysRevA.25.1315); and J. D. Cresser, ''Unequal Time EM Field Commutators in Quantum Optics,'' *Physica Scripta* **T21**, 52 (1988) (doi: 10.1088/0031-8949/1988/T21/010).

mode functions that have amplitudes that are independent of $\mathbf{r}$. To avoid problems with the normalization, we will impose a fictitious array of boxes filling free space, each of volume $V = L^3$, with periodic boundary conditions on the vector potential

$$\mathbf{A}(x+L,y,z,t) = \mathbf{A}(x,y+L,z,t) = \mathbf{A}(x,y,z+L,t) = \mathbf{A}(x,y,z,t), \tag{8.65}$$

which is satisfied by the mode functions

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \frac{1}{\sqrt{V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{8.66}$$
(mode functions, free space)

where the components of the wave vector $\mathbf{k}$ is given by

$$k_{x,y,z} = \frac{2\pi n_{x,y,z}}{L} = \frac{2\pi n_{x,y,z}}{\sqrt[3]{V}}, \tag{8.67}$$

and the n_{x^j} are *any* integers.

We can write out the potential explicitly here as

$$\begin{aligned}\mathbf{A}_{\mathbf{k},\zeta}(\mathbf{r},t) &= i\sqrt{\frac{\hbar}{2\omega\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\ &= i\sqrt{\frac{\hbar}{2\omega\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}a_{\mathbf{k},\zeta}(0) + \text{H.c.},\end{aligned} \tag{8.68}$$

and the electric field similarly becomes

$$\begin{aligned}\mathbf{E}_{\mathbf{k},\zeta}(\mathbf{r},t) &= \sqrt{\frac{\hbar\omega}{2\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\ &= \sqrt{\frac{\hbar\omega}{2\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}a_{\mathbf{k},\zeta}(0) + \text{H.c.}.\end{aligned} \tag{8.69}$$

Strictly speaking, we must let $V \longrightarrow \infty$ in any calculation in free space, unless the problem obeys periodic boundary conditions. This limit is straightforward, where, for example, $\delta^3_{\mathbf{k},\mathbf{k}'} \longrightarrow \delta^3(\mathbf{k}-\mathbf{k}')$, and the sum over modes changes to an integral.

8.4.3 Example: Quantization in Half Space

A case intermediate to the above two is the case of a perfectly conducting plane defined by $z = 0$, where we quantize the half-space $z > 0$. The parallel components of the field must vanish on the plane, and so we choose

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \sqrt{\frac{2}{V}}\left(\hat{\varepsilon}_{\mathbf{k},\zeta,\parallel}\sin k_z z - i\hat{\varepsilon}_{\mathbf{k},\zeta,z}\cos k_z z\right)e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}, \tag{8.70}$$
(half-space mode functions)

where the subscript $\parallel$ denotes the part of the vector parallel to the surface, $\mathbf{v}_\parallel = [\hat{x}(\hat{x}\cdot\mathbf{v}) + \hat{y}(\hat{y}\cdot\mathbf{v})]$, while the z subscript denotes the part of the vector perpendicular to the surface, $\mathbf{v}_z = \hat{z}(\hat{z}\cdot\mathbf{v})$. Here we have quantized inside a cube of length $L = V^{1/3}$, imposing periodic boundary conditions in the x- and y-directions and conducting boundary conditions in the z-direction. The components of the wave vector are then given by

$$k_x = \frac{2\pi n_x}{L}, \qquad k_y = \frac{2\pi n_y}{L}, \qquad k_z = \frac{\pi n_z}{L}, \tag{8.71}$$

where n_x and n_y are any integers, and n_z is nonnegative. The transverse condition $\nabla\cdot\mathbf{E} = 0$ then implies

$$n_x(\hat{x}\cdot\hat{\varepsilon}_{\mathbf{k},\zeta}) + n_y(\hat{y}\cdot\hat{\varepsilon}_{\mathbf{k},\zeta}) + n_z(\hat{z}\cdot\hat{\varepsilon}_{\mathbf{k},\zeta}) = 0, \tag{8.72}$$

again restricting the number of distinct polarizations to two.

Note that we may also write Eq. (8.70) in the form

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = -\frac{i}{\sqrt{2V}}\left(\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}} - \hat{\varepsilon}_{\mathbf{k}^-,\zeta}e^{i\mathbf{k}^-\cdot\mathbf{r}}\right), \tag{8.73}$$

where

$$\mathbf{k}^- = k_x\hat{x} + k_y\hat{y} - k_z\hat{z} \tag{8.74}$$

is the wave vector reflected through the conducting plane, and

$$\hat{\varepsilon}_{\mathbf{k}^-,\zeta} = \hat{\varepsilon}_{\mathbf{k},\zeta,x}\hat{x} + \hat{\varepsilon}_{\mathbf{k},\zeta,y}\hat{y} - \hat{\varepsilon}_{\mathbf{k},\zeta,z}\hat{z} \tag{8.75}$$

is the unit polarization vector with the same reflection. In this form, it is clear that each mode consists of a plane wave propagating along $\mathbf{k}$, along with its reflection off the mirror, which travels along $\mathbf{k}^-$. For the two polarizations, it is conventional to choose one polarization to be oriented parallel to the mirror. This is the **transverse electric (TE) polarization**—or **S-polarization**, for *senkrecht* or perpendicular to the plane of incidence of the incident wave—and is given by

$$\mathbf{f}_{\mathbf{k}}^{(\mathrm{TE})}(\mathbf{r}) = \sqrt{\frac{2}{V}}\left(\hat{k}_\parallel \times \hat{z}\sin k_z z\right)e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}. \tag{8.76}$$
(TE polarization)

Here $\hat{k}_\parallel$ is the projection of the $\mathbf{k}$ into the plane of the conductor, renormalized to unit length. The other polarization, the **transverse magnetic (TM)**—or **P-polarization**, for parallel to the plane of incidence—is orthogonal to both the TE polarization vector and $\mathbf{k}$, and is given by

$$\mathbf{f}_{\mathbf{k}}^{(\mathrm{TM})}(\mathbf{r}) = \sqrt{\frac{2}{V}}\left(\hat{k}_\parallel\frac{k_z}{k}\sin k_z z + i\hat{z}\frac{k_\parallel}{k}\cos k_z z\right)e^{i\mathbf{k}_\parallel\cdot\mathbf{r}}. \tag{8.77}$$
(TM polarization)

It thus always suffices to assume that the unit polarization vectors $\hat{\varepsilon}_{\mathbf{k},\zeta}$ are real.

8.4.4 Example: Quantization in a Spherical Cavity

An important but much more complicated cavity than the rectangular one is the spherical cavity. In spherical coordinates, the Laplacian in the Helmholtz equation (8.47) is

$$\begin{aligned}\nabla^2 &= \frac{1}{r^2}\partial_r r^2\partial_r + \frac{1}{r^2\sin\theta}\partial_\theta\sin\theta\,\partial_\theta + \frac{1}{r^2\sin^2\theta}\partial_\phi^2 \\ &= \frac{1}{r}\partial_r^2 r + \frac{1}{r^2\sin\theta}\partial_\theta\sin\theta\,\partial_\theta + \frac{1}{r^2\sin^2\theta}\partial_\phi^2,\end{aligned} \tag{8.78}$$

where the derivative operators are understood to operate on everything to the right, including an arbitrary test function.

8.4.4.1 Scalar Field

In the scalar case, we can separate the Helmholtz equation $(\nabla^2 + k^2)\psi = 0$ by taking the solution to be the product $\psi(\mathbf{r}) = R(r)\Theta(\theta)\Phi(\phi)$. Substitution of this *ansatz* into the scalar Helmholtz equation yields the equations

$$\begin{gathered}r^2\partial_r^2 R + 2r\partial_r R + (k^2r^2 - c_1^2)R = 0 \\ \frac{1}{\sin\theta}\partial_\theta\left(\sin\theta\,\partial_\theta\Theta\right) + \left(c_1^2 - \frac{c_2^2}{\sin^2\theta}\right)\Theta = 0 \\ \partial_\phi^2\Phi + c_2^2\Phi = 0,\end{gathered} \tag{8.79}$$

where c_1 and c_2 are separation constants [e.g., if $R(r) = \Theta(\theta)$, then there is a constant c such that $R(r) = c = \Theta(\theta)$]. The last equation is easiest to solve, giving

$$\Phi(\phi) = e^{\pm ic_2\phi}. \tag{8.80}$$

Since $\Phi(\phi)$ must be 2π periodic, clearly, c_2 must be an integer, $c_2 = m$, so that

$$\Phi_m(\phi) = e^{im\phi}, \tag{8.81}$$

where m can be positive or negative (or zero). The second separation equation thus becomes, letting $x = \cos\theta$ and regarding Θ to be function of x,

$$\left(1 - x^2\right)\partial_x^2\Theta - 2x\partial_x\Theta + \left(c_1^2 - \frac{m^2}{1-x^2}\right)\Theta = 0. \tag{8.82}$$

Taking $c_1^2 = l(l+1)$, this equation becomes the **general Legendre equation**,

$$\left(1 - x^2\right)\partial_x^2\Theta - 2x\partial_x\Theta + \left(l(l+1) - \frac{m^2}{1-x^2}\right)\Theta = 0. \tag{8.83}$$

which has nondivergent solutions on the domain $[-1, 1]$ only if l is a nonnegative integer and $|m| \le l$. These solutions are the **associated Legendre functions**, denoted by $P_l^m(x)$. They are given explicitly by

$$\begin{aligned} P_l^m(x) &= \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\partial_x^{l+m}(x^2-1)^l \qquad (m \ge 0) \\ P_l^m(x) &= (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x). \end{aligned} \tag{8.84}$$

Clearly, $P_l^m(x)$ is a polynomial if m is even, and $P_l^0(x)$ is an ordinary Legendre polynomial. The $P_l^m(x)$ obey the orthogonality condition

$$\int_{-1}^{1} P_l^m(x)\, P_{l'}^m(x)\, dx = \frac{2(l+m)!}{(2l+1)(l-m)!}\delta_{ll'} \qquad (m \ge 0). \tag{8.85}$$

The full solution to the Helmholtz equation is also orthogonal for different values of m, due to the form of $\Phi(\phi)$ above.

The angular solutions are generally combined, and thus the solution $\Theta(\theta)\Phi(\phi)$ is given by the **spherical harmonics**

$$Y_l^m(\theta,\phi) := \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}\, P_l^m(\cos\theta)\, e^{im\phi}, \tag{8.86}$$
(spherical harmonic)

which are more conveniently normalized such that

$$\begin{aligned} Y_l^{m*}(\theta,\phi) &= (-1)^m Y_l^{-m}(\theta,\phi) \\ \int d\Omega\, Y_l^m(\theta,\phi)\, Y_{l'}^{m'}(\theta,\phi) &= \delta_{ll'}\delta_{mm'}. \end{aligned} \tag{8.87}$$
(orthonormality relations)

(See also Section 7.3.2 for a more quantum-mechanical introduction.) They also obey the sum rule

$$\sum_{m=-l}^{l} |Y_l^m(\theta,\phi)|^2 = \frac{2l+1}{4\pi}, \tag{8.88}$$
(sum rule)

showing that the m "quantum number" determines the orientation of the modes; summing over it results in an isotropic angular distribution. Some examples of the lowest few (monopole and dipole) spherical harmonics are

$$Y_0^0(\theta,\phi) = \frac{1}{\sqrt{4\pi}}, \quad Y_1^0(\theta,\phi) = \sqrt{\frac{3}{4\pi}}\cos\theta, \quad Y_1^{\pm1}(\theta,\phi) = \mp\sqrt{\frac{3}{8\pi}}\sin\theta\, e^{\pm i\phi}, \tag{8.89}$$

The spherical harmonics form a complete set for the angular dependence of the scalar-field solutions.

As for the radial dependence, the equation for the radial function $R(r)$ becomes

$$r^2\partial_r^2 R + 2r\partial_r R + [k^2r^2 - l(l+1)]R = 0. \tag{8.90}$$

Changing variables by setting $\chi(r) := \sqrt{kr}\,R(r)$ leads to

$$r^2\partial_r^2\chi + r\partial_r\chi + \left[k^2r^2 - \left(l+\frac{1}{2}\right)^2\right]\chi = 0. \tag{8.91}$$

This is **Bessel's equation** (with independent variable kr), and the solutions are ordinary Bessel functions of the first kind, $J_{l+1/2}(kr)$, of order $l+1/2$, as well as the ordinary Bessel functions of the second kind, $Y_{l+1/2}(kr)$, of the same order. The solutions $R(r)$ are thus generally written as **spherical Bessel functions** of the first and second kind, defined by

$$\begin{aligned} j_l(r) &:= \sqrt{\frac{\pi}{2r}}\,J_{l+1/2}(r) \\ y_l(r) &:= \sqrt{\frac{\pi}{2r}}\,Y_{l+1/2}(r), \end{aligned} \tag{8.92}$$
(spherical Bessel functions)

respectively. Near the origin, these functions have the asymptotic forms[7]

$$\begin{aligned} j_l(r) &\approx \frac{r^l}{(2l+1)!!} \\ y_l(r) &\approx -\frac{(2l-1)!!}{r^{l+1}}, \end{aligned} \tag{8.93}$$

where $n!! = 1\cdot3\cdot5\cdots n$. The $y_l(r)$ thus correspond to singular modes, and we can henceforth dump them. Technically, the $y_l(r)$ are not even square-normalizable over the cavity for $l>0$, but $y_0(r)$ can be normalized, so we can't necessarily discard it based on normalizability or finite-energy arguments. However, when we go over to the vector-field case, the derivatives involved will also make it non-normalizable. In any case, our desired radial solutions are $R(r) = j_l(kr)$. They form a complete set, as follows for example from the representation[8]

$$e^{i\mathbf{k}\cdot\mathbf{r}} = \sum_{l=0}^{\infty}(2l+1)i^l j_l(kr)P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{r}}), \tag{8.94}$$

where $P_l(x) = P_l^0(x)$ is a Legendre polynomial. Since an arbitrary plane wave may be decomposed into spherical Bessel functions of the first kind, and plane waves are complete, so are the $j_1(r)$.

Of course, the radial solutions must satisfy the boundary condition $j_l(kR) = 0$, where R is now the radius of the spherical cavity. But $j_l(r)$ is an oscillatory function, and so there is a countable infinity of k values where the boundary condition is satisfied. Thus, we will define these k values by

$$j_l(k_{nl}R) = 0, \tag{8.95}$$
(transcendental equation for k_{nl})

[7] For this and other properties see Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), pp. 437-41.

[8] Eugen Merzbacher, *Quantum Mechanics*, 3rd ed. (Wiley, 1998), p. 261.

where the solution is commonly written

$$k_{nl} = \frac{a_{nl}}{R}, \tag{8.96}$$
(allowed wave numbers)

where a_{nl} is the nth positive zero of $j_l(r)$. We may thus write the scalar-field, spherical cavity modes as

$$\psi_{nlm}(\mathbf{r}) = \mathcal{N}_{nl}\, j_l(k_{nl}r)\, Y_l^m(\theta,\phi), \tag{8.97}$$
(scalar modes of spherical cavity)

where the radial normalization factor is given by

$$\mathcal{N}_{nl} := \frac{1}{\sqrt{\displaystyle\int_0^R dr\, r^2\, j_l^2(k_{nl}r)}} \tag{8.98}$$
(radial normalization factor)

The integral in the normalization factor can be performed analytically, with the result (Problem 8.15)

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\, r^2\, j_l^2(k_{nl}r) = \frac{R^3}{2}\left[j_l'(k_{nl}R)\right]^2 = \frac{R^3}{2} j_{l+1}^2(k_{nl}R),$$
(radial normalization factor) (8.99)

We see that the modes are parameterized by three indices (quantum numbers, in the case of a quantum particle in a spherical cavity), as we expect for three dimensions: a radial number n, and two angular numbers l and m.

8.4.4.2 Vector Field

Now given the solutions $\psi(\mathbf{r})$ to the scalar Helmholtz equation $(\nabla^2 + k^2)\psi = 0$, we can construct the solutions to the *vector* Helmholtz equation $(\nabla^2 + k^2)\mathbf{f} = 0$ by simply differentiating ψ.[9] In fact, three independent vector solutions are (Problem 8.14)

$$\mathbf{L} = \nabla\psi, \qquad \mathbf{M} = \nabla\times(\mathbf{r}\psi), \qquad \mathbf{N} = \frac{1}{k}\nabla\times\mathbf{M}. \tag{8.100}$$

Clearly $\mathbf{L}$ is longitudinal (as defined below), since $\nabla\times\mathbf{L} = \nabla\times\nabla\psi = 0$, whereas $\mathbf{M}$ and $\mathbf{N}$ are both transverse (as defined below), since $\nabla\cdot\mathbf{M} = \nabla\cdot\mathbf{N} = 0$. Thus, $\mathbf{M}$ and $\mathbf{N}$ are the ones that we're interested in for the cavity modes; as we will see later, only transverse fields transport energy. Alternately, starting from $\mathbf{N} = (1/k)\nabla\times\mathbf{M}$, we can compute the curl to find $\nabla\times\mathbf{N} = (1/k)\nabla\times\nabla\times\mathbf{M} = (1/k)[\nabla(\nabla\cdot\mathbf{M}) - \nabla^2\mathbf{M}] = k\mathbf{M}$, where we used $(\nabla^2 + k^2)\mathbf{M} = 0$, and thus

$$\mathbf{M} = \frac{1}{k}\nabla\times\mathbf{N}. \tag{8.101}$$

We can therefore see that $\mathbf{M}$ and $\mathbf{N}$ are proportional to each others' curl, and thus are obvious candidates to represent $\mathbf{E}$ and $\mathbf{H}$. In general, we will only use the field modes to represent transverse waves, as is consistent with the above use of plane and standing waves in the free space and the rectangular cavity, so we need not consider the $\mathbf{L}$ field. Furthermore, $\mathbf{L}$ is orthogonal to $\mathbf{M}$, since

$$\mathbf{M} = \nabla\psi\times\mathbf{r} = \mathbf{L}\times\mathbf{r}, \tag{8.102}$$

and evidently $\mathbf{M}$ is also orthogonal to $\mathbf{r}$.

Writing out the first field, using the form (8.97) for the scalar solution (though relaxing for the moment the boundary conditions, which we will apply directly to the vector solution),

$$\mathbf{M}_{nlm}(\mathbf{r}) = \mathcal{N}_{nl}\nabla\times[\mathbf{r}\, j_l(k_{nl}r)\, Y_l^m(\theta,\phi)]. \tag{8.103}$$

[9] Julius Adams Stratton, *Electromagnetic Theory* (McGraw–Hill, 1941).

Noting that

$$\begin{aligned}\nabla \times [\mathbf{r}\,\psi(\mathbf{r})] &= \nabla\psi \times \mathbf{r} + \psi\nabla \times \mathbf{r} \\ &= -\mathbf{r} \times \nabla\psi \\ &= -\mathbf{r} \times \left(\hat{r}\partial_r\psi + \hat{\theta}\frac{1}{r}\partial_\theta\psi + \hat{\phi}\frac{1}{r\sin\theta}\partial_\phi\psi\right) \\ &= -\hat{\phi}\partial_\theta\psi + \hat{\theta}\frac{1}{\sin\theta}\partial_\phi\psi.\end{aligned} \tag{8.104}$$

Also noting that the following derivative of an angular function has the same form,

$$\begin{aligned}\mathbf{r} \times \nabla g(\theta,\phi) &= \mathbf{r} \times \left(\hat{\theta}\frac{1}{r}\partial_\theta g + \hat{\phi}\frac{1}{r\sin\theta}\partial_\phi g\right) \\ &= -\hat{\phi}\partial_\theta g + \hat{\theta}\frac{1}{\sin\theta}\partial_\phi g,\end{aligned} \tag{8.105}$$

we can thus write

$$\mathbf{M}_{nlm}(\mathbf{r}) = \mathcal{N}_{nl}\, j_l(k_{nl}r)\, \mathbf{r} \times \nabla Y_l^m(\theta,\phi). \tag{8.106}$$

The angular part here is often written in normalized form as a **vector spherical harmonic**

$$\mathbf{X}_l^m(\theta,\phi) := -\frac{i}{\sqrt{l(l+1)}}\,\mathbf{r} \times \nabla Y_l^m(\theta,\phi), \tag{8.107}$$
(vector spherical harmonic)

which may also be written

$$\mathbf{X}_l^m(\theta,\phi) = -\frac{i}{\sqrt{l(l+1)}}\left(\hat{\theta}\frac{m}{\sin\theta} + i\hat{\phi}\,\partial_\theta\right) Y_l^m(\theta,\phi) \tag{8.108}$$
(vector spherical harmonic)

after writing out the gradient and cross product. These angular vector fields obey the orthonormality relations[10]

$$\begin{aligned}\int d\Omega\, \mathbf{X}_{l'}^{m'*}(\theta,\phi)\cdot\mathbf{X}_l^m(\theta,\phi) &= \delta_{ll'}\delta_{mm'} \\ \int d\Omega\, \mathbf{X}_{l'}^{m'*}(\theta,\phi)\cdot[\mathbf{r}\times\mathbf{X}_l^m(\theta,\phi)] &= 0,\end{aligned} \tag{8.109}$$
(orthonormality relations)

as well as the sum rule

$$\sum_{m=-l}^{l} |\mathbf{X}_l^m(\theta,\phi)|^2 = \frac{2l+1}{4\pi}, \tag{8.110}$$
(sum rule)

which follows from the scalar sum rule (8.88). Thus, we finally write this solution as the **transverse electric (TE)** mode

$$\mathbf{f}_{nlm}^{(\mathrm{TE})}(\mathbf{r}) \equiv \mathbf{M}_{nlm}(\mathbf{r}) = \mathcal{N}_{nl}\, j_l(k_{nl}r)\, \mathbf{X}_l^m(\theta,\phi), \tag{8.111}$$
(TE mode)

so called because the polarization vector of $\mathbf{f}(\mathbf{r})$ (the same as for $\mathbf{E}$) is parallel to the cavity surface and orthogonal to $\hat{r}$, as we can see from Eq. (8.104) or from Eq. (8.102). This is consistent with our previous notation in the half-space case of Section 8.4.3. Since the transverse component of the electric field vanishes at the cavity surface, the allowed wave numbers are identical to the scalar case,

$$j_l(k_{nl}R) = 0. \tag{8.112}$$
(allowed wave numbers, TE mode)

[10] See John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), p. 431.

and the normalization is likewise the same as for the scalar case:

$$\mathcal{N}_{nl} := \frac{1}{\sqrt{\int_0^R dr\, r^2\, j_l^{\,2}(k_{nl}r)}}. \tag{8.113}$$
(radial normalization factor)

Since the radial boundary condition is the same as for the scalar case, the integral in the normalization factor can again be performed analytically, with the result (Problem 8.15)

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\, r^2\, j_l^{\,2}(k_{nl}r) = \frac{R^3}{2}\left[j_l'(k_{nl}R)\right]^2 = \frac{R^3}{2} j_{l+1}^2(k_{nl}R),$$
(radial normalization factor, TE mode) (8.114)

However, the angular dependence is somewhat more complicated than for the scalar case, due to the vector nature of the field.

Of course, we must deal with the other solutions $\mathbf{N}(\mathbf{r})$ to the vector Helmholtz equation. To do this, we essentially just compute the curls of the TE modes

$$\mathbf{N}_{nlm}(\mathbf{r}) = \frac{1}{k_{nl}} \nabla \times \mathbf{M}_{nlm}(\mathbf{r}). \tag{8.115}$$

Using once again the vector identity $\nabla \times (\psi \mathbf{A}) = \nabla\psi \times \mathbf{A} + \psi \nabla \times \mathbf{A}$, this solution becomes

$$\begin{aligned} \mathbf{N}_{nlm}(\mathbf{r}) &= \frac{\mathcal{N}_{nl}}{k_{nl}} \nabla \times \left[j_l(k_{nl}r)\, \mathbf{X}_l^m(\theta,\phi) \right] \\ &= \frac{\mathcal{N}_{nl}}{k_{nl}} \left[\nabla j_l(k_{nl}r) \times \mathbf{X}_l^m(\theta,\phi) + j_l(k_{nl}r) \nabla \times \mathbf{X}_l^m(\theta,\phi) \right] \\ &= \mathcal{N}_{nl} \left[j_l'(k_{nl}r)\, \hat{r} \times \mathbf{X}_l^m(\theta,\phi) + \frac{1}{k_{nl}} j_l(k_{nl}r)\, \nabla \times \mathbf{X}_l^m(\theta,\phi) \right]. \end{aligned} \tag{8.116}$$

The first term is clearly transverse to the cavity surface, but the second isn't necessarily. However, if this mode represents the electric field, then the orientation of the corresponding magnetic field is of the form

$$\nabla \times \nabla \times \mathbf{X}_l^m(\theta,\phi) = \nabla[\nabla \cdot \mathbf{X}_l^m(\theta,\phi)] - \nabla^2 \mathbf{X}_l^m(\theta,\phi), \tag{8.117}$$

which has no component along $\hat{r}$. Thus, the $\mathbf{N}_{nlm}(\mathbf{r})$ modes are called the **transverse magnetic (TM) modes**,

$$\mathbf{f}_{nlm}^{(\mathrm{TM})}(\mathbf{r}) \equiv \mathbf{N}_{nlm}(\mathbf{r}) = \frac{\mathcal{N}_{nl}}{k_{nl}} \nabla \times \left[j_l(k_{nl}r)\, \mathbf{X}_l^m(\theta,\phi) \right]. \tag{8.118}$$
(TM mode)

The TM mode function must satisfy the boundary condition that the $\hat{\theta}$ and the $\hat{\phi}$ components must vanish at the surface of the cavity. Noting that in spherical coordinates,

$$\nabla \times \mathbf{A} = \hat{r}\, \frac{1}{r \sin\theta} \left[\partial_\theta (\sin\theta\, A_\phi) - \partial_\phi A_\theta \right] + \hat{\theta} \left[\frac{1}{r\sin\theta} \partial_\phi A_r - \frac{1}{r} \partial_r (r A_\phi) \right] + \hat{\phi}\, \frac{1}{r} \left[\partial_r (r A_\theta) - \partial_\theta A_r \right], \tag{8.119}$$

we see that the $\partial_r(rA_\theta)$ and $\partial_r(rA_\phi)$ terms vanish at the boundary provided

$$\partial_r \left[r\, j_l(k_{nl}r) \right]\big|_{r=R} = 0. \tag{8.120}$$
(allowed wave numbers, TM mode)

The other terms of the form $\partial_\phi A_r$ and $\partial_\theta A_r$ vanish automatically, since $\hat{r} \cdot \mathbf{X}_l^m(\theta,\phi) = 0$. The normalization

factor in the integral form (8.113) is the same as for the TE mode, as we can verify by integrating by parts:

$$\begin{aligned}\int d^3r\ \left|\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r})\right|^2 &= \frac{1}{k_{nl}^2}\int d^3r\ |\nabla\times\mathbf{M}_{nlm}(\mathbf{r})|^2 \\ &= \frac{1}{k_{nl}^2}\int d^3r\,\mathbf{M}^*_{nlm}(\mathbf{r})\cdot\nabla\times\nabla\times\mathbf{M}_{nlm}(\mathbf{r}) + \oint d\mathbf{a}\cdot\mathbf{M}^*_{nlm}(\mathbf{r})\times\nabla\times\mathbf{M}_{nlm}(\mathbf{r}) \\ &= -\frac{1}{k_{nl}^2}\int d^3r\,\mathbf{M}^*_{nlm}(\mathbf{r})\cdot\nabla^2\mathbf{M}_{nlm}(\mathbf{r}) \\ &= \int d^3r\,\mathbf{M}^*_{nlm}(\mathbf{r})\cdot\mathbf{M}_{nlm}(\mathbf{r}) \\ &= 1.\end{aligned} \tag{8.121}$$

Here, we have used $\nabla\cdot\mathbf{M}=0$. We have also discarded the surface term, since it amounts to the surface integral of $\mathbf{M}^*\times\mathbf{N}$, or in other words the Poynting vector, and for cavity modes no energy is transported across the cavity boundary. However, the analytic solution for the integral has a somewhat different form,

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\,r^2\,j_l^2(k_{nl}r) = \frac{R^3}{2}\left(1-\frac{l(l+1)}{k_{nl}^2R^2}\right)j_l^2(k_{nl}R), \tag{8.122}$$

(radial normalization factor, TM mode)

because the boundary condition here is different from the TE case (Problem 8.15).

In either the TE or the TM case, note that since the lowest scalar spherical harmonic $Y_0^0(\theta,\phi)=1/\sqrt{4\pi}$ is constant, the corresponding vector spherical harmonic vanishes, $\mathbf{X}_0^0(\theta,\phi)=0$, being the curl of the scalar version. Thus, the lowest-order vector fields have $l=0$, essentially because there are no monopolar vector waves. Note that near the center of the sphere, $j_1(r)\approx r^l/(2l+1)!!$ as we noted before, so that $j_1'(r)=lr^{l-1}/(2l+1)!!$. If an atom is at the center of the spherical cavity, evidently it only has nonvanishing coupling to the $l=1$ TM modes, since their mode functions involve $j_l'(r)$.

It is also useful to represent the vector solutions in terms of the scalar solution $\psi_{nlm}(\mathbf{r})$, Eq. (8.97). Using the expression (8.100) for the TE mode $\mathbf{M}$, as well as the expression (8.119) for the curl, we may write

$$\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r}) = \frac{1}{\sqrt{l(l+1)}}\left(\hat\theta\frac{1}{\sin\theta}\partial_\phi - \hat\phi\partial_\theta\right)\psi_{nlm}(\mathbf{r}), \tag{8.123}$$

(TE mode)

where recall that we need the extra factor of $\sqrt{l(l+1)}$ to normalize the vector angular distribution. Writing out the scalar solution,

$$\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r}) = \frac{\mathcal{N}^{(\mathrm{TE})}_{nl}}{\sqrt{l(l+1)}}j_l(k_{nl}r)\left(\hat\theta\frac{1}{\sin\theta}\partial_\phi - \hat\phi\partial_\theta\right)Y_l^m(\theta,\phi), \tag{8.124}$$

(TE mode)

which is essentially what we have already written out in terms of the vector spherical harmonic. However, for the TM case, we can obtain a relatively simple explicit expression, compared to what we could otherwise get by expanding the above expressions. To write it, we use the expression (8.100) for $\mathbf{N}$, again with the curl (8.119), to write

$$\sqrt{l(l+1)}\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r}) = -\frac{\hat r}{kr\sin\theta}\left[\partial_\theta(\sin\theta\,\partial_\theta\psi_{nlm}) + \frac{1}{\sin^2\theta}\partial_\phi^2\psi_{nlm}\right] + \frac{\hat\theta}{kr}\partial_\theta\partial_r(r\psi_{nlm}) + \frac{\hat\phi}{kr\sin\theta}\partial_\phi\partial_r(r\psi_{nlm}). \tag{8.125}$$

Now we use the fact that ψ_{nlm} satisfies Eqs. (8.79), and thus we may simplify the r component to the form

$$\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r}) = \hat r\frac{\sqrt{l(l+1)}}{kr}\psi_{nlm} + \frac{\hat\theta}{\sqrt{l(l+1)}\,kr}\partial_\theta\partial_r(r\psi_{nlm}) + \frac{\hat\phi}{\sqrt{l(l+1)}\,kr\sin\theta}\partial_\phi\partial_r(r\psi_{nlm}), \tag{8.126}$$

(TM mode)

so that we obtain the explicit form

$$\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r}) = \mathcal{N}^{(\mathrm{TM})}_{nl}\left[\hat{r}\sqrt{l(l+1)}\frac{j_l(k_{nl}r)}{kr}Y^m_l(\theta,\phi) + \frac{\partial_r[rj_l(k_{nl}r)]}{\sqrt{l(l+1)}\,kr}\left(\hat{\theta}\partial_\theta Y^m_l(\theta,\phi) + \hat{\phi}\frac{\partial_\phi Y^m_l(\theta,\phi)}{\sin\theta}\right)\right]. \quad \text{(TM mode)} \tag{8.127}$$

after writing out the scalar solution. It is once again clear here that the TM solutions have a stronger presence near the origin than do the TE modes.

8.4.4.3 Asymptotics

In general, the spherical Bessel functions make things somewhat difficult to work with. For example, we can't write down analytic expressions for the allowed wave numbers k_{nl}. However, for very large spherical cavities, such that for a given wavelength many modes can be excited, it is useful to use the asymptotic forms of the spherical Bessel functions with large arguments. From **Rayleigh's formula**

$$j_l(z) = z^l\left(-\frac{1}{z}\partial_z\right)^l\frac{\sin z}{z}, \tag{8.128}$$

we evidently have the asymptotic form

$$j_l(z) \sim (-1)^l\frac{\partial_z^l\sin z}{z} + O(z^{-2}) = \frac{1}{z}\sin\left(z - \frac{l\pi}{2}\right) + O(z^{-2}). \tag{8.129}$$

This asymptotic form also follows from the asymptotic form for the ordinary (cylindrical) Bessel function,

$$J_\alpha(z) \sim \sqrt{\frac{2}{\pi z}}\cos\left(z - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right), \tag{8.130}$$

along with the definition of $j_l(z)$ in Eq. (8.92). This form has zeroes whenever the argument of the sin is equal to $n\pi$ for integer n, and thus for the TE modes leads from Eq. (8.112) to the asymptotic condition

$$k_{nl}R = \pi n + \frac{\pi}{2}l. \quad \text{(allowed radial TE modes, } k_{nl}R \gg 1) \tag{8.131}$$

Since the first positive zeros always happen away from $z = 0$ in $j_l(z)$, the radial quantum number here obviously has $n > 0$. On the other hand, for the TM mode, the allowed modes from the condition (8.120) has a function of the form

$$\partial_z[zj_l(z)] \sim \partial_z\sin\left(z - \frac{l\pi}{2}\right) + O(z^{-1}) = -\sin\left(z - \frac{(l+1)\pi}{2}\right) + O(z^{-1}), \tag{8.132}$$

and thus gives the condition.

$$k_{nl}R = \pi n + \frac{\pi}{2}(l+1) \quad \text{(allowed radial TM modes, } k_{nl}R \gg 1) \tag{8.133}$$

The first positive zeros here occur for any $n \geq 0$, unlike the TE case.

In this asymptotic regime, we can also analytically evaluate the normalization constant, since from Eq. (8.113), the radial normalization integral becomes

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\, r^2\, j_l^2(k_{nl}r) \approx \frac{1}{k_{nl}^2}\int_0^R dr\,\sin^2\left(k_{nl}r - \frac{l\pi}{2}\right) = \frac{R}{2k_{nl}^2}, \tag{8.134}$$

and thus the normalization factor is

$$\mathcal{N}_{nl} \approx k_{nl}\sqrt{\frac{2}{R}}, \quad \text{(radial normalization factor)} \tag{8.135}$$

with the appropriate value of k_{nl} for the TE or TM modes.

8.4.4.4 Vector Multipole Modes

Now that we have the mathematical apparatus, we may as well generalize the above spherical-cavity modes a bit. In the above treatment of the spherical cavity modes, we excluded the solutions $y_l(r)$ because they were divergent at the origin. However, if we consider the *exterior* modes of a spherical cavity, or we simply exclude from consideration a neighborhood around the origin, then these divergent modes are perfectly acceptable, because the divergence is removed. When they apply, it is common to define the **spherical Hankel functions**

$$\begin{aligned} h_l^{(1)}(z) &= j_l(z) + iy_l(z) \\ h_l^{(2)}(z) &= j_l(z) - iy_l(z). \end{aligned} \tag{8.136}$$

Again from **Rayleigh's formulae**

$$\begin{aligned} j_l(z) &= z^l \left(-\frac{1}{z}\partial_z\right)^l \frac{\sin z}{z} \\ y_l(z) &= -z^l \left(-\frac{1}{z}\partial_z\right)^l \frac{\cos z}{z}, \end{aligned} \tag{8.137}$$

we can write the corresponding formulae for the spherical Hankel functions

$$\begin{aligned} h_l^{(1)}(z) &= -iz^l \left(-\frac{1}{z}\partial_z\right)^l \frac{e^{iz}}{z} \\ h_l^{(2)}(z) &= iz^l \left(-\frac{1}{z}\partial_z\right)^l \frac{e^{-iz}}{z}. \end{aligned} \tag{8.138}$$

The phase dependence clearly indicates that $h_l^{(1)}(kr)$ represents an *outgoing* wave, while $h_l^{(2)}(kr)$ represents an *ingoing* wave. The $j_l(kr)$ and $y_l(kr)$ are thus spherical-coordinate analogues to the standing waves $\sin(kx)$ and $\cos(kx)$ in Cartesian coordinates, while the $h_l^{(1)}(kr)$ and $h_l^{(2)}(kr)$ are analogues to the traveling waves $\exp(ikx)$ and $\exp(-ikx)$.

Then proceeding as above, but in free space, we may write the TE modes, but now separating them into ingoing and outgoing parts, as

$$\begin{aligned} \mathbf{f}_{klm}^{(\mathrm{TE})\rightarrow}(\mathbf{r}) &= h_l^{(1)}(kr)\,\mathbf{X}_l^m(\theta,\phi) \\ \mathbf{f}_{klm}^{(\mathrm{TE})\leftarrow}(\mathbf{r}) &= h_l^{(2)}(kr)\,\mathbf{X}_l^m(\theta,\phi), \end{aligned} \tag{8.139}$$

(TE modes)

where k is a positive, real number, while l and m are still integer indices, with l positive and m nonnegative. Of course, there may be further restrictions on k if there are boundary conditions, such as when treating the exterior modes of a spherical, conducting shell. Correspondingly, the outgoing and ingoing TM modes are

$$\begin{aligned} \mathbf{f}_{klm}^{(\mathrm{TM})\rightarrow}(\mathbf{r}) &= \frac{1}{k}\nabla \times \left[h_l^{(1)}(kr)\,\mathbf{X}_l^m(\theta,\phi)\right] \\ \mathbf{f}_{klm}^{(\mathrm{TM})\leftarrow}(\mathbf{r}) &= \frac{1}{k}\nabla \times \left[h_l^{(2)}(kr)\,\mathbf{X}_l^m(\theta,\phi)\right]. \end{aligned} \tag{8.140}$$

(TM modes)

If we work out the outgoing $l = 1, m = 0$ TE mode, we find for the radial part

$$h_1^{(1)}(r) = i\partial_r \frac{e^{ir}}{r} = -i\frac{e^{ir}}{r^2} - \frac{e^{ir}}{r} = -\frac{e^{ir}}{r}\left(1 + \frac{i}{r}\right), \tag{8.141}$$

and the angular part,

$$\begin{aligned} \mathbf{X}_1^0(\theta,\phi) &= -\frac{i}{\sqrt{2}}\mathbf{r} \times \nabla\sqrt{\frac{3}{4\pi}}\cos\theta \\ &= -i\sqrt{\frac{3}{8\pi}}\mathbf{r} \times \nabla\cos\theta \\ &= \hat{\phi}\, i\sqrt{\frac{3}{8\pi}}\sin\theta, \end{aligned} \tag{8.142}$$

so that the complete mode function is

$$\mathbf{f}_{k10}^{(\mathrm{TE})\rightarrow}(\mathbf{r}) = \hat{\phi}\sqrt{\frac{3}{8\pi}}\frac{e^{ikr}}{k^2}\left(\frac{1}{r^2} - i\frac{k}{r}\right)\sin\theta. \tag{8.143}$$

Similarly, computing the outgoing $l = 1, m = 0$ TM mode, using Eq. (8.119) for the curl, we need the derivatives

$$\begin{aligned}\frac{1}{r\sin\theta}\partial_\theta \sin\theta i\sqrt{\frac{3}{8\pi}}\sin\theta &= i\sqrt{\frac{3}{2\pi}}\frac{\cos\theta}{r} \\ -\frac{1}{r}\partial_r r\left[-\frac{e^{ikr}}{kr}\left(1+\frac{i}{kr}\right)\right] &= -i\frac{e^{ikr}}{k^2}\left(\frac{1}{r^3} - i\frac{k}{r^2} - \frac{k^2}{r}\right),\end{aligned} \tag{8.144}$$

so that we find the mode function

$$\mathbf{f}_{k10}^{(\mathrm{TM})\rightarrow}(\mathbf{r}) = \sqrt{\frac{3}{8\pi}}\frac{e^{ikr}}{k^2}\left[\hat{r}\,2\left(\frac{1}{r^3} - i\frac{k}{r^2}\right)\cos\theta + \hat{\theta}\left(\frac{1}{r^3} - i\frac{k}{r^2} - \frac{k^2}{r}\right)\sin\theta\right]. \tag{8.145}$$

Recalling the electric-dipole fields from Eq. (1.42),

$$\begin{aligned}\mathbf{E}^{(+)}(\mathbf{r},t) &= \frac{1}{4\pi\epsilon_0}\left[3(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}\right]\left[\frac{d^{(+)}(t_\mathrm{r})}{r^3} + \frac{\dot{d}^{(+)}(t_\mathrm{r})}{cr^2}\right] + \frac{1}{4\pi\epsilon_0}\left[(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}\right]\frac{\ddot{d}^{(+)}(t_\mathrm{r})}{c^2 r} \\ \mathbf{H}^{(+)}(\mathbf{r},t) &= \frac{c}{4\pi}(\hat{\varepsilon}\times\hat{r})\left[\frac{\dot{d}^{(+)}(t_\mathrm{r})}{cr^2} + \frac{\ddot{d}^{(+)}(t_\mathrm{r})}{c^2 r}\right],\end{aligned} \tag{8.146}$$

which for monochromatic fields become,

$$\begin{aligned}\mathbf{E}^{(+)}(\mathbf{r},t) &= \frac{d^{(+)}(\omega)}{4\pi\epsilon_0}e^{ikr}\left\{\left[3(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}\right]\left(\frac{1}{r^3} - i\frac{k}{r^2}\right) - \left[(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}\right]\frac{k^2}{r}\right\} \\ \mathbf{H}^{(+)}(\mathbf{r},t) &= -i\frac{cd^{(+)}(\omega)}{4\pi}e^{ikr}(\hat{\varepsilon}\times\hat{r})\left(\frac{k}{r^2} - i\frac{k^2}{r}\right).\end{aligned} \tag{8.147}$$

Comparing to the above mode functions for $\hat{\varepsilon} = \hat{z}$, we see the dipole fields may be written as

$$\begin{aligned}\mathbf{E}^{(+)}(\mathbf{r},t) &= \frac{1}{\sqrt{6\pi}}\frac{\omega^2 d^{(+)}(\omega)}{\epsilon_0 c^2}\,\mathbf{f}_{k10}^{(\mathrm{TM})\rightarrow}(\mathbf{r}) \\ \mathbf{H}^{(+)}(\mathbf{r},t) &= -i\frac{\omega^2 d^{(+)}(\omega)}{c\sqrt{6\pi}}\,\mathbf{f}_{k10}^{(\mathrm{TE})\rightarrow}(\mathbf{r}).\end{aligned} \tag{8.148}$$

Similarly, $\mathbf{f}_{k10}^{(\mathrm{TE})\rightarrow}(\mathbf{r})$ represents the dimensionless *electric*-field mode profile due to a *magnetic* dipole, and $\mathbf{f}_{k10}^{(\mathrm{TM})\rightarrow}(\mathbf{r})$ represents the dimensionless *magnetic*-field mode profile due to a magnetic dipole.

In general, the $\mathbf{f}_{klm}^{(\mathrm{TM})\rightarrow}(\mathbf{r})$ represent the dimensionless electric-field mode profiles due to electric multipoles of order l, and the $\mathbf{f}_{klm}^{(\mathrm{TE})\rightarrow}(\mathbf{r})$ represent the dimensionless magnetic-field mode profiles due to electric multipoles of order l. For magnetic multipoles, the profiles exchange identity. From the form of $h_l^{(1)}(kr)$ in Eq. (8.138), we can see that in the far field, all the multipole fields decay as r^{-1}, corresponding to radiation. Similarly, in the near field, for example, the electric field due to an electric multipole of order l goes as r^{l+2}, while the magnetic field goes as r^{l+1}.

8.5 Transverse and Longitudinal Fields

8.5.1 Helmholtz Theorem

Before continuing, it is convenient to distinguish between *transverse* and *longitudinal* part of a vector field[11]. For an arbitrary field $\mathbf{C}(\mathbf{r})$, the **Helmholtz theorem** states that there is a *unique* decomposition

$$\mathbf{C}(\mathbf{r}) = \mathbf{C}^{\perp}(\mathbf{r}) + \mathbf{C}^{\parallel}(\mathbf{r}) \tag{8.149}$$
(Helmholtz theorem)

such that the **transverse field** is divergenceless,

$$\nabla \cdot \mathbf{C}^{\perp}(\mathbf{r}) = 0, \tag{8.150}$$
(transverse field condition)

and the **longitudinal field** is irrotational,

$$\nabla \times \mathbf{C}^{\parallel}(\mathbf{r}) = 0. \tag{8.151}$$
(longitudinal field condition)

We can see this by starting with the delta-function identity

$$\mathbf{C}(\mathbf{r}) = \int d^3r' \mathbf{C}(\mathbf{r}')\delta^3(\mathbf{r} - \mathbf{r}'). \tag{8.152}$$

Thus using the delta-function identity

$$\nabla^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -4\pi\delta^3(\mathbf{r} - \mathbf{r}') \tag{8.153}$$

(which is essentially Poisson's equation $\nabla^2\phi = -\rho/\epsilon_0$ for a point charge) with the vector identity

$$\nabla^2\mathbf{C} = \nabla(\nabla \cdot \mathbf{C}) - \nabla \times (\nabla \times \mathbf{C}), \tag{8.154}$$

we can write

$$\begin{aligned}
\mathbf{C}(\mathbf{r}) &= -\frac{1}{4\pi}\int d^3r' \mathbf{C}(\mathbf{r}')\nabla^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} \\
&= \frac{1}{4\pi}\nabla \times \nabla \times \int d^3r' \frac{\mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{4\pi}\nabla \int d^3r' \nabla \cdot \frac{\mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\
&= \frac{1}{4\pi}\nabla \times \nabla \times \int d^3r' \frac{\mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{4\pi}\nabla \int d^3r' \, \mathbf{C}(\mathbf{r}') \cdot \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} \\
&= \frac{1}{4\pi}\nabla \times \nabla \times \int d^3r' \frac{\mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} - \frac{1}{4\pi}\nabla \int d^3r' \, \frac{\nabla' \cdot \mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\
&= \mathbf{C}^{\perp}(\mathbf{r}) + \mathbf{C}^{\parallel}(\mathbf{r}),
\end{aligned} \tag{8.155}$$

where we have assumed that the boundary terms vanish, and we have defined

$$\begin{aligned}
\mathbf{C}^{\perp}(\mathbf{r}) &:= \frac{1}{4\pi}\nabla \times \nabla \times \int d^3r' \frac{\mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \\
\mathbf{C}^{\parallel}(\mathbf{r}) &:= -\frac{1}{4\pi}\nabla \int d^3r' \, \frac{\nabla' \cdot \mathbf{C}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}
\end{aligned} \tag{8.156}$$
(transverse and longitudinal fields)

[11] For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Appendix F, p. 501.

From their forms, the two components clearly have the desired properties.

Now for uniqueness of the decomposition. Once the divergence and curl of a vector field are specified, along with its boundary conditions, the vector field itself is uniquely specified. That is, suppose that

$$\nabla \cdot \mathbf{C}_1(\mathbf{r}) = \nabla \cdot \mathbf{C}_2(\mathbf{r}) \tag{8.157}$$

$$\nabla \times \mathbf{C}_1(\mathbf{r}) = \nabla \times \mathbf{C}_2(\mathbf{r}). \tag{8.158}$$

Then the *difference* field satisfies

$$\nabla \cdot (\mathbf{C}_2 - \mathbf{C}_1) = 0, \qquad \nabla \times (\mathbf{C}_2 - \mathbf{C}_1) = 0. \tag{8.159}$$

Because the curl of the difference vanishes, we can write it as the gradient of a scalar function,

$$\mathbf{C}_2 - \mathbf{C}_1 = \nabla h(\mathbf{r}), \tag{8.160}$$

which, with the fact that the divergence vanishes, implies Laplace's equation for h:

$$\nabla^2 h(\mathbf{r}) = 0. \tag{8.161}$$

Further, if $\mathbf{C}_1 = \mathbf{C}_2$ on the boundary (in our argument above, this was at arbitrarily large distances), then $h(\mathbf{r})$ vanishes on the boundary. But solutions to Laplace's equation have no maxima or minima, so $h(\mathbf{r}) = 0$. To establish this formally, consider the divergence theorem

$$\int_S \mathbf{K} \cdot d\mathbf{a} = \int_V d^3r\, \nabla \cdot \mathbf{K}. \tag{8.162}$$

Then letting $\mathbf{K} = h\nabla h$, the surface integral becomes

$$\int_S \mathbf{K} \cdot d\mathbf{a} = \int_S h(\nabla h) \cdot d\mathbf{a} = \int_S h(\mathbf{C}_2 - \mathbf{C}_1) \cdot d\mathbf{a} = 0, \tag{8.163}$$

if we assume that the *normal* component of $\mathbf{C}_1$ and $\mathbf{C}_2$ are equal on the boundary. (We need not specify the equality of the *transverse* components.) This implies that the volume integral vanishes,

$$\int_V d^3r\, \nabla \cdot \mathbf{K} = 0, \tag{8.164}$$

so that we can use

$$\nabla \cdot (h\nabla h) = h\nabla \cdot \nabla h + (\nabla h)^2 \tag{8.165}$$

to write

$$\begin{aligned} \int_V d^3r\, (\nabla h)^2 &= -\int_V d^3r\, h\nabla \cdot \nabla h \\ &= -\int_V d^3r\, h\nabla \cdot (\mathbf{C}_2 - \mathbf{C}_1) \\ &= 0, \end{aligned} \tag{8.166}$$

since the divergences of the fields are equal. Finally, we have

$$\int_V d^3r\, (\mathbf{C}_2 - \mathbf{C}_1)^2 = 0, \tag{8.167}$$

and since $(\mathbf{C}_2 - \mathbf{C}_1)^2 \geq 0$, the only way to satisfy this constraint is for $\mathbf{C}_1 = \mathbf{C}_2$. Thus, the Helmholtz decomposition (8.156) is unique.

8.5.1.1 Coulomb Gauge

In the Coulomb gauge, the transverse and longitudinal components of the fields are easy to identify. The vector field is completely transverse, since $\nabla \cdot \mathbf{A} = 0$. The magnetic field is similarly transverse, as fundamentally $\nabla \cdot \mathbf{B} = 0$. The electric field has both components; the source-free part due to $\mathbf{A}$ is $\mathbf{E}^\perp = -\partial_t \mathbf{A}$, which is clearly transverse. The part due to a source charge is $\mathbf{E}^\parallel = -\nabla\phi$, which is clearly longitudinal.

8.5.2 Transverse and Longitudinal Delta Functions

Just as we can use the Kronecker delta and the delta function as a projection operator for a component of the field,

$$C_\alpha(\mathbf{r}) = \int d^3r' \delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}')\, C_\beta(\mathbf{r}') \tag{8.168}$$

(summation is implied here by repeated indices), we can also define projection operators for the transverse and longitudinal parts of the field. The **transverse delta function**, defined by

$$C^\perp_\alpha(\mathbf{r}) = \int d^3r' \delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}')\, C_\beta(\mathbf{r}'),$$

(defining relation, transverse delta function) (8.169)

projects out the transverse part of the field, while the **longitudinal delta function**, defined by

$$C^\parallel_\alpha(\mathbf{r}) = \int d^3r' \delta^\parallel_{\alpha\beta}(\mathbf{r}-\mathbf{r}')\, C_\beta(\mathbf{r}'),$$

(defining relation, longitudinal delta function) (8.170)

projects out the longitudinal part of the field.

8.5.2.1 Momentum Representation

We will use the following normalization convention for the Fourier transform and inverse transform of the field:

$$\begin{aligned}\tilde{\mathbf{C}}(\mathbf{k}) &= \frac{1}{(2\pi)^{3/2}}\int d^3r\, \mathbf{C}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \\ \mathbf{C}(\mathbf{r}) &= \frac{1}{(2\pi)^{3/2}}\int d^3k\, \tilde{\mathbf{C}}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}}.\end{aligned} \tag{8.171}$$

The vector identity

$$\begin{aligned}\tilde{\mathbf{C}}(\mathbf{k}) &= -\frac{1}{k^2}\mathbf{k}\times[\mathbf{k}\times\tilde{\mathbf{C}}(\mathbf{k})] + \frac{1}{k^2}\mathbf{k}[\mathbf{k}\cdot\tilde{\mathbf{C}}(\mathbf{k})] \\ &= \tilde{\mathbf{C}}^\perp(\mathbf{k}) + \tilde{\mathbf{C}}^\parallel(\mathbf{k})\end{aligned} \tag{8.172}$$

is the $\mathbf{k}$-space version of Eq. (8.155). The transverse and longitudinal components are easy to identify here from the conditions

$$\mathbf{k}\cdot\tilde{\mathbf{C}}^\perp(\mathbf{k}) = 0, \qquad \mathbf{k}\times\tilde{\mathbf{C}}^\parallel(\mathbf{k}) = 0, \tag{8.173}$$

which are the Fourier-space versions of Eqs. (8.150) and (8.151). Then we can compute the inverse transform of the longitudinal part:

$$\begin{aligned}\mathbf{C}^\parallel(\mathbf{r}) &= \frac{1}{(2\pi)^{3/2}}\int d^3k\, \frac{1}{k^2}\mathbf{k}[\mathbf{k}\cdot\tilde{\mathbf{C}}(\mathbf{k})] e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \frac{1}{(2\pi)^3}\int d^3r' \int d^3k\, \frac{1}{k^2}\mathbf{k}[\mathbf{k}\cdot\mathbf{C}(\mathbf{r}')] e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}.\end{aligned} \tag{8.174}$$

We can write this relation in components to find

$$C^\parallel_\alpha(\mathbf{r}) = \int d^3r' \frac{1}{(2\pi)^3}\int d^3k\, \frac{1}{k^2} k_\alpha k_\beta C_\beta(\mathbf{r}') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}. \tag{8.175}$$

Comparing this to Eq. (8.170), we can write the longitudinal delta function as

$$\delta^{\parallel}_{\alpha\beta}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\,\frac{k_\alpha k_\beta}{k^2}e^{i\mathbf{k}\cdot\mathbf{r}}.$$

(longitudinal delta function, momentum representation) (8.176)

Eq. (8.149) implies that

$$\delta_{\alpha\beta}\delta^3(\mathbf{r}) = \delta^{\perp}_{\alpha\beta}(\mathbf{r}) + \delta^{\parallel}_{\alpha\beta}(\mathbf{r}), \tag{8.177}$$

so the transverse delta function simply becomes

$$\delta^{\perp}_{\alpha\beta}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}},$$

(transverse delta function, momentum representation) (8.178)

where we used the representation

$$\delta^3(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\,e^{i\mathbf{k}\cdot\mathbf{r}} \tag{8.179}$$

of the delta function in three dimensions.

8.5.2.2 Position Representation

We can also evaluate the integrals above to obtain direct expressions for the transverse and longitudinal delta functions. Starting with Eq. (8.178), we note that the integral is not strictly convergent, so we will insert a convergence factor $e^{-k\lambda}$, letting $\lambda \longrightarrow 0$ after the integration. This procedure is effectively equivalent to assuming that the fields on which the projectors operate are reasonably smooth (i.e., bandwidth-limited). Then

$$\begin{aligned}
\delta^{\perp}_{\alpha\beta}(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}}e^{-k\lambda} \\
&= \frac{1}{(2\pi)^3}\left(-\delta_{\alpha\beta}\nabla^2 + \partial_\alpha\partial_\beta\right)\int d^3k\,\frac{e^{i\mathbf{k}\cdot\mathbf{r}-k\lambda}}{k^2} \\
&= \frac{1}{(2\pi)^2}\left(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\right)\int_0^\infty dk\int_0^\pi d\theta\,\sin\theta\,e^{ikr\cos\theta-k\lambda} \\
&= \frac{1}{(2\pi)^2}\left(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\right)\int_0^\infty dk\,\frac{2\sin kr}{kr}e^{-k\lambda} \\
&= \frac{1}{2\pi^2}\left(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\right)\frac{\tan^{-1}(r/\lambda)}{r}.
\end{aligned} \tag{8.180}$$

Letting $\lambda \longrightarrow 0$, we find

$$\delta^{\perp}_{\alpha\beta}(\mathbf{r}) = \frac{1}{4\pi}\left(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\right)\frac{1}{r}. \tag{8.181}$$

We can then use $\nabla^2(1/r) = -4\pi\delta^3(\mathbf{r})$ and the relation

$$\partial_\alpha\partial_\beta\left(\frac{1}{r}\right) = -\partial_\alpha\frac{r_\beta}{r^3} = -\frac{4\pi}{3}\delta_{\alpha\beta}\delta^3(\mathbf{r}) - \frac{1}{r^3}\delta_{\alpha\beta} + \frac{3r_\alpha r_\beta}{r^5}, \tag{8.182}$$

where the delta function arises since for $\alpha = \beta$ we effectively have $1/3$ of $\nabla^2(1/r)$ (the other terms arise from straightforward differentiation). The result is

$$\delta^{\perp}_{\alpha\beta}(\mathbf{r}) = \frac{2}{3}\delta_{\alpha\beta}\delta^3(\mathbf{r}) - \frac{1}{4\pi r^3}\left(\delta_{\alpha\beta} - \frac{3r_\alpha r_\beta}{r^2}\right).$$

(transverse delta function) (8.183)

Using Eq. (8.177), we can also write the corresponding expression

$$\delta^{\parallel}_{\alpha\beta}(\mathbf{r}) = \frac{1}{3}\delta_{\alpha\beta}\delta^3(\mathbf{r}) + \frac{1}{4\pi r^3}\left(\delta_{\alpha\beta} - \frac{3r_\alpha r_\beta}{r^2}\right)$$

(longitudinal delta function) (8.184)

for the longitudinal delta function.

8.6 Field Commutators

In view of the bosonic commutator (8.60), the field operators will not in general commute.[12] As these commutation relations are useful, we will spend some time exploring these. We start with the relations (8.56) for the quantum fields. We can then write, for example,

$$\begin{aligned}[A_\alpha(\mathbf{r},t), A_\beta(\mathbf{r}',t')] &= \sum_{\mathbf{k},\zeta} \frac{\hbar}{2\omega_\mathbf{k}\epsilon_0} f_{\mathbf{k},\zeta,\alpha}(\mathbf{r}) f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}') e^{-i\omega_\mathbf{k}(t-t')} - \text{c.c.} \\ &= \frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\frac{1}{\omega_\mathbf{k}}\text{Im}\left[f_{\mathbf{k},\zeta,\alpha}(\mathbf{r}) f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}') e^{-i\omega_\mathbf{k}(t-t')}\right].\end{aligned}$$

(vector-potential commutator) (8.185)

In this form, the commutator is not easy to interpret, but it is clear that the commutator depends on the boundary conditions that determine the mode functions $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$.

Other useful relations include

$$\begin{aligned}[E_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] &= \frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\omega_\mathbf{k}\,\text{Im}\left[f_{\mathbf{k},\zeta,\alpha}(\mathbf{r}) f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}') e^{-i\omega_\mathbf{k}(t-t')}\right] \\ [B_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t')] &= -\frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\frac{1}{\omega_\mathbf{k}}\varepsilon_{\alpha\mu\nu}\varepsilon_{\beta\sigma\tau}\,\text{Im}\left[\left(\partial_\mu f_{\mathbf{k},\zeta,\nu}(\mathbf{r})\right)\left(\partial'_\sigma f^*_{\mathbf{k},\zeta,\tau}(\mathbf{r}')\right) e^{-i\omega_\mathbf{k}(t-t')}\right] \\ [A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] &= -\frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\text{Re}\left[f_{\mathbf{k},\zeta,\alpha}(\mathbf{r}) f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}') e^{-i\omega_\mathbf{k}(t-t')}\right] \\ [E_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t')] &= -\frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\text{Re}\left[f_{\mathbf{k},\zeta,\alpha}(\mathbf{r})\varepsilon_{\beta\mu\nu}\partial'_\mu f^*_{\mathbf{k},\zeta,\nu}(\mathbf{r}') e^{-i\omega_\mathbf{k}(t-t')}\right].\end{aligned}$$

(various field commutators) (8.186)

Again, the interpretation here is not transparent, so we will consider their specific form in free space.

8.6.1 Free-Space Commutators

We will now use the free-space mode functions $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = V^{-1/2}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}$ of Section 8.4.2, to write out the commutators in the free-space case. Recall that they satisfy $\mathbf{k}\cdot\hat{\varepsilon}_{\mathbf{k},\zeta} = 0$ (because $\nabla\cdot\mathbf{E} = 0$).

8.6.1.1 Direction Cosines

To evaluate the summations over the polarization index ζ in the above commutators, we will need to compute the sum

$$\sum_\zeta \left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha\right)\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta\right). \tag{8.187}$$

[12] For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Section 2.8, p. 59; Leonard Mandel and Emil Wolf, *Optical Coherence and Quantum Optics* (Cambridge, 1995), Section 10.8, p. 500; and P. W. Milonni, "Casimir forces without the vacuum radiation field," *Physical Review A*, **25**, 1315 (1982) (doi: 10.1103/PhysRevA.25.1315).

To evaluate this sum, recall the **direction cosines** for the vector $\mathbf{r}$. This vector makes angles θ_α with the respective r_α-axes. The direction cosines are defined as $\gamma_\alpha = \cos\theta_\alpha = \mathbf{r}\cdot\hat{r}_\alpha/r$, and thus they satisfy

$$\sum_\alpha \gamma_\alpha^2 = \sum_\alpha \frac{r_\alpha^2}{r^2} = 1. \tag{8.188}$$

More generally, if we have two orthonormal cartesian bases $\hat{r}_\alpha$ and $\hat{r}'_\alpha$, then we can define direction cosines between the coordinate systems of $\gamma_{\alpha\beta} := \hat{r}_\alpha\cdot\hat{r}'_\beta$. Then it follows from the orthogonality of the basis vectors that

$$\sum_\mu \gamma_{\alpha\mu}\gamma_{\beta\mu} = \sum_\mu \left(\hat{r}_\alpha\cdot\hat{r}'_\mu\right)\left(\hat{r}_\beta\cdot\hat{r}'_\mu\right) = \hat{r}_\alpha\cdot\hat{r}_\beta = \delta_{\alpha\beta}, \tag{8.189}$$

which for the case $\alpha=\beta$ is equivalent to Eq. (8.188). Noting that $\hat{\varepsilon}_{\mathbf{k},\zeta}$ and $\hat{k}$ form one orthonormal basis, with $\hat{r}_\alpha$ forming another, we can apply the result (8.189) to obtain

$$\sum_\zeta \left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha\right)\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta\right) = \delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}, \tag{8.190}$$

which will prove to be a very useful relation in mode-summation problems.

8.6.1.2 Evaluation

From Eq. (8.185), we can write

$$\begin{aligned}[A_\alpha(\mathbf{r},t), A_\beta(\mathbf{r}',t')] &= \frac{i\hbar}{c\epsilon_0 V}\sum_{\mathbf{k},\zeta}\frac{1}{k}\mathrm{Im}\left[e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_\mathbf{k}(t-t')}\right]\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha\right)\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta\right)\\ &= -\frac{i\hbar}{c\epsilon_0 V}\sum_{\mathbf{k}}\frac{1}{k}e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\sin[\omega_\mathbf{k}(t-t')]\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right),\end{aligned} \tag{8.191}$$

where we used the fact that $\int d^3k\exp(i\mathbf{k}\cdot\mathbf{r})$ is real. In free space, we take the limit of large quantization volume ($V\longrightarrow\infty$), and the spacing between the modes becomes correspondingly small. In this limit, an integral of a function is equivalent to a sum weighted by the mode spacings. Thus we can write

$$\sum_\mathbf{k} f(\mathbf{k})\,\Delta k_x\,\Delta k_y\,\Delta k_z \longrightarrow \int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\int_{-\infty}^{\infty}dk_z\,f(\mathbf{k}) \tag{8.192}$$

for an arbitrary function $f(\mathbf{k})$. Since

$$\Delta k_{x,y,z} = \frac{2\pi}{V^{1/3}}, \tag{8.193}$$

we can thus make the formal replacement

$$\sum_\mathbf{k} \longrightarrow \frac{V}{(2\pi)^3}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\int_{-\infty}^{\infty}dk_z. \tag{8.194}$$

Thus, we write

$$[A_\alpha(\mathbf{r},t), A_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{c\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\,\frac{1}{k}e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\sin[\omega_\mathbf{k}(t-t')]\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right)$$

(vector-potential commutator, free space) (8.195)

for the vector-potential commutator.

Similarly, for the electric field we can write

$$[E_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] = -\frac{i\hbar c}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\,k e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\sin[\omega_\mathbf{k}(t-t')]\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right). \tag{8.196}$$

We can simplify this commutator somewhat by introducing the singular D function:

$$\begin{aligned} D(\mathbf{r},t) &:= -\frac{1}{(2\pi)^3}\int d^3k\,\frac{1}{k}e^{i\mathbf{k}\cdot\mathbf{r}}\sin\omega_{\mathbf{k}}t \\ &= -\frac{1}{(2\pi)^2}\int_0^\infty dk\int_0^\pi d\theta\sin\theta\,ke^{ikr\cos\theta}\sin\omega_{\mathbf{k}}t \\ &= -\frac{1}{(2\pi)^2}\int_0^\infty dk\,\frac{2\sin kr}{r}\sin\omega_{\mathbf{k}}t \\ &= \frac{1}{8\pi^2 r}\int_0^\infty dk\,\left(e^{ikr}-e^{-ikr}\right)\left(e^{ickt}-e^{-ickt}\right) \\ &= \frac{1}{8\pi^2 r}\int_{-\infty}^\infty dk\,\left(e^{ik(r+ct)}-e^{ik(r-ct)}\right) \\ &= \frac{1}{4\pi r}\left[\delta(r+ct)-\delta(r-ct)\right]. \end{aligned} \tag{8.197}$$

Notice that $D(\mathbf{r},t)$ vanishes away from the light cone $r=\pm ct$. Then the commutator (8.196) becomes

$$[E_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t')] = \frac{i\hbar c}{\epsilon_0}\left[\frac{\delta_{\alpha\beta}}{c^2}\partial_t\partial_{t'}-\partial_\alpha\partial'_\beta\right]D(\mathbf{r}-\mathbf{r}',t-t'),$$

(electric-field commutator, free space) (8.198)

where $\partial'_\alpha := \partial/\partial r'_\alpha$. This and the two field commutators that follow are the **Jordan–Pauli commutators**.[13] The interpretation of this commutator is the electric field can be measured at two spacetime points, so long as they are not on the same light cone.[14] More intuitively, making an electric-field measurement at $(\mathbf{r},t)$ disturbs the field, and in vacuum the disturbance propagates occupies any spacetime point $(\mathbf{r}',t')$ on the "future" light cone $|\mathbf{r}'-\mathbf{r}|=c(t'-t)$, causing measurement problems at any of those points. Similarly, any measurement event on the "past" light cone locus of points $(\mathbf{r}',t')$ satisfying $|\mathbf{r}'-\mathbf{r}|=c(t-t')$ causes a disturbance that propagates to $(\mathbf{r},t)$, disturbing a measurement there.

A similar calculation shows that the magnetic field has a commutator of almost the same form,

$$[B_\alpha(\mathbf{r},t),B_\beta(\mathbf{r}',t')] = \frac{i\hbar}{\epsilon_0 c}\left[\frac{\delta_{\alpha\beta}}{c^2}\partial_t\partial_{t'}-\partial_\alpha\partial_\beta\right]D(\mathbf{r}-\mathbf{r}',t-t'),$$

(magnetic-field commutator, free space) (8.199)

and so the same interpretation applies to the magnetic field. The mixed commutator for the electric and magnetic fields is slightly different. Using the same procedure as above, we can write

$$\begin{aligned} [E_\alpha(\mathbf{r},t),B_\beta(\mathbf{r}',t')] &= -\frac{i\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\sum_\zeta\int d^3k\,\mathrm{Re}\left[\varepsilon_{\beta\mu\nu}\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha\right)\left(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\nu\right)(ik_\mu)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_{\mathbf{k}}(t-t')}\right] \\ &= \frac{\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\,\mathrm{Re}\left[\varepsilon_{\beta\mu\nu}\left(\delta_{\alpha\nu}-\frac{k_\alpha k_\nu}{k^2}\right)k_\mu e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_{\mathbf{k}}(t-t')}\right] \\ &= \frac{\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\,\mathrm{Re}\left[\varepsilon_{\alpha\beta\mu}k_\mu e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_{\mathbf{k}}(t-t')}\right] \\ &= \frac{\hbar}{\epsilon_0}\varepsilon_{\alpha\beta\mu}\frac{1}{(2\pi)^3}\int d^3k\,k_\mu e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\cos\left[\omega_{\mathbf{k}}(t-t')\right], \end{aligned} \tag{8.200}$$

[13] P. Jordan and W. Pauli, Jr., "Zur Quantenelektrodynamik ladungsfreier Felder," *Zeitschrift für Physik* **47**, 151 (1928) (doi: 10.1007/BF02055793) (doi: 10.1007/BF02055793).

[14] The seminal discussion of simultaneous measurability of the quantum electric and magnetic fields can be found in Niels Bohr and Léon Rosenfeld, "Zur Frage der Messbarkeit der elektromagnetischen Feldgrössen," *Mathematisk-Fysiske Meddelelser* **12** (1933); translation reprinted in *Quantum Theory and Measurement*, John Archibld Wheeler and Wojciech Hubert Zurek, Eds. (Princeton, 1983) p. 479.

and thus

$$[E_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{\epsilon_0}\varepsilon_{\alpha\beta\mu}\partial_t\partial_\mu D(\mathbf{r}-\mathbf{r}',t-t').$$

(electric-magnetic-field commutator, free space) (8.201)

(Note that this commutator vanishes when $\alpha = \beta$.) Again, the same measurability comment applies here. Finally, we have the commutator between the conjugate fields $\mathbf{A}$ and $\mathbf{E}$,

$$[A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\sum_\zeta \int d^3k\, \mathrm{Re}\left[(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)\,(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta)\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_\mathbf{k}(t-t')}\right], \tag{8.202}$$

so that

$$[A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\, \mathrm{Re}\left[e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}e^{-i\omega_\mathbf{k}(t-t')}\right]\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right).$$

(potential–field commutator) (8.203)

This relation does not simplify as the ones for the $\mathbf{E}$ and $\mathbf{B}$ fields, a point that we will return to below.

8.6.1.3 Equal-Time Commutators in Free Space

From Eq. (8.196), we can see that at equal times,

$$[E_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t)] = 0. \tag{8.204}$$

(equal times)

Similarly, we can see that

$$\begin{aligned}[B_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t)] &= 0\\ [A_\alpha(\mathbf{r},t), A_\beta(\mathbf{r}',t)] &= 0.\end{aligned} \tag{8.205}$$

(equal times)

Thus, the fields can be measured at two different locations at the same time. However, for the mixed commutator for the electric and magnetic fields,

$$\begin{aligned}[E_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t)] &= \frac{\hbar}{\epsilon_0}\varepsilon_{\alpha\beta\mu}\frac{1}{(2\pi)^3}\int d^3k\, k_\mu e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\\ &= -\frac{i\hbar}{\epsilon_0}\varepsilon_{\alpha\beta\mu}\partial_\mu\delta^3(\mathbf{r}-\mathbf{r}').\end{aligned} \tag{8.206}$$

(equal times)

The electric and magnetic fields thus cannot be measured at the same location at equal times. For the electric field and vector potential, we arrive at the important commutation relation

$$\begin{aligned}[A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t)] &= -\frac{i\hbar}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right)\\ &= -\frac{i\hbar}{\epsilon_0}\delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}'),\end{aligned} \tag{8.207}$$

(equal times)

which is nonzero even off of the light cone. The vector potential is evidently not a local field in the same sense as the electric and magnetic fields (we will see this again in the Aharonov–Bohm effect in Section 9.4). Finally, it turns out that

$$[A_\alpha(\mathbf{r},t), B_\beta(\mathbf{r}',t)] = 0, \tag{8.208}$$

(equal times)

so that the potential is more ‘‘compatible’’ with the magnetic field.

8.6.2 Half-Space Commutators

As we noted above, the field commutators depend on the boundary conditions through the form of the mode profile functions $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$. We can see this by considering the half-space mode functions of Section 8.4.3. Using the mode functions in the form of Eq. (8.73), and putting them into Eqs. (8.186), the mixed commutator for the vector potential and electric field becomes

$$\begin{aligned}[A_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{2\epsilon_0 V}\sum_{\mathbf{k},\zeta}\mathrm{Re}\Big[&(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta)\,e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + (\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\beta)\,e^{i\mathbf{k}^-\cdot(\mathbf{r}-\mathbf{r}')}\\ &-(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\beta)\,e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}'^-)} - (\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta)\,e^{i\mathbf{k}\cdot(\mathbf{r}^--\mathbf{r}')}\Big]\\ &\times e^{-i\omega_{\mathbf{k}}(t-t')}.\end{aligned} \tag{8.209}$$

In the expression here $\mathbf{r}^- := x\hat{x}+y\hat{y}-z\hat{z}$, and $\mathbf{r}^--\mathbf{r}' = \mathbf{r}-\mathbf{r}'-2z\hat{z}$. Carrying out the sum over polarizations as usual, using the result (8.190), while noting the appropriate sign changes when the reflected polarization vectors are involved,

$$\sum_\zeta(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\beta) = \delta^-_{\alpha\beta} - \frac{k_\alpha k^-_\beta}{k^2}, \tag{8.210}$$

we can write

$$\begin{aligned}[A_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{2\epsilon_0 V}\sum_{\mathbf{k}}\mathrm{Re}\Big[&\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + \left(\delta_{\alpha\beta}-\frac{k^-_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}^-\cdot(\mathbf{r}-\mathbf{r}')}\\ &-\left(\delta^-_{\alpha\beta}-\frac{k_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}'^-)} - \left(\delta^-_{\alpha\beta}-\frac{k^-_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}^--\mathbf{r}')}\Big]e^{-i\omega_{\mathbf{k}}(t-t')}.\end{aligned} \tag{8.211}$$

Here $\delta^-_{\alpha\beta}$ is the same as the usual Kronecker delta, except that $\delta^-_{zz} = -1$. We can simplify this expression by considering the commutator only at equal times. Owing to the form of the wave vector in half-space, we can make the formal replacement

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{4\pi^3}\int d^3k, \tag{8.212}$$

where the integration is over half of reciprocal space. We can then extend the integration over all space and add a factor of $1/2$ to eliminate the real-part operator, with the result

$$\begin{aligned}[A_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t)] = -\frac{i\hbar}{2(2\pi)^3\epsilon_0}\int d^3k\Big[&\left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + \left(\delta_{\alpha\beta}-\frac{k^-_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}^-\cdot(\mathbf{r}-\mathbf{r}')}\\ &-\left(\delta^-_{\alpha\beta}-\frac{k_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}^-\cdot(\mathbf{r}^--\mathbf{r}')} - \left(\delta^-_{\alpha\beta}-\frac{k^-_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}^--\mathbf{r}')}\Big].\end{aligned} \tag{8.213}$$

If we define the "reflected" transverse delta function by

$$\begin{aligned}\delta^\top_{\alpha\beta}(\mathbf{r}) &:= \frac{1}{2(2\pi)^3}\int d^3k\left(\delta^-_{\alpha\beta}-\frac{k_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}^-\cdot\mathbf{r}} + \frac{1}{2(2\pi)^3}\int d^3k\left(\delta^-_{\alpha\beta}-\frac{k^-_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}}\\ &= \frac{1}{(2\pi)^3}\int d^3k\left(\delta^-_{\alpha\beta}-\frac{k^-_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}},\end{aligned} \tag{8.214}$$

this commutator simplifies to

$$[A_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t)] = -\frac{i\hbar}{\epsilon_0}\Big[\delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}') - \delta^\top_{\alpha\beta}(\mathbf{r}^--\mathbf{r}')\Big]. \tag{8.215}$$

(equal times, half-space)

Noting again that $\mathbf{r}^- - \mathbf{r}' = \mathbf{r} - \mathbf{r}' - 2z\hat{z} = \hat{x}(x-x') + \hat{y}(y-y') - \hat{z}(z+z')$, we see that now the commutator contains contributions from two intervals: the first is the *direct* separation $\mathbf{r} - \mathbf{r}'$, which is the same as in the free-space case, while the second is the separation including one bounce from the conducting surface (mirror). The second contribution also contains a flipped orientation of the delta function, which accounts for the reversal of an image-charge distribution with respect to the source-charge distribution. This seems a physically reasonable modification to the free-space commutator.

Similarly, we can write the electric-field commutator as

$$\begin{aligned}[E_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] &= \frac{i\hbar c}{2\epsilon_0 V}\sum_{\mathbf{k},\zeta} k\,\mathrm{Im}\Big[(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta)\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + (\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\beta)\, e^{i\mathbf{k}^-\cdot(\mathbf{r}-\mathbf{r}')} \\ &\qquad - (\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\beta)\, e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}'^-)} - (\hat{\varepsilon}_{\mathbf{k}^-,\zeta}\cdot\hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\hat{r}_\beta)\, e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}'^-)}\Big] \\ &\qquad \times e^{-i\omega_{\mathbf{k}}(t-t')} \\ &= \frac{i\hbar c}{\epsilon_0}\frac{1}{(2\pi)^3}\int d^3k\, k\Bigg[\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} + \left(\delta_{\alpha\beta} - \frac{k^-_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}^-\cdot(\mathbf{r}-\mathbf{r}')} \\ &\qquad - \left(\delta^-_{\alpha\beta} - \frac{k_\alpha k^-_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}'^-)} - \left(\delta^-_{\alpha\beta} - \frac{k^-_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot(\mathbf{r}^--\mathbf{r}')}\Bigg] \\ &\qquad \times \sin[\omega_{\mathbf{k}}(t-t')].\end{aligned} \tag{8.216}$$

Carrying out the same procedure as before, we can write this commutator in terms of the singular D function as

$$[E_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] = \frac{i\hbar c}{\epsilon_0}\left[\frac{\delta_{\alpha\beta}}{c^2}\partial_t\partial_{t'} - \partial_\alpha\partial'_\beta\right]D(\mathbf{r}-\mathbf{r}', t-t') - \frac{i\hbar c}{\epsilon_0}\left[\frac{\delta^-_{\alpha\beta}}{c^2}\partial_t\partial_{t'} - \partial_\alpha\partial'_\beta\right]D(\mathbf{r}^- - \mathbf{r}', t-t'), \tag{8.217}$$

(half-space)

where again $\partial'_\alpha := \partial/\partial r'_\alpha$. Again, we see that we have the same form as before, but now the simultaneous measureability is excluded also by an additional term corresponding to the light cone that includes a bounce off of the mirror.[15] The second term has precisely the same form as before, except for sign differences in the z-related components. Recall that the orientation of a dipole image has similar sign modifications compared to the original.

8.7 Unconfined Mode Functions

Recall from Section 8.4 that we quantized electromagnetic field modes inside a quantization volume. In situations without a cavity, such as in the Weisskopf–Wigner calculation for the rate of spontaneous emission in free space (Chapter 11), the quantization volume corresponds to a fictitious cavity. Generally, the quantization volume cancels in the relevant physical quantities for these calculations. For example, the free-space mode functions from Section 8.4.2 are

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \frac{1}{\sqrt{V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{8.218}$$

where again $\mathbf{k}$ takes on discrete values due to the periodic boundary conditions, and the mode functions are normalized according to

$$\int_V d^3r\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\cdot\mathbf{f}^*_{\mathbf{k}',\zeta'}(\mathbf{r}) = \delta^3_{\mathbf{k},\mathbf{k}'}\delta_{\zeta,\zeta'}. \tag{8.219}$$

[15] P. W. Milonni, "Casimir forces without the vacuum radiation field," *Physical Review A* **25**, 1315 (1982) (doi: 10.1103/PhysRevA.25.1315).

The sum over modes amounts to something of the form [c.f. Eq. (11.23)]

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{(2\pi)^3}\int d^3k, \tag{8.220}$$

where the integration extends over all possible orientations of $\mathbf{k}$. Generally, physical quantities involve an integrand quadratic in the mode function, and thus the factors of V cancel. The subsequent limit $V \longrightarrow \infty$ for true free space is then trivial.

Thus, for problems where no physical cavity is involved, it is convenient to define mode functions where the limit $V \longrightarrow \infty$ is *already* taken. We can take the large-volume limit of Eq (8.219) by noting that in this limit the spacings between adjacent values of k_x, k_y, and k_z become small (scaling as $2\pi/\sqrt[3]{V}$), and so we can make the replacement

$$\delta^3_{\mathbf{k},\mathbf{k}'} \longrightarrow (\Delta k)^3\delta^3(\mathbf{k}-\mathbf{k}') = \frac{(2\pi)^3}{V}\delta^3(\mathbf{k}-\mathbf{k}'). \tag{8.221}$$

This is again so that the sum of the left-hand behaves as the integral over the right-hand side. Then we can write

$$\int_V d^3r\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\cdot\mathbf{f}^*_{\mathbf{k}',\zeta'}(\mathbf{r}) = \frac{(2\pi)^3}{V}\delta^3(\mathbf{k}-\mathbf{k}')\delta_{\zeta,\zeta'} \tag{8.222}$$

in the large-volume limit. Then absorbing a factor of $\sqrt{V}$ into the definition of the mode function by the rescaling

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) \longrightarrow \frac{1}{\sqrt{V}}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}), \tag{8.223}$$

the normalization for unbounded mode functions becomes

$$\int d^3r\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\cdot\mathbf{f}^*_{\mathbf{k}',\zeta'}(\mathbf{r}) = (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}')\,\delta_{\zeta,\zeta'}. \tag{8.224}$$

The sum over modes, assuming a summand quadratic in the mode functions (as is usually the case), then is given by the correspondence

$$\sum_{\mathbf{k}} \longrightarrow \frac{1}{(2\pi)^3}\int d^3k. \tag{8.225}$$

The free-space mode functions are then

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{8.226}$$

but now a continuous vector index $\mathbf{k}$. In general, the mode functions for *any* situation (e.g., half-space) will be independent of V, and can be obtained by setting $V \longrightarrow 1$ in the expressions for the functions quantized in a finite volume.

8.8 Hamiltonian Viewpoint of Electromagnetic Gauge Freedom

Finally, we will return to the Hamiltonian structure of the electromagnetic field, and examine more closely the gauge freedom that we swept under the rug by choosing a particular gauge.[16] We start with the Lagrangian for the mass-free electromagnetic field, with sources, including both the vector potential $\mathbf{A}$ and the scalar potential ϕ:

$$\begin{aligned} L &= L_{\text{free}} + L_{\text{source}} \\ L_{\text{free}} &= \frac{\epsilon_0}{2}\int d^3r\left[(\nabla\phi+\partial_t\mathbf{A})^2 - c^2(\nabla\times\mathbf{A})^2\right] \\ L_{\text{source}} &= -\int d^3r\left[\rho\phi - \mathbf{j}\cdot\mathbf{A}\right]. \end{aligned}$$

(electromagnetic Lagrangian) (8.227)

[16] This is essentially Dirac's treatment of Hamiltonian dynamics with constraints, P. A. M. Dirac, "Generalized Hamiltonian dynamics," *Canadian Journal of Mathematics* **2**, 129 (1950) (doi: 10.4153/CJM-1950-012-1); but applied to the electromagnetic field, as in P. A. M. Dirac, "The Hamiltonian form of field dynamics," *Canadian Journal of Mathematics* **3**, 1 (1951) (doi: 10.4153/CJM-1951-001-2). I learned the following treatment from Tanmoy Bhattacharya (unpublished).

Here, we have broken the Lagrangian into the free part and a source part that represents the coupling to the source fields ρ (charge) and $\mathbf{j}$ (current density). We will take this Lagrangian to be the fundamental starting point. However, we can also note that the form of the Lagrangian motivates the definitions

$$\begin{aligned}\mathbf{E} &:= -\nabla\phi - \partial_t \mathbf{A} \\ \mathbf{B} &:= \nabla \times \mathbf{A}\end{aligned} \tag{8.228}$$

(electromagnetic fields)

for the electric and magnetic fields. Then in terms of these fields, the simple quadratic nature of the Lagrangian is more apparent:

$$L_{\text{free}} = \frac{\epsilon_0}{2} \int d^3r \left[\mathbf{E}^2 - c^2 \mathbf{B}^2\right].$$

(free electromagnetic Lagrangian) (8.229)

We will also soon see that the electric field is again essentially the momentum field conjugate to the coordinate field $\mathbf{A}$. The minus sign in the definition of the electric field simply makes the gradient of the potential agree with the usual mechanical potential, $\dot{\mathbf{p}} = -\nabla V(\mathbf{q})$.

8.8.1 Hamiltonian

To obtain the Hamiltonian for the electromagnetic field, we functionally differentiate the Lagrangian:

$$\begin{aligned}\Pi_0 &:= \frac{\delta L}{\delta \dot{\phi}} = 0 \\ \mathbf{\Pi} &:= \frac{\delta L}{\delta \dot{\mathbf{A}}} = \epsilon_0(\nabla\phi + \dot{\mathbf{A}}) = -\epsilon_0 \mathbf{E}.\end{aligned} \tag{8.230}$$

(conjugate momenta)

The fact that the momentum Π_0 conjugate to ϕ vanishes indicates some funniness, and will ultimately lead to constraints and the gauge freedom on the field. Now the Hamiltonian is given by

$$H = \int d^3r \left[\Pi_0 \dot{\phi} + \mathbf{\Pi} \cdot \dot{\mathbf{A}}\right] - L. \tag{8.231}$$

We can drop the vanishing Π_0 piece, and simplifying, we can write the Hamiltonian as

$$\begin{aligned}H &= H_{\text{free}} + H_{\text{source}} \\ H_{\text{free}} &= \int d^3r \left[\frac{\mathbf{\Pi}^2}{2\epsilon_0} - \mathbf{\Pi} \cdot \nabla\phi + \frac{1}{2}\epsilon_0 c^2 (\nabla \times \mathbf{A})^2\right] \\ H_{\text{source}} &= \int d^3r \left[\rho\phi - \mathbf{j} \cdot \mathbf{A}\right].\end{aligned}$$

(electromagnetic Hamiltonian) (8.232)

Note again that while ϕ appears here, the conjugate momentum does not. In fact, since we are constrained to $\Pi_0 = 0$, the Hamiltonian is arbitrary up to a term proportional to Π_0 anyway. Thus we should generalize the Hamiltonian so that

$$\begin{aligned}H &= H_{\text{free}} + H_{\text{source}} + H_{\text{gauge}} \\ H_{\text{gauge}} &= \int d^3r \, \Pi_0 g,\end{aligned} \tag{8.233}$$

(electromagnetic Hamiltonian)

where $g(\Pi_0, \mathbf{\Pi}, \phi, \mathbf{A}; \mathbf{r}, t)$ is an arbitrary function, and will represent part of the gauge freedom of the field.

8.8.2 Hamilton Equations and Gauge Freedom

Now we can work out the dynamics of the field in terms of Hamilton's equations. First we start with the equation of motion for the scalar momentum:

$$\dot{\Pi}_0 = -\frac{\delta H}{\delta \phi} = -\nabla \cdot \mathbf{\Pi} - \rho. \tag{8.234}$$

However, $\Pi_0 = 0$, so the last expression here must vanish, leading to one of the Maxwell equations:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \tag{8.235}$$
(Gauss' law)

Note that there is no time derivative here: this is a *constraint* equation, not an evolution equation. We ostensibly started with an evolution equation, but constraining $\Pi_0 = 0$ turned this into a constraint.

Next, the evolution equation for the scalar potential is

$$\dot{\phi} = \frac{\delta H}{\delta \Pi_0} = g + \Pi_0 \frac{\delta g}{\delta \Pi_0} = g. \tag{8.236}$$

The last equality follows by assuming the derivative of g to be finite—then the last term vanishes, since it includes a factor of $\Pi_0 = 0$. The time derivative of ϕ is thus given by a completely arbitrary function, and so ϕ itself may be arbitrarily chosen. This is our first gauge freedom. A common choice is to use this freedom to set $\phi = 0$, which implies the choice $g = 0$. Note that we can already introduce a second gauge freedom, as we have the conserved quantity $(\nabla \cdot \mathbf{\Pi} + \rho) = 0$ from the constraint (Maxwell equation) above. Thus, we modify the gauge Hamiltonian according to

$$H_{\text{gauge}} = \int d^3r \Big[\Pi_0 g + (\nabla \cdot \mathbf{\Pi} + \rho)h\Big], \tag{8.237}$$
(gauge Hamiltonian)

where $h(\Pi_0, \mathbf{\Pi}, \phi, \mathbf{A}; \mathbf{r}, t)$ is another arbitrary function. Note that introducing the h term changes nothing that we have done so far, since we have not yet differentiated with respect to $\mathbf{\Pi}$.

Working out the equation of motion for $\mathbf{\Pi}$,

$$\dot{\mathbf{\Pi}} = -\frac{\delta H}{\delta \mathbf{A}} = -\epsilon_0 c^2 \nabla \times \nabla \times \mathbf{A} + \mathbf{j}. \tag{8.238}$$

Converting to the usual fields, we find the next Maxwell equation

$$\nabla \times \mathbf{B} = \frac{1}{c^2}\dot{\mathbf{E}} + \mu_0 \mathbf{j}, \tag{8.239}$$
(Ampère's law)

noting that $c^2 = 1/\mu_0\epsilon_0$. The other two Maxwell equations follow from the definition of the fields: $\nabla \cdot \mathbf{B} = \nabla \times \nabla \cdot \mathbf{A} = 0$ is the magnetic-monopole law, and $\nabla \times \mathbf{E} = \nabla \times (-\nabla\phi - \partial_t \mathbf{A}) = -\partial_t \nabla \times \mathbf{A} = -\partial_t \mathbf{B}$ is Faraday's law.

The equation of motion for the vector potential $\mathbf{A}$ is

$$\dot{\mathbf{A}} = -\frac{\delta H}{\delta \mathbf{\Pi}} = \frac{\mathbf{\Pi}}{\epsilon_0} - \nabla\phi - \nabla h. \tag{8.240}$$

The interpretation here is that $\dot{\mathbf{A}}$ is arbitrary up to a a gradient of a function, which has the same form as the arbitrariness in ϕ. If f is the antiderivative of h (i.e., $\dot{f} = h$), then in addition to the first gauge freedom above,

$$\phi \text{ can be modified arbitrarily,} \tag{8.241}$$
(first gauge freedom)

this second gauge freedom amounts to freedom to modify the vector potential under transformations of the form

$$\mathbf{A} \longrightarrow \mathbf{A} + \nabla f. \tag{8.242}$$
(second gauge freedom)

In starting with four coordinate fields $(\phi, \mathbf{A})$ and four momentum fields $(\Pi_0, \mathbf{\Pi})$, we eliminate two momenta via $\Pi_0 = 0$ and $\nabla \cdot \mathbf{\Pi} = -\rho$, and by finding that we can introduce two arbitrary fields, ϕ and f, we essentially eliminate two of the coordinate fields. Thus, the remaining independent canonical coordinates—two momenta and two configuration coordinates—correspond to two independent degrees of freedom, corresponding to the two independent polarizations in the transverse fields. The longitudinal fields are counted separately, since we have implicitly introduced them via the source charge ρ.

To connect this gauge freedom with the more usual treatment, we can solve (8.240) for the momentum to obtain

$$\mathbf{\Pi} = \epsilon_0(\dot{\mathbf{A}} + \nabla\phi + \nabla h). \tag{8.243}$$

This expression is the same as in Eq. (8.230), except for the presence of the gradient of h. In the Hamiltonian formulation, this is not an inconsistency: after deriving Eq. (8.230) for the momentum field and the Hamiltonian, we have modified the Hamiltonian by introducing H_{gauge}, and so in this case the momentum changes to reflect this. However, to compare more directly to the Lagrangian picture, we can require that the momentum field not change. In this case, we should use the arbitrariness in the scalar potential to modify it to cancel the extra h term, which would mean setting $\phi \longrightarrow \phi - h$, to correspond to $\dot{\mathbf{A}} \longrightarrow \dot{\mathbf{A}} + \nabla h$. Thus, we have a combined gauge invariance under the combined transformation

$$\begin{aligned} \mathbf{A} &\longrightarrow \mathbf{A} + \nabla f \\ \phi &\longrightarrow \phi - \partial_t f. \end{aligned} \tag{8.244}$$
(Lagrangian gauge freedom)

However, in the Hamiltonian picture, this link between the transformations is not explicitly enforced; if ϕ is set to zero, then for example to model the effects of a static charge, we must still pick an appropriate form of $\mathbf{A}$.

8.8.3 Continuity Constraint

One last detail comes from again considering Eq. (8.234), where the right-hand-side expression $\nabla \cdot \mathbf{\Pi} + \rho$ vanishes, and is thus a constant of the motion. We can then differentiate it,

$$\partial_t(\nabla \cdot \mathbf{\Pi} + \rho) = \nabla \cdot \dot{\mathbf{\Pi}} + \dot{\rho} = \nabla \cdot \mathbf{j} + \dot{\rho}, \tag{8.245}$$

where we have used Eq. (8.238) for $\dot{\mathbf{\Pi}}$. This quantity vanishes, and we are left with

$$\partial_t \rho + \nabla \cdot \mathbf{j} = 0, \tag{8.246}$$
(continuity constraint)

which is the usual continuity condition for the source fields.

8.9 Quantization with Dielectric Media

We have so far restricted quantization to free space, possibly with boundary conditions that modify the spatial mode functions. Here we will generalize field quantization to handle dielectrics modeled by permittivities $\epsilon(\mathbf{r})$. We will *not* handle dispersion here, though, which causes considerably more complication for general quantization. We will handle dispersion only in the linear-response regime in Chapter 14, where we will not need to explicitly quantize the field.

8.9.1 Fresnel Coefficients: Modes of a Planar, Dielectric Interface

As another example of unconfined mode functions in the sense of the last section, and as a starting point for field quantization with dielectric media, we will consider the field modes in the presence of an interface between two dielectric media. We will take Maxwell's equations for the electromagnetic fields in a medium,

but without sources, to be our starting point:

$$\begin{aligned}\nabla \cdot \mathbf{D} &= 0\\ \nabla \cdot \mathbf{B} &= 0\\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B}\\ \nabla \times \mathbf{H} &= \partial_t \mathbf{D}.\end{aligned} \tag{8.247}$$
(Maxwell's equations)

These equations are covered in more detail later (see Section 14.1, so we will use them without much discussion. Here, $\mathbf{D}$ is the **electric flux density** or **electric displacement**, $\mathbf{B}$ is the **magnetic flux density**, and $\mathbf{E}$ and $\mathbf{H}$ are the usual electric and magnetic fields, respectively. We will ignore magnetic effects, so that

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{8.248}$$

and the electric fields are related by

$$\mathbf{D} = \epsilon \mathbf{E}, \tag{8.249}$$

with $\epsilon = \epsilon_0(1 + \chi)$, in terms of the **susceptibility** χ of the dielectric medium. Both $\epsilon(\mathbf{r})$ and $\chi(\mathbf{r})$ are functions of space for an inhomogeneous optical medium.

Now we will decouple the equations to derive a wave equation for the electric field. The curl of the third Maxwell equation is

$$\nabla \times (\nabla \times \mathbf{E}) = -\partial_t \nabla \times \mathbf{B}, \tag{8.250}$$

and with $\mathbf{B} = \mu_0 \mathbf{H}$ and the fourth Maxwell equation, we arrive at the wave equation

$$\nabla \times (\nabla \times \mathbf{E}) = -\mu_0 \epsilon \partial_t^2 \mathbf{E}, \tag{8.251}$$
(electric-field wave equation)

after decoupling the electric displacement with $\mathbf{D} = \epsilon \mathbf{E}$. For a *homogeneous* medium of constant ϵ, this simplifies by using the identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ to replace the iterated curl, where then the result simplifies by using $\nabla(\nabla \cdot \mathbf{E}) = \nabla(\nabla \cdot \mathbf{D})/\epsilon = 0$ to obtain

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon \partial_t^2 \mathbf{E}.$$
(electric-field wave equation, homogeneous medium) (8.252)

The homogeneous wave equation has plane-wave solutions of the form

$$\mathbf{E}^{(+)}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}, \tag{8.253}$$
(plane-wave solution)

provided

$$\frac{\omega}{k} = \frac{1}{\sqrt{\mu_0 \epsilon}}. \tag{8.254}$$
(dispersion relation/phase velocity)

This is also the phase velocity of the wave, denoted c, which is commonly written in terms of the **refractive index** $n := \sqrt{\epsilon/\epsilon_0}$. The last Maxwell equation implies an accompanying magnetic field of the same form, with the same phase, but orthogonal to $\mathbf{E}$. More generally, the transverse nature of the plane wave solutions is apparent from the last two Maxwell equations, from which it follows that $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{k}$ form a mutually orthogonal vector basis, provided the medium is homogeneous and isotropic. Further, $\mathbf{E} \times \mathbf{B}$ points in the direction of $\mathbf{k}$.

8.9.1.1 Boundary Conditions

To analyze the electric field at the interface, we will need to consider the boundary conditions on the field at the material discontinuity at the interface. Because the interface produces only a *finite* discontinuity in the

wave equation, we will use this to enforce continuity conditions across the interface. Since we are considering the inhomogeneous part of space, we can write out Eq. (8.251) which for the monochromatic mode becomes

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\mu_0 \epsilon \omega^2 \mathbf{E}, \tag{8.255}$$

It is convenient to separate this equation into normal and transverse components. The normal component is

$$\partial_z(\nabla \cdot \mathbf{E}) - \nabla^2 E_z = -\mu_0 \epsilon \omega^2 E_z, \tag{8.256}$$

or since $E_z = 0$

$$\partial_z(\nabla \cdot \mathbf{E}) - \nabla^2 E_z = -\mu_0 \epsilon \omega^2 E_z, \tag{8.257}$$

and canceling the z derivatives,

$$\partial_z(\nabla_{\mathrm{T}} \cdot \mathbf{E}) - \nabla_{\mathrm{T}}^2 E_z = -\mu_0 \epsilon \omega^2 E_z, \tag{8.258}$$

where $\nabla_{\mathrm{T}}^2 := \partial_x^2 + \partial_y^2$ is the transverse Laplacian. Examining the behavior of this equation in the z direction, since ϵ only makes a finite step, we can integrate this equation from $z = -\delta$ to δ, letting $\delta \longrightarrow 0$. Only the first term survives the limit, in which case $\nabla_{\mathrm{T}} \cdot \mathbf{E}(\delta) = \nabla_{\mathrm{T}} \cdot \mathbf{E}(-\delta)$. Thus, the first boundary condition is that $\nabla_{\mathrm{T}} \cdot \mathbf{E}$ is continuous across the boundary. To obtain a second boundary condition, we return to Eq. (8.255), considering now the components parallel to the interface:

$$\nabla_{\mathrm{T}}(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}_{\|} = -\mu_0 \epsilon \omega^2 \mathbf{E}_{\|}. \tag{8.259}$$

Separating out the z derivatives, we then have

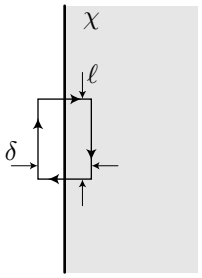
$$\nabla_{\mathrm{T}}(\partial_z E_z) + \nabla_{\mathrm{T}}(\nabla_{\mathrm{T}} \cdot \mathbf{E}_{\|}) - \nabla_{\mathrm{T}}^2 \mathbf{E}_{\|} - \partial_z^2 \mathbf{E}_{\|} = -\mu_0 \epsilon \omega^2 \mathbf{E}_{\|}. \tag{8.260}$$

Again, we consider this to be a differential equation in z, where the only z derivatives appear in the first and fourth terms. Since the coefficients of the equation are bounded, we can assume all the terms of this equation are also bounded in the neighborhood of the interface, which holds so long as the field itself is bounded in the region near the boundary, which is a reasonable physical assumption in the absence of localized sources there. Then either $\partial_z E_z$ is finitely discontinuous or it is continuous. In either case, we can integrate over a vanishingly small interval in z across the boundary, in which case we obtain

$$\nabla_{\mathrm{T}}(\delta E_z) - \delta \partial_z \mathbf{E}_{\|} = 0, \tag{8.261}$$

where the δ symbols here indicate reference the change in the following quantity over the interface. This means that the change in the derivative of $\mathbf{E}_{\|}$ is bounded, which means that $\mathbf{E}_{\|}$ itself is continuous across the boundary, which is our second condition. Actually, this turns out to be equivalent to the first condition, since $\nabla_{\perp}$ will be equivalent to $\mathbf{k}_{\|}$ at the interface, and we will show below that all parallel components of the wave vectors are the same at the boundary. Thus, we will allow the second condition to supersede the first one, keeping the simpler requirement of continuity of $\mathbf{E}_{\|}$.

To obtain the other boundary condition we will need, we can return to the wave equation (8.251), and integrate around a rectangular contour in a plane normal to the (local) interface that encompasses the boundary as shown below.



The width δ of the loop will contract to zero, while we assume the length ℓ of the loop is short compared to any length scale of the interface (in the case of a nonplanar interface) or field. Then integrating $\nabla \times \mathbf{E}$ around this contour, we can then use Stokes' theorem and then Eq. (8.251) to find

$$\oint (\nabla \times \mathbf{E}) \cdot d\mathbf{r} = \int [\nabla \times (\nabla \times \mathbf{E})] \cdot d\mathbf{a} = -\mu_0 \int (\epsilon \partial_t^2 \mathbf{E}) \cdot d\mathbf{a}, \tag{8.262}$$

where $\mathbf{a}$ is the area-normal vector, and the integrals run over the area enclosed by the contour. As $\delta \longrightarrow 0$, the last integral vanishes, and only the segments of length ℓ contribute to the contour integral on the left-hand side, such that the contributions of the two segments cancel. Since this argument can be repeated in any plane normal to the interface it implies that the component of $\nabla \times \mathbf{E}$ parallel to the interface is continuous across the boundary. Thus, our second boundary condition is that $(\nabla \times \mathbf{E})_\parallel$ is continuous across the interface.

It is possible to derive another boundary condition, though it is one we will not use here. Computing the divergence of Eq. (8.251) for a monochromatic field, we have

$$\nabla \cdot \epsilon \mathbf{E} = 0. \tag{8.263}$$

This is a statement that according to the wave equation, $\epsilon\mathbf{E}$ must be a transverse field, or that $\mathbf{E}$ must be "ϵ-transverse" in this sense. We note that the only term involving a z derivative is $\partial_z \epsilon E_z$, so integrating this equation over a short interval across the interface in the z direction shows that ϵE_z must be continuous across the boundary.

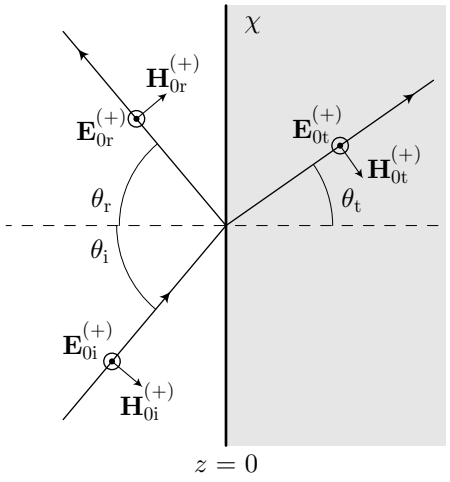
In summary, we have

$$\mathbf{E}_\parallel \text{ is continuous}, \qquad (\nabla \times \mathbf{E})_\parallel \text{ is continuous} \tag{8.264}$$
(boundary conditions)

for the boundary conditions at the interface for the solution $\mathbf{E}$ of the vector wave equation. Note that in this form, the boundary conditions *only* refer to the electric field itself, and is thus more "self-contained" than boundary conditions that are typically written for both $\mathbf{E}$ and $\mathbf{H}$. (Though note that $\mathbf{B}$ satisfies the same boundary conditions, since it satisfies the same wave equation as does $\mathbf{E}$.)

8.9.1.2 TE Modes

Now we will consider the field modes in the presence of a dielectric interface, where the dielectric covers $z < 0$ and $z > 0$ is vacuum. Since the permeability is constant in each region, the solutions are plane waves (or superpositions of plane waves) in each region. First, we will make the *ansatz* of a solution that consists of three parts: an incident plane wave and a reflected plane wave on the vacuum side, and a transmitted wave on the dielectric side. We will take all of the propagation vectors to lie in the same plane. The transverse electric field also has two independent components, which we will treat separately. The two polarizations that respect the symmetry of the surface are TE (transverse electric, meaning that the electric field is parallel to the interface) and TM (transverse magnetic, with the magnetic field parallel to the interface) polarizations. First, we will consider TE polarization, and sketch our *ansatz* below.



Although we will refer only to the electric field in the following, we have sketched in the proper relative orientations of the magnetic fields relative to the (electric-field) polarization vectors. Writing out our assumptions mathematically, we have the monochromatic TE field mode

$$\mathbf{E}_{\mathbf{k},\mathrm{TE}}^{(+)} = E_{0\mathrm{i}}^{(+)}\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\left[e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(z) + r_{\mathrm{TE}}e^{i\mathbf{k}_\mathrm{r}\cdot\mathbf{r}}\Theta(z) + t_{\mathrm{TE}}e^{i\mathbf{k}_\mathrm{t}\cdot\mathbf{r}}\Theta(-z)\right], \tag{8.265}$$

where $\mathbf{k}$, $\mathbf{k}_\mathrm{r}$, and $\mathbf{k}_\mathrm{t}$ are the incident, reflected, and transmitted wave vectors, $E_{0\mathrm{i}}$, $E_{0\mathrm{r}}$, and $E_{0\mathrm{t}}$ are the respective field amplitudes, and we have defined $r_{\mathrm{TE}} := E_{0\mathrm{r}}/E_{0\mathrm{i}}$ and $t_{\mathrm{TE}} := E_{0\mathrm{t}}/E_{0\mathrm{i}}$ as the field reflection and transmission coefficients, respectively. A common time dependence of $\exp(-i\omega t)$ is implied for all the terms. But while ω is the same on both sides of the boundary, in view of the dispersion relation (8.254), k should be viewed as a function of ϵ ($k \propto \sqrt{\epsilon} \propto n$).

The first boundary condition (8.264), continuity of the surface-parallel component $\mathbf{E}_\parallel$, gives simply for the mode (8.265)

$$e^{i\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel} + r_{\mathrm{TE}}e^{i\mathbf{k}_{\mathrm{r}\parallel}\cdot\mathbf{r}_\parallel} = t_{\mathrm{TE}}e^{i\mathbf{k}_{\mathrm{t}\parallel}\cdot\mathbf{r}_\parallel}, \tag{8.266}$$

since the polarizations are all already transverse. Here, the notation $\mathbf{r}_\perp$ refers to only the transverse part of the vector (i.e., only the x and y components, since $z = 0$ here). The only way for the phases to match at every point on the interfaces is for the transverse parts of all the wave vectors themselves to be separately equal,

$$\mathbf{k}_\parallel = \mathbf{k}_{\mathrm{r}\parallel} = \mathbf{k}_{\mathrm{t}\parallel}. \tag{8.267}$$

The first condition here leads to

$$\mathbf{k}_\mathrm{r} = \mathbf{k}^-, \tag{8.268}$$

where $\mathbf{k}^- = k_x\hat{x} + k_y\hat{y} - k_z\hat{z}$, since we assumed the incident and reflected waves to propagate toward and away from the interface, respectively. This in turn leads to the usual reflection law $\theta_\mathrm{i} = \theta_\mathrm{r}$. The other equality in the phase condition,

$$\mathbf{k}_\parallel = \mathbf{k}_{\mathrm{t}\parallel}, \tag{8.269}$$

can be rewritten as

$$k_\mathrm{i}\sin\theta_\mathrm{i} = k_\mathrm{t}\sin\theta_\mathrm{t}, \tag{8.270}$$

or more familiarly as Snell's law,

$$n_\mathrm{i}\sin\theta_\mathrm{i} = n_\mathrm{t}\sin\theta_\mathrm{t}, \tag{8.271}$$

if we write the relation in terms of the refractive index. The first boundary condition (8.266) then becomes

$$1 + r_{\mathrm{TE}} = t_{\mathrm{TE}}. \tag{8.272}$$

The second of the boundary conditions (8.264), i.e., enforcing the continuity of $(\nabla\times\mathbf{E})_\parallel$, becomes

$$(\mathbf{k}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel + (\mathbf{k}_\mathrm{r}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel r_{\mathrm{TE}} = (\mathbf{k}_\mathrm{t}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel t_{\mathrm{TE}}, \tag{8.273}$$

or working out the cross products,

$$k_z(1 - r_{\text{TE}}) = k_{\text{t},z} t_{\text{TE}}. \tag{8.274}$$

Eqs. (8.272) and (8.274) have the solutions

$$r_{\text{TE}} = \frac{k_z - k_{\text{t},z}}{k_z + k_{\text{t},z}}, \qquad t_{\text{TE}} = \frac{2k_z}{k_z + k_{\text{t},z}}. \tag{8.275}$$

(TE Fresnel coefficients)

Another useful representation of these coefficients comes from noting that $k_z = k\cos\theta_{\text{i}}$ and

$$\begin{aligned} k_{\text{t},z} &= k_{\text{t}} \cos\theta_{\text{t}} \\ &= k\sqrt{1+\chi}\sqrt{1 - \sin^2\theta_{\text{t}}} \\ &= k\sqrt{1+\chi}\sqrt{1 - \frac{1}{(1+\chi)}\sin^2\theta_{\text{i}}} \\ &= k\sqrt{1 + \chi - \sin^2\theta_{\text{i}}} \\ &= k\sqrt{\chi + \cos^2\theta_{\text{i}}}, \end{aligned} \tag{8.276}$$

so that

$$r_{\text{TE}} = \frac{\xi - \sqrt{\chi + \xi^2}}{\xi + \sqrt{\chi + \xi^2}}, \qquad t_{\text{TE}} = \frac{2\xi}{\xi + \sqrt{\chi + \xi^2}}, \tag{8.277}$$

(TE Fresnel coefficients)

where $\xi = \cos\theta_{\text{i}}$. Still another form comes from writing out the angles more explicitly, and using $k_{\text{i}} = n_1 k_0$ and $k_{\text{t}} = n_2 k_0$, where k_0 is the vacuum wave number, and n_1 and n_2 are the refractive indices on the left- and right-and sides of the interface, respectively, so that

$$r_{\text{TE}} = \frac{n_1\cos\theta_{\text{i}} - n_2\cos\theta_{\text{t}}}{n_1\cos\theta_{\text{i}} + n_2\cos\theta_{\text{t}}}, \qquad t_{\text{TE}} = \frac{2n_1\cos\theta_{\text{i}}}{n_1\cos\theta_{\text{i}} + n_2\cos\theta_{\text{t}}}.$$

(TE Fresnel coefficients) (8.278)

(While we are nominally assuming a vacuum on one side of the interface, the k-vector treatment is actually more general than this.) Fixing these coefficients also fixes the form of the mode ansatz (8.265). We can then define a normalized mode function

$$\mathbf{f}_{\mathbf{k},\text{TE}} = \hat{\varepsilon}_{\mathbf{k},\text{TE}} \left[e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(z) + r_{\text{TE}} e^{i\mathbf{k}_{\text{r}}\cdot\mathbf{r}}\Theta(z) + t_{\text{TE}} e^{i\mathbf{k}_{\text{t}}\cdot\mathbf{r}}\Theta(-z) \right].$$

(normalized mode function) (8.279)

To work out the normalization of this mode function, first we write out

$$\begin{aligned} &\int d^3r\, \frac{\epsilon(\mathbf{r})}{\epsilon_0}\, \mathbf{f}^*_{\mathbf{k},\text{TE}}(\mathbf{r}) \cdot \mathbf{f}_{\mathbf{k}',\text{TE}}(\mathbf{r}) \\ &\qquad = \int d^3r \left\{ e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}}\Theta(z) + |r_{\text{TE}}|^2 e^{i(\mathbf{k}'_{\text{r}}-\mathbf{k}_{\text{r}})\cdot\mathbf{r}}\Theta(z) + \left[r_{\text{TE}} e^{i(\mathbf{k}'-\mathbf{k}_{\text{r}})\cdot\mathbf{r}} + r^*_{\text{TE}} e^{i(\mathbf{k}-\mathbf{k}'_{\text{r}})\cdot\mathbf{r}} \right]\Theta(z) \right. \\ &\qquad\qquad \left. + \frac{\epsilon}{\epsilon_0} |t_{\text{TE}}|^2 e^{i(\mathbf{k}'_{\text{t}}-\mathbf{k}_{\text{t}})\cdot\mathbf{r}}\Theta(-z) \right\}, \end{aligned} \tag{8.280}$$

assuming the two waves are incident from the vacuum side. If we consider the real part of this equation, the conjugate terms are equivalent to the same terms already present, but with $\Theta(z) \longrightarrow \Theta(-z)$. Thus, we can

carry out the integrals to obtain δ functions, with the result

$$\begin{aligned}
&2\mathrm{Re}\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\mathrm{TE}}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\mathrm{TE}}(\mathbf{r})\\
&\qquad=(2\pi)^3\delta^3(\mathbf{k}'-\mathbf{k})+|r_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{r}})+(2\pi)^3\left[r_{\mathrm{TE}}\delta^3(\mathbf{k}'-\mathbf{k}_{\mathrm{r}})+r^*_{\mathrm{TE}}\delta^3(\mathbf{k}-\mathbf{k}'_{\mathrm{r}})\right]\\
&\qquad\quad+\frac{\epsilon}{\epsilon_0}\,|t_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{t}})\\
&\qquad=(2\pi)^3\delta^3(\mathbf{k}'-\mathbf{k})+|r_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{r}})+\frac{\epsilon}{\epsilon_0}\,|t_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{t}}),
\end{aligned}\tag{8.281}$$

where we have removed the δ functions that are always zero due to the relative orientations of incident and reflected waves, under the assumption that both modes are incident from the same direction. To simplify this expression further, we should write the wave vectors in the last term in terms of the incident wave vector. This procedure is slightly involved, and goes as follows.[17] First, recalling that

$$k_{\mathrm{t}}^2=\frac{\epsilon}{\epsilon_0}k^2\tag{8.282}$$

and from Snell's law,

$$k_{\mathrm{t}\parallel}=k_\parallel,\tag{8.283}$$

then we will now relate the z components of the incident and transmitted wave vectors. The incident version satisfies

$$k_z^2=k^2-k_\parallel^2,\tag{8.284}$$

while the transmitted version satisfies

$$k_{\mathrm{t},z}^2=k_{\mathrm{t}}^2-k_{\mathrm{t}\parallel}^2=\frac{\epsilon}{\epsilon_0}k^2-k_\parallel^2.\tag{8.285}$$

Rearranging and dividing these last two equations gives

$$\frac{k_{\mathrm{t},z}^2+k_\parallel^2}{k_z^2+k_\parallel^2}=\frac{\epsilon}{\epsilon_0}.\tag{8.286}$$

If we instead introduce a subtraction with these same two relations in terms of $\mathbf{k}'$ before division, we obtain

$$\frac{({k'}_{\mathrm{t},z}^2-k_{\mathrm{t},z}^2)+({k'}_\parallel^2-k_\parallel^2)}{({k'}_z^2-k_z^2)+({k'}_\parallel^2-k_\parallel^2)}=\frac{\epsilon}{\epsilon_0}.\tag{8.287}$$

Then using this relation, we have the relation

$$\begin{aligned}
\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{t}})&=\delta^2(\mathbf{k}'_{\mathrm{t}\parallel}-\mathbf{k}_{\mathrm{t}\parallel})\,\delta(k'_{\mathrm{t},z}-k_{\mathrm{t},z})\\
&=2|k_{\mathrm{t},z}|\delta^2(\mathbf{k}'_{\mathrm{t}\parallel}-\mathbf{k}_{\mathrm{t}\parallel})\,\delta[(k'_{\mathrm{t},z}-k_{\mathrm{t},z})(k'_{\mathrm{t},z}+k_{\mathrm{t},z})]\\
&=2|k_{\mathrm{t},z}|\delta^2(\mathbf{k}'_{\mathrm{t}\parallel}-\mathbf{k}_{\mathrm{t}\parallel})\,\delta({k'}_{\mathrm{t},z}^2-k_{\mathrm{t},z}^2)\\
&=2\frac{\epsilon_0}{\epsilon}|k_{\mathrm{t},z}|\delta^2(\mathbf{k}'_{\mathrm{t}\parallel}-\mathbf{k}_{\mathrm{t}\parallel})\,\delta({k'}_z^2-k_z^2)\\
&=\frac{\epsilon_0}{\epsilon}\left|\frac{k_{\mathrm{t},z}}{k_z}\right|\delta^2(\mathbf{k}'_{\mathrm{t}\parallel}-\mathbf{k}_{\mathrm{t}\parallel})\,\delta(k'_z-k_z)\\
&=\frac{\epsilon_0}{\epsilon}\left|\frac{k_{\mathrm{t},z}}{k_z}\right|\delta^2(\mathbf{k}'_\parallel-\mathbf{k}_\parallel)\,\delta(k'_z-k_z)\\
&=\frac{\epsilon_0}{\epsilon}\frac{k_{\mathrm{t},z}}{k_z}\delta^3(\mathbf{k}'-\mathbf{k}),
\end{aligned}\tag{8.288}$$

[17] C. K. Carniglia and L. Mandel, "Quantization of Evanescent Electromagnetic Waves," *Physical Review D* **3**, 280 (1971) (doi: 10.1103/PhysRevD.3.280).

where in the last step we noted that $k_{\mathrm{t},z}$ and k_z have the same sign by construction. Eq. (8.281) then becomes

$$2\mathrm{Re}\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\mathrm{TE}}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\mathrm{TE}}(\mathbf{r}) = (2\pi)^3\delta^3(\mathbf{k}'-\mathbf{k}) + |r_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{r}}) + \frac{k_{\mathrm{t},z}}{k_z}|t_{\mathrm{TE}}|^2(2\pi)^3\delta^3(\mathbf{k}'-\mathbf{k}). \tag{8.289}$$

To simplify this even further, we can multiply Eq. (8.272) with Eq. (8.274) to give

$$1 = r_{\mathrm{TE}}^2 + \frac{k_{\mathrm{t},z}}{k_z}t_{\mathrm{TE}}^2 =: R_{\mathrm{TE}} + T_{\mathrm{TE}}. \tag{8.290}$$
(reflectance and transmittance)

The interpretation here regarding complex-valued Fresnel coefficients is a bit subtle. Briefly, r_{TE} is real unless its modulus is unity. In the former case, t_{TE} is real. In the latter case, the transmitted wave is evanescent (decaying with z), T_{TE} is defined to vanish, and the t_{TE} wave is normalizable, so it does not contribute to the orthonormality relation. which we can interpret as conservation of energy: R and T here are the reflection and transmission coefficients of the *intensity*, respectively called the **reflectance** and **transmittance**. In this case, Eq. (8.289) reduces to

$$\mathrm{Re}\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\mathrm{TE}}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\mathrm{TE}}(\mathbf{r}) = (2\pi)^3\delta^3(\mathbf{k}'-\mathbf{k}). \tag{8.291}$$

Repeating the above argument for the imaginary part is simple: while the real part gave integrals over cosines leading to δ functions, the imaginary part gives corresponding integrals over sines, which simply vanish. Therefore, our orthonormality relation is

$$\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\mathrm{TE}}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\mathrm{TE}}(\mathbf{r}) = (2\pi)^3\delta^3(\mathbf{k}-\mathbf{k}'). \tag{8.292}$$
(orthonormality relation)

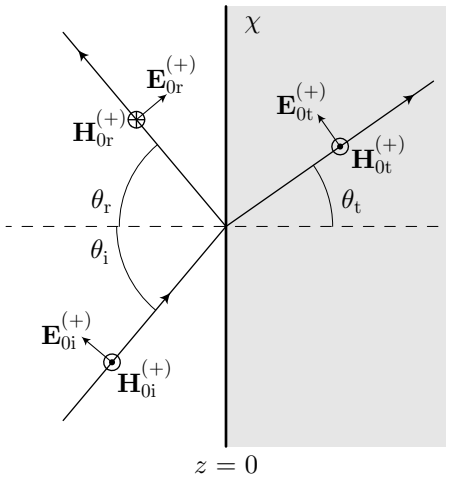
Note that for $\epsilon=\epsilon_0$, this reduces to the relation we expect for free-space mode functions. Technically, we need to verify this relation in a couple more situations. First, we should consider the case where both waves are incident from the dielectric side. This goes through in the same way as we have shown. (Though note that evanescent waves are possible in this case, in which case $\mathbf{k}_{\mathrm{t}}$ becomes complex.) Also, we should verify the case where one wave is incident from each side. In this case, Eq. (8.280) becomes

$$\begin{aligned}
&\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\mathrm{TE}}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\mathrm{TE}}(\mathbf{r})\\
&\quad= \int d^3r\left\{t_{\mathrm{TE}}e^{i(\mathbf{k}'_{\mathrm{t}}-\mathbf{k})\cdot\mathbf{r}}\Theta(z) + r^*_{\mathrm{TE}}t_{\mathrm{TE}}e^{i(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{r}})\cdot\mathbf{r}}\Theta(z)\right.\\
&\qquad\qquad\left. + \frac{\epsilon}{\epsilon_0}t^*_{\mathrm{TE}}e^{i(\mathbf{k}'-\mathbf{k}_{\mathrm{t}})\cdot\mathbf{r}}\Theta(-z) + \frac{\epsilon}{\epsilon_0}r_{\mathrm{TE}}t^*_{\mathrm{TE}}e^{i(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{t}})\cdot\mathbf{r}}\Theta(-z)\right\}\\
&\quad= \frac{(2\pi)^3}{2}\left\{t_{\mathrm{TE}}\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}) + r^*_{\mathrm{TE}}t_{\mathrm{TE}}\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{r}}) + \frac{\epsilon}{\epsilon_0}t^*_{\mathrm{TE}}\delta^3(\mathbf{k}'-\mathbf{k}_{\mathrm{t}}) + \frac{\epsilon}{\epsilon_0}r_{\mathrm{TE}}t^*_{\mathrm{TE}}\delta^3(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{t}})\right\}\\
&\quad= \frac{(2\pi)^3}{2}\left\{r^*_{\mathrm{TE}}t_{\mathrm{TE}}\delta^3(\mathbf{k}'_{\mathrm{t}}-\mathbf{k}_{\mathrm{r}}) + \frac{\epsilon}{\epsilon_0}r_{\mathrm{TE}}t^*_{\mathrm{TE}}\delta^3(\mathbf{k}'_{\mathrm{r}}-\mathbf{k}_{\mathrm{t}})\right\},
\end{aligned} \tag{8.293}$$

where we have assumed $\mathbf{k}$ to be incident from the vacuum side, and $\mathbf{k}'$ to be incident from the dielectric side. We have also dropped the interaction terms between incident and transmitted waves that always vanish. Now we must also be careful with the interpretation of the reflection and transmitted coefficients. The conjugated coefficients are incident from the vacuum side, whereas the normal ones are incident from the dielectric side. From Eq. (8.275), this means in particular that $r^*_{\mathrm{TE}} = -r_{\mathrm{TE}}$ and $t^*_{\mathrm{TE}} = (k_{\mathrm{t},z}/k_z)t_{\mathrm{TE}}$. Then the remaining two terms here cancel by applying Eq. (8.288) to the second δ function. Thus, oppositely-propagating modes are always orthogonal.

8.9.1.3 TM Modes

The other (TM) polarization follows in essentially the same wave as for the TE polarization. We will again sketch our *ansatz* and sign convention for the fields below, again recalling that the dielectric covers the region $z < 0$.



Again, we will refer only to the electric field in the following, but we have sketched in the proper relative orientations of the magnetic fields relative to the (electric-field) polarization vectors. Mathematically, we have the monochromatic TM field mode

$$\mathbf{E}_{\mathbf{k},\mathrm{TM}}^{(+)} = E_{0\mathrm{i}}^{(+)} \left[(\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}} \times \hat{k})\, e^{i\mathbf{k}\cdot\mathbf{r}} \Theta(z) + (\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}} \times \hat{k}_\mathrm{r})\, r_\mathrm{TM} e^{i\mathbf{k}_\mathrm{r}\cdot\mathbf{r}} \Theta(z) + (\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}} \times \hat{k}_\mathrm{t})\, t_\mathrm{TM} e^{i\mathbf{k}_\mathrm{t}\cdot\mathbf{r}} \Theta(-z) \right]. \quad (8.294)$$

Clearly, this mode is orthogonal to any TM mode. Applying the first of the boundary conditions (8.264) again implies the phase conditions (8.267)-(8.271). For the vector part of the mode, continuity of $\mathbf{E}_\parallel$ implies

$$(\hat{k} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel + (\hat{k}_\mathrm{r} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel r_\mathrm{TM} = (\hat{k}_\mathrm{t} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})_\parallel t_\mathrm{TM}, \quad (8.295)$$

[Note the similarity to Eq. (8.273) here.] Working out the cross products, we have

$$\frac{k_z}{k}(1 + r_\mathrm{TM}) = \frac{k_{\mathrm{t},z}}{k_\mathrm{t}} t_\mathrm{TM}. \quad (8.296)$$

The second boundary condition, continuity of $(\nabla \times \mathbf{E})_\parallel$ gives

$$[\mathbf{k} \times (\hat{k} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})]_\parallel + [\mathbf{k}_\mathrm{r} \times (\hat{k}_\mathrm{r} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})]_\parallel r_\mathrm{TM} = [\mathbf{k}_\mathrm{t} \times (\hat{k}_\mathrm{t} \times \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}})]_\parallel t_\mathrm{TM}, \quad (8.297)$$

so that working out the cross products and dividing through by k gives

$$1 - r_\mathrm{TM} = \frac{k_\mathrm{t}}{k} t_\mathrm{TM}. \quad (8.298)$$

Eqs. (8.296) and (8.298) have the solutions

$$r_\mathrm{TM} = \frac{k^2 k_{\mathrm{t},z} - k_\mathrm{t}^2 k_z}{k^2 k_{\mathrm{t},z} + k_\mathrm{t}^2 k_z}, \qquad t_\mathrm{TM} = \frac{2k k_\mathrm{t} k_z}{k^2 k_{\mathrm{t},z} + k_\mathrm{t}^2 k_z}. \quad (8.299)$$

(TM Fresnel coefficients)

which is the general form for a dielectric–dielectric interface, or using $k_\mathrm{t}/k = \sqrt{\epsilon}$ for a vacuum-dielectric interface,

$$r_\mathrm{TM} = \frac{k_{\mathrm{t},z} - \epsilon k_z}{k_{\mathrm{t},z} - \epsilon k_z}, \qquad t_\mathrm{TM} = \frac{2\sqrt{\epsilon} k_z}{k_{\mathrm{t},z} + \epsilon k_z}. \quad (8.300)$$

(TM Fresnel coefficients)

We can also write this in terms of $\xi = \cos\theta_\mathrm{i}$ using Eq. (8.276) as

$$r_{\mathrm{TM}} = \frac{\sqrt{\chi+\xi^2}-(1+\chi)\xi}{\sqrt{\chi+\xi^2}+(1+\chi)\xi}, \qquad t_{\mathrm{TM}} = \frac{2\sqrt{1+\chi}\xi}{\sqrt{\chi+\xi^2}+(1+\chi)\xi}.$$

(TM Fresnel coefficients) (8.301)

Additionally, we can write

$$r_{\mathrm{TM}} = \frac{n_1\cos\theta_\mathrm{t}-n_2\cos\theta_\mathrm{i}}{n_1\cos\theta_\mathrm{t}+n_2\cos\theta_\mathrm{i}}, \qquad t_{\mathrm{TM}} = \frac{2n_1\cos\theta_\mathrm{i}}{n_1\cos\theta_\mathrm{t}+n_2\cos\theta_\mathrm{i}}$$

(TM Fresnel coefficients) (8.302)

for a dielectric–dielectric interface with refractive indices n_1 and n_2 on the incident and transmitted side, respectively.

Now if we multiply together Eqs. (8.296) and (8.298), we obtain

$$1 = r_{\mathrm{TM}}^2 + \frac{k_{\mathrm{t},z}}{k_z}t_{\mathrm{TM}}^2 =: R_{\mathrm{TM}} + T_{\mathrm{TM}},$$

(8.303) (reflectance and transmittance)

where we used that k_t is real so long as $t_{\mathrm{TM}} \neq 0$. This is equivalent to the TE result (8.290), and thus the reflectance and transmittance for intensities has the same definition in terms of the respective reflection and transmission coefficients as for TE polarization. Then defining the mode function

$$\mathbf{f}_{\mathbf{k},\mathrm{TM}} = (\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\times\hat{k})\,e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(z) + (\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\times\hat{k}^-)\,r_{\mathrm{TM}}e^{i\mathbf{k}_\mathrm{r}\cdot\mathbf{r}}\Theta(z) + (\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\times\hat{k}_\mathrm{t})\,t_{\mathrm{TM}}e^{i\mathbf{k}_\mathrm{t}\cdot\mathbf{r}}\Theta(-z),$$

(normalized mode function) (8.304)

and noting that the phase behavior is the same as for the TE modes, we see that the orthonormality argument for the TE case goes through here as well. Furthermore, by construction the TE modes are orthogonal to the TM modes, so we have the general orthonormality relation

$$\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\,\mathbf{f}^*_{\mathbf{k},\zeta}(\mathbf{r})\cdot\mathbf{f}_{\mathbf{k}',\zeta'}(\mathbf{r}) = (2\pi)^3\delta_{\zeta\zeta'}\delta^3(\mathbf{k}-\mathbf{k}'),$$

(8.305) (orthonormality relation)

where ζ is either TE or TM.

8.9.2 Quantum Fields at a Planar Dielectric Interface

With the mode functions in hand, we can establish the form for the quantized fields, just as we did in free space. The derivation proceeds mostly in the same way as before, but with a few modifications.[18]

8.9.2.1 Hamiltonian Structure

The free-space Lagrangian (8.28) had the form

$$L = \frac{1}{2}\int d^3r\left(\epsilon_0E^2 - \frac{1}{\mu_0}B^2\right) \tag{8.306}$$

in terms of the electromagnetic fields, and has the obvious generalization

$$L = \frac{1}{2}\int d^3r\left(\epsilon(\mathbf{r})E^2 - \frac{1}{\mu_0}B^2\right)$$

(dielectric electromagnetic Lagrangian) (8.307)

[18] Here we are following the lucid treatment of Roy J. Glauber and M. Lewenstein, "Quantum optics of dielectric media," *Physical Review A* **43**, 467 (1991), Eq. (2.16a) (doi: 10.1103/PhysRevA.43.467).

in the presence of a dielectric. Again taking $\mathbf{A}$ to be the generalized coordinate, the conjugate momentum is

$$:= \frac{\delta L}{\delta(\partial_t \mathbf{A})} = \epsilon(\mathbf{r})\partial_t \mathbf{A} = -\epsilon(\mathbf{r})\mathbf{E} = -\mathbf{D}. \tag{8.308}$$
(conjugate momentum to $\mathbf{A}$)

This leads to the Hamiltonian

$$H = \int d^3r \left(\frac{\Pi^2}{2\epsilon(\mathbf{r})} + \frac{1}{2\mu_0}(\nabla \times \mathbf{A})^2 \right) = \frac{1}{2}\int d^3r \left(\epsilon(\mathbf{r})E^2 + \frac{1}{\mu_0}B^2 \right), \tag{8.309}$$
(dielectric electromagnetic Hamiltonian)

Here, the Hamilton equation $\partial_t \mathbf{\Pi} = -\delta H/\delta \mathbf{A}$ leads to the last Maxwell equation (8.247). In terms of the vector potential, this gives the wave equation

$$\nabla \times (\nabla \times \mathbf{A}) + \frac{\epsilon(\mathbf{r})}{\epsilon_0 c^2}\partial_t^2 \mathbf{A} = 0. \tag{8.310}$$
(dielectric wave equation)

The first Maxwell equation ($\nabla \cdot \mathbf{D} = 0$) follows most naturally from the generalized Coulomb-gauge condition

$$\nabla \cdot [\epsilon(\mathbf{r})\mathbf{A}] = 0, \tag{8.311}$$
(generalized Coulomb gauge)

although the Maxwell equation is gauge-invariant. More fundamentally, this transverse-field condition arises in the form

$$\nabla \cdot \mathbf{\Pi} = -\nabla \cdot \mathbf{D} = 0 \tag{8.312}$$
(dielectric wave equation)

by introducing the scalar potential ϕ, and considering its Hamilton equation, as in Eq. (8.234). The second Maxwell equation follows as before from the definition $\mathbf{B} = \nabla \times \mathbf{A}$ from Eqs. (8.2), and the third equation follows as before from differentiating this definition and the definition $\mathbf{E} = -\partial_t \mathbf{A}$.

8.9.2.2 Mode Functions

As before, we can separate variables in the wave equation by assuming a harmonic time dependence

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) &= \alpha(t)\mathbf{f}(\mathbf{r}) + \text{c.c.} \\ &= \alpha(0)e^{-i\omega t}\mathbf{f}(\mathbf{r}) + \text{c.c.}, \end{aligned} \tag{8.313}$$

such that the mode functions satisfy the wave equation (8.310) in the form

$$\nabla \times \nabla \times \mathbf{f}(\mathbf{r}) = \frac{\epsilon(\mathbf{r})}{\epsilon_0 c^2}\omega^2 \mathbf{f}(\mathbf{r}), \tag{8.314}$$
(generalized Helmholtz equation)

and the mode functions are constrained by Eq. (8.312) such that

$$\nabla \cdot [\epsilon(\mathbf{r})\mathbf{f}(\mathbf{r})] = 0. \tag{8.315}$$
(mode-function transverse constraint)

We will assume the mode functions to be normalized such that

$$\int d^3r \, \frac{\epsilon(\mathbf{r})}{\epsilon_0} |\mathbf{f}(\mathbf{r})|^2 = 1. \tag{8.316}$$

This is consistent with the orthonormality relation (8.305) for a planar dielectric interface. In the multimode case, note that if we choose

$$\mathbf{f}(\mathbf{r}) = \frac{1}{\sqrt{\epsilon(\mathbf{r})}}\mathbf{g}(\mathbf{r}), \tag{8.317}$$

the wave equation (8.314) becomes

$$\frac{1}{\sqrt{\epsilon(\mathbf{r})}}\nabla\times\nabla\times\frac{1}{\sqrt{\epsilon(\mathbf{r})}}\mathbf{g}(\mathbf{r})=\frac{\omega^2}{\epsilon_0 c^2}\mathbf{g}(\mathbf{r}), \tag{8.318}$$

in which case $g(\mathbf{r})$ is an eigenfunction of a Hermitian operator. Then we can assume the full set of solutions $g_{\mathbf{k},\zeta}(\mathbf{r})$ to be orthonormal,

$$\int d^3r\,\mathbf{g}_{\mathbf{k},\zeta}(\mathbf{r})\,\mathbf{g}^*_{\mathbf{k}',\zeta'}(\mathbf{r})=\delta_{\zeta\zeta'}\delta^3_{\mathbf{k}\mathbf{k}'}, \tag{8.319}$$

in which case we have the orthonormality relation

$$\int d^3r\,\frac{\epsilon(\mathbf{r})}{\epsilon_0}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,\mathbf{f}^*_{\mathbf{k}',\zeta'}(\mathbf{r})=\delta_{\zeta\zeta'}\delta^3_{\mathbf{k}\mathbf{k}'},$$

(orthonormality of mode functions) (8.320)

which again is consistent with the special case (8.305) when the $\mathbf{k}$ sum is compared to an integral.

8.9.2.3 Quantized Fields

In the more general separation of variables, we can write the (classical) multimode field as

$$\mathbf{A}(\mathbf{r},t)=\frac{1}{\sqrt{\epsilon_0}}\sum_{\mathbf{k},\zeta}q_{\mathbf{k},\zeta}(t)\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}), \tag{8.321}$$

where the $q_{\mathbf{k},\zeta}(t)=q_{\mathbf{k},\zeta}(0)\,e^{-i\omega_{\mathbf{k}}t}$ express the time dependence and weight of each mode in the sum. The second term in the Hamiltonian (8.309) then reduces to

$$\begin{aligned}\frac{1}{2\mu_0}\int d^3r\,(\nabla\times\mathbf{A})^2&=\frac{1}{2\mu_0}\int d^3r\,\mathbf{A}\cdot\nabla\times\nabla\times\mathbf{A}\\&=\frac{1}{2\mu_0\epsilon_0}\sum_{\mathbf{k},\zeta}\sum_{\mathbf{k}',\zeta'}q_{\mathbf{k},\zeta}q^*_{\mathbf{k}',\zeta'}\int d^3r\,\mathbf{f}_{\mathbf{k},\zeta}\cdot\nabla\times\nabla\times\mathbf{f}^*_{\mathbf{k}',\zeta'}\\&=\frac{1}{2\mu_0\epsilon_0}\sum_{\mathbf{k},\zeta}\sum_{\mathbf{k}',\zeta'}q_{\mathbf{k},\zeta}q^*_{\mathbf{k}',\zeta'}\int d^3r\,\frac{\epsilon}{\epsilon_0 c^2}\omega^2_{\mathbf{k}'}\mathbf{f}_{\mathbf{k},\zeta}\mathbf{f}^*_{\mathbf{k}',\zeta'}\\&=\frac{1}{2}\sum_{\mathbf{k},\zeta}\sum_{\mathbf{k}',\zeta'}\omega^2_{\mathbf{k}'}q_{\mathbf{k},\zeta}q^*_{\mathbf{k}',\zeta'}\int d^3r\,\frac{\epsilon}{\epsilon_0}\mathbf{f}_{\mathbf{k},\zeta}\mathbf{f}^*_{\mathbf{k}',\zeta'}\\&=\frac{1}{2}\omega^2_{\mathbf{k}}q^2_{\mathbf{k},\zeta},\end{aligned} \tag{8.322}$$

where we used Eqs. (8.314) and (8.320). Similarly, we can write the conjugate field as a superposition of mode functions,

$$\mathbf{\Pi}(\mathbf{r},t)=\frac{\epsilon(\mathbf{r})}{\sqrt{\epsilon_0}}\sum_{\mathbf{k},\zeta}p_{\mathbf{k},\zeta}(t)\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}), \tag{8.323}$$

where the $p_{\mathbf{k},\zeta}(t)=p_{\mathbf{k},\zeta}(0)\,e^{-i\omega_{\mathbf{k}}t}$ similarly express the time dependence of the mode components, and the $\epsilon(\mathbf{r})$ here is necessary to satisfy the constraint (8.312). The first term in the Hamiltonian (8.309) then reduces to

$$\begin{aligned}\int d^3r\frac{\Pi^2}{2\epsilon(\mathbf{r})}&=\frac{1}{2}\sum_{\mathbf{k},\zeta}\sum_{\mathbf{k}',\zeta'}p_{\mathbf{k},\zeta}p^*_{\mathbf{k}',\zeta'}\int d^3r\,\frac{\epsilon}{\epsilon_0}\mathbf{f}_{\mathbf{k},\zeta}\mathbf{f}^*_{\mathbf{k}',\zeta'}\\&=\frac{1}{2}p^2_{\mathbf{k},\zeta}.\end{aligned} \tag{8.324}$$

Thus, the total electromagnetic Hamiltonian becomes

$$H=\sum_{\mathbf{k},\zeta}\left(\frac{p^2_{\mathbf{k},\zeta}}{2}+\frac{1}{2}\omega^2_{\mathbf{k}}q^2_{\mathbf{k},\zeta}\right), \tag{8.325}$$

which has the form of a set of uncoupled (complex) harmonic oscillators of unit mass and frequencies $\omega_{\mathbf{k}}$. Then quantization of the field amounts to promoting the canonical normal coordinates to operators, with the usual relations (for unit mass):

$$\begin{aligned} q_{\mathbf{k},\zeta} &= \sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}}\left(a_{\mathbf{k},\zeta} + a^\dagger_{\mathbf{k},\zeta}\right) \\ p_{\mathbf{k},\zeta} &= \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2}}\left(\frac{a_{\mathbf{k},\zeta} - a^\dagger_{\mathbf{k},\zeta}}{i}\right). \end{aligned} \tag{8.326}$$

With these operator forms, the conjugate fields (8.321) and (8.323) become

$$\begin{aligned} \mathbf{A}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\left(a_{\mathbf{k},\zeta} + a^\dagger_{\mathbf{k},\zeta}\right) \\ \mathbf{\Pi}(\mathbf{r},t) &= -i\epsilon(\mathbf{r})\sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\left(a_{\mathbf{k},\zeta} - a^\dagger_{\mathbf{k},\zeta}\right), \end{aligned} \tag{8.327}$$

Since the fields are real, but the mode functions in general are not, any complex mode must be accompanied by its complex conjugate as another mode. Thus, by rearranging terms among the modes, we can alternately write our fields as

$$\begin{aligned} \mathbf{A}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,a_{\mathbf{k},\zeta} + \text{H.c.} \\ \mathbf{\Pi}(\mathbf{r},t) &= \epsilon(\mathbf{r})\sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,a_{\mathbf{k},\zeta} + \text{H.c.} \end{aligned} \tag{8.328}$$

We have also shifted our phase convention by letting $a_{\mathbf{k},\zeta} \longrightarrow i a_{\mathbf{k},\zeta}$. With these changes, the $\mathbf{A}$ field has the same form as in free space, Eq. (8.61), and the momentum field is consistent with the electric field from before via Eq. (8.308). Thus, the quantized fields in the presence of an inhomogeneous dielectric, have the same form as before,

$$\begin{aligned} \mathbf{A}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\ \mathbf{E}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} -\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,a_{\mathbf{k},\zeta}(t) + \text{H.c.} \\ \mathbf{B}(\mathbf{r},t) &= \sum_{\mathbf{k},\zeta} i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\,[\nabla\times\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})]\,a_{\mathbf{k},\zeta}(t) + \text{H.c.}, \end{aligned} \tag{8.329}$$

(quantized fields, with dielectric)

except that the mode functions are modified by the presence of dielectric, as we saw in the dielectric-interface example of Section 8.9.1.

8.9.3 Transverse and Longitudinal Fields: Dielectric

In the free-space case, it was useful to introduce the concept of transverse and longitudinal fields, for example, to represent commutators of the quantum fields. It is still useful to do this with an inhomogeneous dielectric, but the definitions must be modified somewhat.

8.9.3.1 Scalar Green Function

First, a small digression to recall a tool we will need to generalize the notation for the transverse and longitudinal projection operators. From the Maxwell equation $\nabla\cdot\mathbf{E} = \rho/\epsilon_0$ and the potential relation

$\mathbf{E} = -\partial_t \mathbf{A} - \nabla\phi = -\nabla\phi$ for a static field, we have the Poisson equation

$$\nabla^2 \phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon_0}. \tag{8.330}$$

For a point charge q at $\mathbf{r}_0$,

$$\rho(\mathbf{r}) = q\delta^3(\mathbf{r} - \mathbf{r}_0), \tag{8.331}$$

the well-known potential in free space is

$$\phi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \int d^3r' \frac{1}{|\mathbf{r} - \mathbf{r}_0|}. \tag{8.332}$$

The scalar Green function $G_0(\mathbf{r}, \mathbf{r}')$ is the solution to the Poisson equation (8.330) when the charge density $\rho(\mathbf{r})/\epsilon_0$ is replaced by a unit point charge at $\mathbf{r}'$:

$$\nabla^2 G_0(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}').$$

(scalar Green function, defining equation) (8.333)

By comparing this expression to the solution (8.332), we can identify the free-space scalar Green function

$$G_0(\mathbf{r}, \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}. \tag{8.334}$$

(scalar Green function, free space)

Multiplying through by $\rho(\mathbf{r}')$ and integrating in Eq. (8.333), and then comparing to Eq. (8.330) shows that the solution for a general charge density is

$$\phi(\mathbf{r}) = \frac{1}{\epsilon_0} \int d^3r'\, G_0(\mathbf{r}, \mathbf{r}')\, \rho(\mathbf{r}').$$

(solution in terms of Green function) (8.335)

Note in particular that we recover the solution (8.332) in the case of a point charge (8.331).

8.9.3.2 Transverse and Longitudinal Fields: Free Space

Recall [Section 8.5] that according to the Helmholtz theorem, any vector field can be decomposed into transverse and longitudinal parts,

$$\mathbf{C}(\mathbf{r}) = \mathbf{C}^{\perp}(\mathbf{r}) + \mathbf{C}^{\parallel}(\mathbf{r}), \tag{8.336}$$

where $\nabla \cdot \mathbf{C}^{\perp}(\mathbf{r}) = 0$ and $\nabla \times \mathbf{C}^{\parallel}(\mathbf{r}) = 0$. As an alternate way to construct these components, consider that according to these properties,

$$\nabla \cdot \mathbf{C}(\mathbf{r}) = \nabla \cdot \mathbf{C}^{\parallel}(\mathbf{r}), \tag{8.337}$$

since the transverse part vanishes under the divergence operator. We can formally invert this equation as

$$\mathbf{C}^{\parallel}(\mathbf{r}) = \nabla^{-1} \nabla \cdot \mathbf{C}(\mathbf{r}), \tag{8.338}$$

(longitudinal projection operator)

where the vector integral operator ∇^{-1} is defined such that this reproduces Eq. (8.337) under the divergence operator. Hence, we see that $\nabla^{-1}\nabla$ is a representation of the longitudinal projection operator. To give meaning to the ∇^{-1} operator, note that we can similarly rewrite Eq. (8.333) as

$$\nabla^{-1} \delta(\mathbf{r} - \mathbf{r}') = -\nabla G_0(\mathbf{r}, \mathbf{r}'). \tag{8.339}$$

If we multiply through by a scalar field $f(\mathbf{r}')$ and integrate with respect to $\mathbf{r}'$,

$$\nabla^{-1} f(\mathbf{r}) = -\nabla \int d^3r'\, G_0(\mathbf{r}, \mathbf{r}')\, f(\mathbf{r}'), \tag{8.340}$$

(integral representation of ∇^{-1})

which serves to explicitly define the action of ∇^{-1}. Now suppose instead that we multiply through Eq. (8.339) by $\nabla \cdot \mathbf{C}(\mathbf{r}')$, and then integrate over $\mathbf{r}'$:

$$\nabla^{-1}[\nabla \cdot \mathbf{C}(\mathbf{r})] = -\int d^3r' \, [\nabla G_0(\mathbf{r}, \mathbf{r}')][\nabla' \cdot \mathbf{C}(\mathbf{r}')] = -\nabla \int d^3r' \, G_0(\mathbf{r}, \mathbf{r}')\nabla' \cdot \mathbf{C}(\mathbf{r}'). \tag{8.341}$$

Using Eq. (8.338) on the left-hand side and Eq. (8.334) on the right-hand side, we obtain

$$\mathbf{C}^{\parallel}(\mathbf{r}) = -\nabla \int d^3r' \, \frac{\nabla' \cdot \mathbf{C}(\mathbf{r}')}{4\pi|\mathbf{r} - \mathbf{r}'|}. \tag{8.342}$$

This is the integral representation of the longitudinal field that we derived before in Eq. (8.156).

Because we have the decomposition (8.336), we can then write the transverse field as

$$\mathbf{C}^{\perp}(\mathbf{r}) = \mathbf{C}(\mathbf{r}) - \nabla^{-1}\nabla \cdot \mathbf{C}(\mathbf{r}), \tag{8.343}$$

or

$$\mathbf{C}^{\perp}(\mathbf{r}) = \left(1 - \nabla^{-1}\nabla\right) \cdot \mathbf{C}(\mathbf{r}), \tag{8.344}$$

(transverse projection operator)

where here the "1" is a dyadic identity operator. Now since we have the defining relations [Eqs. (8.169) and (8.170)]

$$\begin{aligned} C^{\perp}_{\alpha}(\mathbf{r}) &= \int d^3r' \delta^{\perp}_{\alpha\beta}(\mathbf{r} - \mathbf{r}') \, C_{\beta}(\mathbf{r}') \\ C^{\parallel}_{\alpha}(\mathbf{r}) &= \int d^3r' \delta^{\parallel}_{\alpha\beta}(\mathbf{r} - \mathbf{r}') \, C_{\beta}(\mathbf{r}'), \end{aligned} \tag{8.345}$$

for the transverse and longitudinal δ functions, we can put Eq. (8.341) into index notation to obtain

$$\begin{aligned} C^{\parallel}_{\alpha}(\mathbf{r}) &= -\partial_{\alpha} \int d^3r' \, G_0(\mathbf{r}, \mathbf{r}')\partial'_{\beta} C_{\beta}(\mathbf{r}') \\ &= \int d^3r' \, [\partial_{\alpha}\partial'_{\beta} G_0(\mathbf{r}, \mathbf{r}')] C_{\beta}(\mathbf{r}') \end{aligned} \tag{8.346}$$

and compare the result to the defining relation to yield

$$\delta^{\parallel}_{\alpha\beta}(\mathbf{r} - \mathbf{r}') = \partial_{\alpha}\partial'_{\beta} G_0(\mathbf{r}, \mathbf{r}'). \tag{8.347}$$

(longitudinal δ function)

The derivatives can then be expanded to give the longitudinal δ function in the position representation [cf. Eqs. (8.182) and (8.184). Similarly, for the transverse projector, we have

$$\delta^{\perp}_{\alpha\beta}(\mathbf{r} - \mathbf{r}') = \delta_{\alpha\beta}\delta(\mathbf{r} - \mathbf{r}') - \partial_{\alpha}\partial'_{\beta} G_0(\mathbf{r}, \mathbf{r}') \tag{8.348}$$

(transverse δ function)

in terms of the free-space scalar Green function.

8.9.3.3 Scalar Green Function: Dielectric Case

In the case of an inhomogeneous dielectric $\epsilon(\mathbf{r})$, we have the Maxwell equation $\nabla \cdot \mathbf{D} = \nabla \cdot \epsilon\mathbf{E} = \rho$, which leads to the generalization

$$(\nabla \cdot \epsilon\nabla)\phi(\mathbf{r}) = -\rho(\mathbf{r}) \tag{8.349}$$

of the Poisson equation. In analogy with the free-space case, we can then define the inhomogeneous scalar Green function by replacing the charge with a δ function:

$$(\nabla \cdot \epsilon\nabla)G(\mathbf{r}, \mathbf{r}') = -\epsilon_0 \delta(\mathbf{r} - \mathbf{r}'), \tag{8.350}$$

(scalar Green function, defining equation)

where we have maintained the same dimensions as in the free-space case in this replacement. For simple dielectric geometries, the solution can be evaluated for example by the method of images. However, this equation cannot be solved analytically for arbitrary $\epsilon(\mathbf{r})$. In any case, the solution for a general charge distribution is still given by Eq. (8.335).

8.9.3.4 Dielectric Transverse and Longitudinal Fields: Green-Function Form

Even in the presence of an inhomogeneous dielectric $\epsilon(\mathrm{r})$, we can decompose a field into transverse and longitudinal parts as in Section 8.9.3.2, since in general the field has nothing to do with the dielectric. However, there is also another sense in which we can decompose fields when an inhomogeneous measure $\epsilon(\mathrm{r})$ is available. For example, in the absence of charge, Maxwell's equations state that $\nabla\cdot\mathbf{D}=0$, so that $\mathbf{D}=\mathbf{D}^{\perp}$ is purely transverse. But in this case, $\mathbf{E}=\mathbf{D}/\epsilon$ is *not* purely transverse. Then what is the decomposition of $\mathbf{E}$ that gives rise to the transverse and longitudinal parts of $\mathbf{D}$? Another example is in the choice of a generalized Coulomb gauge $\nabla\cdot\epsilon\mathbf{A}=0$. In this case, $\nabla\cdot\mathbf{A}\neq 0$ in general. Then what is the decomposition of $\mathbf{A}$ into components, such that one of them vanishes in the sense of $\nabla\cdot\epsilon\mathbf{A}=0$?

Sticking to the gauge example, suppose that we make the (unique) decomposition

$$\mathbf{A}(\mathbf{r})=\mathbf{A}^{\epsilon\perp}(\mathbf{r})+\mathbf{A}^{\epsilon\parallel}(\mathbf{r}), \tag{8.351}$$
(ϵ-Helmholtz decomposition)

such that $\nabla\cdot\epsilon\mathbf{A}^{\epsilon\perp}=0$ and $\nabla\times\epsilon\mathbf{A}^{\epsilon\parallel}=0$. Our goal here is to find the projection operators corresponding to these ''ϵ-transverse'' and ''ϵ-longitudinal'' components. We can begin by noting that to make this decomposition, we can simply decompose $\epsilon\mathbf{A}$ in the usual way, so that our constraints are clearly satisfied by choosing

$$\mathbf{A}^{\epsilon\perp}=\epsilon^{-1}[\epsilon\mathbf{A}]^{\perp},\qquad \mathbf{A}^{\epsilon\parallel}=\epsilon^{-1}[\epsilon\mathbf{A}]^{\parallel}. \tag{8.352}$$

Taking the longitudinal case first, we can use Eq. (8.338) to obtain

$$\begin{aligned}\mathbf{A}^{\epsilon\parallel}(\mathbf{r})&=\epsilon^{-1}\nabla^{-1}\nabla\cdot[\epsilon\mathbf{A}(\mathbf{r})]\\&=-\epsilon^{-1}\nabla\int d^3r'\,G_0(\mathbf{r},\mathbf{r}')\,\nabla'\cdot[\epsilon\mathbf{A}(\mathbf{r}')],\end{aligned} \tag{8.353}$$

where we used Eq. (8.340) for ∇^{-1}. Now multiply through by $\epsilon(\mathbf{r})$ and compute the divergence. Note that we lose no information in doing so, since the vector fields are longitudinal on both sides of the equality:

$$\nabla\cdot[\epsilon\mathbf{A}^{\epsilon\parallel}(\mathbf{r})]=-\nabla^2\int d^3r'\,G_0(\mathbf{r},\mathbf{r}')\,\nabla'\cdot[\epsilon\mathbf{A}(\mathbf{r}')]. \tag{8.354}$$

Then we can combine Eqs. (8.333) and (8.350) to obtain

$$\epsilon_0\nabla^2G_0(\mathbf{r},\mathbf{r}')=\nabla\cdot\epsilon\nabla G(\mathbf{r},\mathbf{r}'), \tag{8.355}$$

so that

$$\nabla\cdot[\epsilon\mathbf{A}^{\epsilon\parallel}(\mathbf{r})]=-\epsilon_0^{-1}\nabla\cdot\epsilon\nabla\int d^3r'\,G(\mathbf{r},\mathbf{r}')\,\nabla'\cdot[\epsilon\mathbf{A}(\mathbf{r}')]. \tag{8.356}$$

Undoing the overall divergence operator (as we justified above), and dividing through by $\epsilon(\mathbf{r})$ gives

$$\mathbf{A}^{\epsilon\parallel}(\mathbf{r})=-\epsilon_0^{-1}\nabla\int d^3r'\,G(\mathbf{r},\mathbf{r}')\,\nabla'\cdot[\epsilon\mathbf{A}(\mathbf{r}')]. \tag{8.357}$$

Writing this out in components,

$$\begin{aligned}A_\alpha^{\epsilon\parallel}(\mathbf{r})&=-\epsilon_0^{-1}\partial_\alpha\int d^3r'\,G(\mathbf{r},\mathbf{r}')\,\partial'_\beta[\epsilon A_\beta(\mathbf{r}')]\\&=\epsilon_0^{-1}\int d^3r'\,\epsilon(\mathbf{r}')\,[\partial_\alpha\partial'_\beta G(\mathbf{r},\mathbf{r}')]\,A_\beta(\mathbf{r}')\end{aligned} \tag{8.358}$$

Therefore, we can define the ϵ-longitudinal projection

$$A_\alpha^{\epsilon\parallel}(\mathbf{r})=\int d^3r'\,\frac{\epsilon(\mathbf{r}')}{\epsilon_0}\,\delta^{\epsilon\parallel}(\mathbf{r},\mathbf{r}')\,A_\beta(\mathbf{r}'), \tag{8.359}$$
(ϵ-longitudinal projection)

where

$$\delta^{\epsilon\parallel}_{\alpha\beta}(\mathbf{r},\mathbf{r}') := \partial_\alpha \partial'_\beta G(\mathbf{r},\mathbf{r}'). \tag{8.360}$$
(ϵ-longitudinal δ function)

Note that this projector is not symmetric, $\delta^{\epsilon\parallel}(\mathbf{r},\mathbf{r}') \neq \delta^{\epsilon\parallel}(\mathbf{r}',\mathbf{r})$, nor is it "relocatable" in the sense $\delta^{\epsilon\parallel}(\mathbf{r},\mathbf{r}') \neq \delta^{\epsilon\parallel}(\mathbf{r}-\mathbf{r}')$. Note also that the projection and the δ function reduce to their free-space counterparts when $\epsilon \longrightarrow \epsilon_0$.

Similarly, the ϵ-transverse projector is given by subtracting the ϵ-longitudinal projector from the identity,

$$\delta^{\epsilon\perp}_{\alpha\beta}(\mathbf{r},\mathbf{r}') := \frac{\epsilon_0}{\epsilon(\mathbf{r})}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}') - \partial_\alpha\partial'_\beta G(\mathbf{r},\mathbf{r}'). \tag{8.361}$$
(ϵ-transverse δ function)

including the correct measure on the identity such that in the projection integral, we have[19]

$$A^{\epsilon\perp}_\alpha(\mathbf{r}) = \int d^3r' \, \frac{\epsilon(\mathbf{r}')}{\epsilon_0}\, \delta^{\epsilon\perp}_{\alpha\beta}(\mathbf{r},\mathbf{r}')\, A_\beta(\mathbf{r}'). \tag{8.362}$$
(ϵ-transverse projection)

We may verify the proper action of the ϵ-transverse projector simply as

$$\begin{aligned}\int d^3r \, \frac{\epsilon(\mathbf{r}')}{\epsilon_0}\, \delta^{\epsilon\perp}_{\alpha\beta}(\mathbf{r},\mathbf{r}')\, A_\beta(\mathbf{r}') &= \int d^3r' \, \frac{\epsilon(\mathbf{r}')}{\epsilon_0}\left[\frac{\epsilon_0}{\epsilon(\mathbf{r})}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}') - \partial_\alpha\partial'_\beta G(\mathbf{r},\mathbf{r}')\right] A_\beta(\mathbf{r}') \\ &= A_\alpha(\mathbf{r}) - A^{\epsilon\parallel}_\alpha(\mathbf{r}) \\ &= A^{\epsilon\perp}_\alpha(\mathbf{r}),\end{aligned} \tag{8.363}$$

using the ϵ-decomposition (8.351).

8.9.3.5 Completeness

The other route to representing the projection operators[20] is to recall that the functions $\mathbf{g}(\mathbf{r})$ are eigenfunctions of a Hermitian operator in Eq. (8.318). We can then choose a set of orthonormal functions via (8.319), and in doing so, we can sum over elements of the form $g_{\mathbf{k},\zeta,\alpha}(\mathbf{r})\, g^*_{\mathbf{k},\zeta,\beta}$ to represent the identity operator. Normally, for a set of complete, vector-valued functions on $\mathbb{R}^3$, the identity operator would be $\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}')$. However, the functions $g_{\mathbf{k},\zeta}(\mathbf{r})$ satisfy the constraint $\nabla\cdot[\sqrt{\epsilon}g(\mathbf{r})] = 0$, and so these elements only form an identity on this restricted subspace. Transforming back to the $\mathbf{f}(\mathbf{r})$ functions, we can therefore write the identity as

$$\delta^{\epsilon\perp}_{\alpha\beta}(\mathbf{r}-\mathbf{r}') = \sum_{\mathbf{k},\zeta} f_{\mathbf{k},\zeta,\alpha}(\mathbf{r})\, f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}'), \tag{8.364}$$
(completeness relation)

which is now an identity on the space of functions satisfying $\nabla\cdot[\epsilon f(\mathbf{r})] = 0$, which is exactly how we defined this projection operator, via the decomposition (8.351).

[19] For an in-depth discussion of the projector in for a half-space dielectric, see Robert Zietal, *Quantum Electrodynamics Near Material Boundaries*, Ph.D. thesis, University of Sussex (2010), Chapter 4.

[20] Roy J. Glauber and M. Lewenstein, *op. cit.*

8.10 Exercises

Problem 8.1

Geometric optics, or ray optics, can be formulated in terms of the action principle (Fermat's principle):

$$\delta \int n(x,y,z)\, ds = 0, \tag{8.365}$$

where $ds^2 = dx^2 + dy^2 + dz^2$ and $n(x,y,z)$ is the index-of-refraction profile that models the optical system. Take the coordinate z to be the "time" variable and the coordinate y to be the position coordinate. Let's simplify things and consider only the two-dimensional case, so x is an ignorable coordinate, and note that z is also ignorable in the sense of being completely determined by x, y, and s.

(a) Draw an analogy to classical mechanics, and write down the ray-optics Lagrangian. Then show that the conjugate momentum p for the generalized coordinate y is $n\,dy/ds$. Finally, write down the ray-optics Hamiltonian, which you should write in terms of the canonical coordinates p and y, but *not* y'.

(b) Make the paraxial approximation (small p, small y), and assume that the refractive index may be written as a small variation on a large baseline, $n = n_0 + \delta n$, where $\delta n/n_0 \ll 1$. Keep only lowest-order terms in p, y, and δn (dropping higher-order cross-terms in these variables), and show that the Hamiltonian takes the form of a classical particle Hamiltonian, with effective mass n_0 and potential $-\delta y$.

Problem 8.2

The usual Euler–Lagrange equation,

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0, \tag{8.366}$$

applies to Lagrangians of the form $L(q,\dot{q};t)$.

(a) Generalize the Euler–Lagrange equation to handle Lagrangians of the form $L(q,\dot{q},\ddot{q},\dddot{q};t)$. Indicate any conditions you impose on the endpoints of the variation.

(b) One might hope to write down a Lagrangian for the Abraham–Lorentz force ($\mathbf{F} \propto \dddot{\mathbf{x}}$) by considering Lagrangians of the form

$$L = \frac{1}{2}m\dot{q}^2 - V(q) + \beta q \dddot{q}. \tag{8.367}$$

Use your result from part (a) to write down the equation of motion for this Lagrangian.

Problem 8.3

Consider a thin string of linear mass density μ, stretched under tension T_0 nearly along the x-axis between positions x_1 and x_2. Consider only small (i.e., linear) vibrations of this string, so that the wave function $y(x,t)$ is always much smaller than the string's length.

(a) Derive the following Lagrangian for the string:

$$L(y, y_t, y_x) = \frac{1}{2}\int_{x_1}^{x_2} \left[\mu y_t^2 - T_0 y_x^2\right] dx. \tag{8.368}$$

To do this, take a small segment of string of length $d\ell$, and compute its kinetic energy, integrating the result to get the total kinetic energy. Then compute the potential energy by computing the length of the string, and then considering what this means in terms of the energy. Assume that both y and y_x are small, and ignore any longitudinal motion of the segment.

(b) Noting that the Lagrangian has the form

$$L(y, y_t, y_x) = \int_{x_1}^{x_2} f(y, y_t, y_x)\, dx, \tag{8.369}$$

derive a suitably generalized Euler–Lagrange equation for the action principle $\delta S = 0$ in terms of the integrand f. Then use your result to derive the wave equation for the string.

Problem 8.4

Compute the functional derivative of

$$S = \int d^3r \int dt \left[i\hbar\phi^*\phi_t - \frac{\hbar^2}{2m}\nabla\phi^*\nabla\phi - \phi^* V(x,t)\phi \right], \tag{8.370}$$

with respect to the fields $\phi^*(\mathbf{r},t)$ and $\phi(\mathbf{r},t)$, ignoring surface terms. What are the equations of motion obtained from the action principles $\delta S/\delta\phi = 0$ and $\delta S/\delta\phi^* = 0$? What is the canonically conjugate momentum field to ϕ? To ϕ^*?

Problem 8.5

In each of the following, you may ignore surface terms.

(a) Compute the functional derivative of

$$S = \int dt\, f(t)\, f^{(201)}(t), \tag{8.371}$$

where $f^{(n)}(t)$ is the nth derivative of $f(t)$.

(b) Compute the functional derivative of

$$S = \int dt\, f^{(199)}(t)\, f^{(201)}(t). \tag{8.372}$$

Problem 8.6

Consider the functional

$$F[a,b] = \int_{-1}^{1} dx\, f(a,b), \tag{8.373}$$

where

$$f(a,b) = \begin{cases} 0 & \text{if } a = b = 0 \\ \dfrac{ab^2}{a^2+b^2} & \text{otherwise} \end{cases} \tag{8.374}$$

for functions $a(x)$ and $b(x)$. Calculate the variation $\delta F[a=0, b=0; \delta a, \delta b]$. What is the pathology of this functional?[21]

Problem 8.7

(a) Show that $\delta^\perp_{\alpha\beta}(\mathbf{r}) = \delta^\perp_{\beta\alpha}(\mathbf{r})$.

(b) Show that $\delta^\perp_{\alpha\beta}(-\mathbf{r}) = \delta^\perp_{\alpha\beta}(\mathbf{r})$.

(c) Show that $\partial_\alpha \delta^\perp_{\alpha\beta}(\mathbf{r}) = \partial_\beta \delta^\perp_{\alpha\beta}(\mathbf{r}) = 0$.

Problem 8.8

The electric field due to an oscillating dipole at frequency $\omega = ck$ has the somewhat messy form

$$\mathbf{E}^{(+)}(\mathbf{r},\omega) = \frac{1}{4\pi\epsilon_0}\left\{ [3(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}]\left[\frac{1}{r^3} - i\frac{k}{r^2}\right] - [(\hat{\varepsilon}\cdot\hat{r})\hat{r} - \hat{\varepsilon}]\,\frac{k^2}{r}\right\} d^{(+)}(\omega) e^{ikr}, \tag{8.375}$$

where $\hat{r}$ is a unit vector in the $\mathbf{r}$ direction, and $\hat{\varepsilon}$ is a unit vector marking the dipole orientation. (The plus superscripts here indicate an implied time dependence of $e^{-i\omega t}$.)

[21] This problem stolen from P. J. Morrison, "Hamiltonian description of the ideal fluid," *Reviews of Modern Physics* **70**, 467 (1998) (doi: 10.1103/RevModPhys.70.467).

A naïve guess for the radiation field is to adapt the *scalar* spherical wave e^{ikr}/r, tacking on the orientation and magnitude of the dipole moment to make a vector field (we can also tack on a factor of $4\pi\epsilon_0$ for good measure):

$$\mathbf{E}^{(+)}_{\text{guess}}(\mathbf{r},\omega) = \frac{1}{4\pi\epsilon_0}\mathbf{d}^{(+)}(\omega)\frac{e^{ikr}}{r}. \tag{8.376}$$

Obviously this is wrong. However, show that the correct dipole field arises by using the transverse projector $(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta)$. That is, show that

$$E^{(+)}_\alpha = (k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta)E^{(+)}_{\text{guess},\beta}. \tag{8.377}$$

Note that we associate this operator with transverse projection by examining the transverse delta function:

$$\delta^\perp_{\alpha\beta}(\mathbf{r}) = \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}} = \frac{1}{(2\pi)^3}\int d^3k\,\frac{1}{k^2}\left(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\right)e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{8.378}$$

Thus, up to a factor of $1/k^2$ the transverse delta function is a Fourier transform of the projection operator $(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta)$.

Problem 8.9

Following the steps in the notes, show that the transverse and longitudinal delta functions,

$$\begin{aligned}\delta^\perp_{\alpha\beta}(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}}\\ \delta^\parallel_{\alpha\beta}(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\,\frac{k_\alpha k_\beta}{k^2}e^{i\mathbf{k}\cdot\mathbf{r}},\end{aligned} \tag{8.379}$$

can be expressed as

$$\begin{aligned}\delta^\perp_{\alpha\beta}(\mathbf{r}) &= \frac{2}{3}\delta_{\alpha\beta}\delta^3(\mathbf{r}) - \frac{1}{4\pi r^3}\left(\delta_{\alpha\beta} - \frac{3r_\alpha r_\beta}{r^2}\right)\\ \delta^\parallel_{\alpha\beta}(\mathbf{r}) &= \frac{1}{3}\delta_{\alpha\beta}\delta^3(\mathbf{r}) + \frac{1}{4\pi r^3}\left(\delta_{\alpha\beta} - \frac{3r_\alpha r_\beta}{r^2}\right)\end{aligned} \tag{8.380}$$

in the position representation.

Problem 8.10

Starting with the general relation

$$[A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t')] = -\frac{i\hbar}{\epsilon_0}\sum_{\mathbf{k},\zeta}\text{Re}\left[f_{\mathbf{k},\zeta,\alpha}(\mathbf{r})f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}')e^{-i\omega_\mathbf{k}(t-t')}\right], \tag{8.381}$$

follow the notes and derive the following commutator in *half space*:

$$[A_\alpha(\mathbf{r},t), E_\beta(\mathbf{r}',t)] = -\frac{i\hbar}{\epsilon_0}\left[\delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}') - \delta^\top_{\alpha\beta}(\mathbf{r}^- - \mathbf{r}')\right]. \tag{8.382}$$

Again, the interpretation is that the vector potential and electric field are ''connected'' at two spacetime points if they lie on the same light cone with respect to paths that are either direct or have one bounce off the mirror.

Problem 8.11

Show by using the momentum-space representation of the transverse delta function that

$$\int d^3r\,\delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}')\delta^\perp_{\beta\gamma}(\mathbf{r}) = \delta^\perp_{\alpha\gamma}(\mathbf{r}'). \tag{8.383}$$

Problem 8.12

The *quantum* coherence functions $g^{(n)}(\tau)$ are defined in terms of the quantum fields in the same way we defined them for the classical counterparts, except that the classical time average is replaced by an expectation value with respect to the state of the field.

(a) Derive expressions for $g^{(1)}(\tau)$ and $g^{(2)}(\tau)$ for a single mode of the electromagnetic field.

(b) Evaluate your expressions from part (a) for a field in a number state (**Fock state**) $|n\rangle$. Show that the first-order coherence can be interpreted classically, but that this state is nonclassical at the second order of coherence.

Problem 8.13

Consider the **squeezed vacuum state**

$$|\zeta\rangle = S(\zeta)|0\rangle, \tag{8.384}$$

of the harmonic oscillator (or a single field mode), where $|0\rangle$ is the vacuum state, and $S(\zeta)$ is the squeezing operator[22]

$$S(\zeta) := \exp\left[\frac{1}{2}\left(\zeta^* a^2 - \zeta a^{\dagger 2}\right)\right]. \tag{8.385}$$

Note that the squeezing operator reduces to the identity for $\zeta = 0$. Like the vacuum state, the squeezed vacuum is a minimum-uncertainty Gaussian state, but with a different set of variances (the vacuum state is literally stretched or "squeezed" into a different Gaussian, keeping the uncertainty product constant in some basis).

Compute the initial value of the second-order correlation function

$$g^{(2)}(0) = \frac{\langle a^\dagger a^\dagger a a\rangle}{\langle a^\dagger a\rangle^2}, \tag{8.386}$$

analogous to the one we studied to find antibunching in the resonance fluorescence in the two-level atom. From this result, what can you conclude about the classicality of the squeezed vacuum? Is there anything physically funny about your solution in the limit $\zeta \longrightarrow 0$?

You may use without proof the transformation rules

$$\begin{aligned} S^\dagger(\zeta) a S(\zeta) &= a\cosh r - a^\dagger e^{i\theta}\sinh r \\ S^\dagger(\zeta) a^\dagger S(\zeta) &= a^\dagger \cosh r - a e^{-i\theta}\sinh r \end{aligned} \tag{8.387}$$

for the ladder operators, where $\zeta = re^{i\theta}$.

Problem 8.14

Given the solution $\psi(\mathbf{r})$ to the scalar Helmholtz equation

$$(\nabla^2 + k^2)\psi = 0, \tag{8.388}$$

show that the vector fields

$$\mathbf{L} = \nabla\psi, \qquad \mathbf{M} = \nabla\times(\mathbf{r}\psi), \qquad \mathbf{N} = \frac{1}{k}\nabla\times\mathbf{M} \tag{8.389}$$

satisfy the vector Helmholtz equation

$$(\nabla^2 + k^2)\mathbf{f} = 0. \tag{8.390}$$

[22] See David Stoler, "Equivalence Classes of Minimum Uncertainty Packets," *Physical Review D* **1**, 3217 (1970) (doi: 10.1103/PhysRevD.1.3217); Carlton M. Caves, "Quantum-mechanical noise in an interferometer," *Physical Review D* **23**, 1693 (1981) (doi: 10.1103/PhysRevD.23.1693).

Problem 8.15

In this problem you will work out the normalization factors for the spherical-cavity modes.

(a) Noting that the ordinary Bessel function $J_n(x)$ satisfies **Bessel's equation**

$$\frac{1}{x}\partial_x\left[xJ_n'(x)\right] + \left(1 - \frac{n^2}{x^2}\right)J_n(x) = 0, \tag{8.391}$$

show that

$$\int_0^R dr\, rJ_n^2(kr) = \frac{R^2}{2}\left\{\left[J_n'(kR)\right]^2 - J_n(kR)J_n''(kR) - \frac{1}{kR}J_n(kR)J_n'(kR)\right\}. \tag{8.392}$$

Do this by multiplying Bessel's equation by $rJ_n(k'r)$, then switching $k' \longleftrightarrow k$ and subtracting the resulting equations. Then integrate by parts and let $k' \longrightarrow k$, being careful to keep the lowest-order nonvanishing terms in $k' - k$.

(b) Use the result of part (a) to show that the normalization integral from Eq. (8.98) or (8.113) is

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\, r^2\, j_l^2(k_{nl}r) = \frac{R^3}{2}\left[j_l'(k_{nl}R)\right]^2 = \frac{R^3}{2}j_{l+1}^2(k_{nl}R), \tag{8.393}$$

for modes subject to the radial boundary condition (8.112)

$$j_l(k_{nl}R) = 0, \tag{8.394}$$

as is appropriate for a perfectly conducting spherical cavity for scalar or TE vector waves.

(c) Show that the same normalization integral, subject to the radial boundary condition (8.120)

$$\partial_r\left[r\, j_l(k_{nl}r)\right]\big|_{r=R} = 0, \tag{8.395}$$

as is appropriate for a perfectly conducting spherical cavity for TM vector waves, becomes

$$\mathcal{N}_{nl}^{-2} = \int_0^R dr\, r^2\, j_l^2(k_{nl}r) = \frac{R^3}{2}\left(1 - \frac{l(l+1)}{k_{nl}^2R^2}\right)j_l^2(kR). \tag{8.396}$$

To start, it will help to use Bessel's equation again to eliminate the $J_n''(x)$ in the result from part (a).

Problem 8.16

Work out the Hamiltonian structure of the *massive* electromagnetic field (**Proca** field), paralleling the *massless* treatment of Section 8.8. Use as your Lagrangian

$$\begin{aligned} L &= L_{\text{free}} + L_{\text{source}} + L_{\text{mass}} \\ L_{\text{free}} &= \frac{\epsilon_0}{2}\int d^3r\left[(\nabla\phi + \partial_t\mathbf{A})^2 - c^2(\nabla\times\mathbf{A})^2\right] \\ L_{\text{source}} &= -\int d^3r\left[\rho\phi - \mathbf{j}\cdot\mathbf{A}\right] \\ L_{\text{mass}} &= -\frac{\epsilon_0}{2}\frac{m^2c^4}{\hbar^2}\int d^3r\left[\mathbf{A}^2 - \frac{\phi^2}{c^2}\right], \end{aligned} \tag{8.397}$$

where L_{free} and L_{source} are defined as before, and notice that the particle mass m enters with the proper dimensions as the Compton length $\hbar/mc$. In your treatment you should cover the following:

(a) Derive the canonical momenta and the Hamiltonian.

(b) Write down Hamilton's equations and the generalized Maxwell equations, treating the explicitly any gauge freedom for this field theory that arises from the structure of the Hamiltonian.

(c) Show that charge is only conserved in the Lorenz gauge, and thus that this field theory is not truly gauge-invariant.

(d) Write down the wave equations for the potentials in the Lorenz gauge.

(e) Analyze a longitudinal solution $\mathbf{A} = \hat{z}\alpha \exp[i(kz - \omega t)]$, $\phi = \beta \exp[i(kz - \omega t)]$, where α and β are real amplitudes. Derive the dispersion relation between ω and k. By considering $(\mathbf{A}, \phi)$ to be the components of a four-dimensional vector, show that in the limit $m \longrightarrow 0$ limit, the longitudinal field becomes orthogonal to the source four-vector (you should work out the form of the source field). Thus, in this limit, the longitudinal field decouples from any sources, leaving only the two transverse fields in massless electromagnetism.

(f) Work out the scalar potential for a static point charge q localized at $\mathbf{r} = 0$. *Hint:* start by showing that away from $\mathbf{r} = 0$ that a solution of the form $r\phi = C \exp(-\mu r)$ satisfies the wave equation.

Chapter 9

Atomic Interaction with the Quantized Field

Up till now, we have been using the dipole interaction Hamiltonian $H_{\text{AF}} = -\mathbf{d}\cdot\mathbf{E}$ to describe the coupling of the atom and field. However, this interaction is approximate, being valid only in the dipole approximation for nearly stationary atoms. Here, we will address the fundamental question, what *is* the fundamental Hamiltonian for the atom–field interaction? The answer turns out to have some subtleties, and this subject has historically been the source of substantial confusion.

9.1 Lorentz Force

The classical force on an electron of charge $q = -e$ in an electromagnetic field is given by[1]

$$\mathbf{F} = -e(\mathbf{E} + \mathbf{v}\times\mathbf{B}). \tag{9.1}$$
(Lorentz force law)

Instead of the fields, we can write this in terms of the vector and scalar potentials $\mathbf{A}$ and ϕ, respectively:

$$\mathbf{F} = e[\nabla\phi + \partial_t\mathbf{A} - \mathbf{v}\times(\nabla\times\mathbf{A})]. \tag{9.2}$$

But the vector identity

$$\nabla(\mathbf{v}\cdot\mathbf{A}) = \mathbf{v}\times(\nabla\times\mathbf{A}) + (\mathbf{v}\cdot\nabla)\mathbf{A} \tag{9.3}$$

gives

$$\mathbf{F} = e[\nabla\phi + \partial_t\mathbf{A} + (\mathbf{v}\cdot\nabla)\mathbf{A} - \nabla(\mathbf{v}\cdot\mathbf{A})]. \tag{9.4}$$

The particular combination

$$\frac{d\mathbf{A}(\mathbf{r},t)}{dt} = \partial_t\mathbf{A} + \frac{\partial\mathbf{A}}{\partial x_\alpha}\frac{dx_\alpha}{dt} = \partial_t\mathbf{A} + (\mathbf{v}\cdot\nabla)\mathbf{A} \tag{9.5}$$

is known as the **convective derivative**, and allows us to write

$$\mathbf{F} = e\left[\nabla\phi + \frac{d\mathbf{A}}{dt} - \nabla(\mathbf{v}\cdot\mathbf{A})\right]. \tag{9.6}$$
(Lorentz force law)

Of course, we also make the identification $\mathbf{F} = m\ddot{\mathbf{r}}$.

[1] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice-Hall, 2013), Section 5.1.2, p. 212.

9.1.1 Lagrangian

We note that we can derive this force law from the Lagrangian

$$L = \frac{1}{2}m\dot{\mathbf{r}}^2 - e\dot{\mathbf{r}}\cdot\mathbf{A}(\mathbf{r}) + e\phi(\mathbf{r}). \tag{9.7}$$
(Lorentz-force Lagrangian)

To see this, we simply evaluate the Euler–Lagrange equation

$$\frac{\partial L}{\partial \mathbf{r}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{r}}} = 0, \tag{9.8}$$

which gives

$$-e\nabla(\dot{\mathbf{r}}\cdot\mathbf{A}) + e\nabla\phi - \frac{d}{dt}(m\dot{\mathbf{r}} - e\mathbf{A}) = 0. \tag{9.9}$$

It is easy to see (Problem 9.2) that this reproduces Eq. (9.6), but we can also write this in the form

$$\frac{d}{dt}(m\dot{\mathbf{r}} - e\mathbf{A}) = e\nabla[\phi - (\mathbf{v}\cdot\mathbf{A})], \tag{9.10}$$

which suggests that $m\dot{\mathbf{r}} - e\mathbf{A}$ plays the role of the momentum, while $-e[\phi - (\mathbf{v}\cdot\mathbf{A})]$ plays the role of the (velocity-dependent) potential.

9.1.2 Hamiltonian

We can see that this is indeed the case by deriving the Hamiltonian. Since the potential is velocity-dependent, the canonical momentum involves the vector potential:

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{r}}} = m\dot{\mathbf{r}} - e\mathbf{A} = \mathbf{p}_{\text{kinetic}} - e\mathbf{A}.$$
(Lorentz-force canonical momentum) (9.11)

Here, $\mathbf{p}_{\text{kinetic}} := m\dot{\mathbf{r}}$ is the usual kinetic momentum. Then, with $\dot{\mathbf{r}} = (\mathbf{p}+e\mathbf{A})/m$, in general the Hamiltonian is the Legendre transform of the Lagrangian,

$$H = \mathbf{p}\cdot\dot{\mathbf{r}} - L, \tag{9.12}$$

so that the Hamiltonian for the Lorentz force is

$$H = \frac{(\mathbf{p}+e\mathbf{A})^2}{2m} - e\phi. \tag{9.13}$$
(Lorentz-force Hamiltonian)

Now the magnetic-field ($\nabla\times\mathbf{A}$) and transverse-electric-field ($\partial_t\mathbf{A}$) parts of the interaction is included in the *kinetic* part, while the longitudinal-electric part of the interaction (due to ϕ) is in the potential term.

9.2 Quantization and Minimal Coupling

The total Hamiltonian for a system of particles of charge q_α and mass m_α interacting with the electromagnetic field is then[2]

$$H = \sum_\alpha \frac{[\mathbf{p}_\alpha - q_\alpha\mathbf{A}(\mathbf{r}_\alpha)]^2}{2m_\alpha} + \frac{\epsilon_0}{2}\int d^3r\left(E^2 + c^2B^2\right), \tag{9.14}$$

where we now explicitly include the field Hamiltonian, and for the moment we do not explicitly consider any contribution due to a scalar potential ϕ, since we have in a sense already included it. This is because of the

[2]For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Section 4.2, p. 115.

Helmholtz theorem, which says that the electric field can be decomposed into transverse and longitudinal components as

$$\mathbf{E} = \mathbf{E}^{\perp} + \mathbf{E}^{\parallel}, \tag{9.15}$$

where $\nabla \cdot \mathbf{E}^{\perp} = 0$ and $\nabla \times \mathbf{E}^{\parallel} = 0$. This decomposition is obvious in the Coulomb gauge, since $\mathbf{E}^{\perp} = -\partial_t \mathbf{A}$ and $\mathbf{E}^{\parallel} = -\nabla\phi$. Then using

$$\int d^3r\, \mathbf{E}^{\perp} \cdot \mathbf{E}^{\parallel} = 0, \tag{9.16}$$

we can write the electric-field contribution to the Hamiltonian as

$$\begin{aligned} \int d^3r E^2 &= \int d^3r \left(E^{\perp 2} + E^{\parallel 2} \right) \\ &= \int d^3r\, E^{\perp 2} + \int d^3r\, (\nabla\phi)^2 \\ &= \int d^3r\, E^{\perp 2} - \int d^3r\, \phi \nabla^2 \phi \\ &= \int d^3r\, E^{\perp 2} + \frac{1}{\epsilon_0} \int d^3r\, \rho\phi, \end{aligned} \tag{9.17}$$

where ρ is the source charge density, and we have dropped surface terms.

Suppose now that the source charges come in the form of localized point particles,

$$\rho = \sum_{\alpha=1}^{N} q_\alpha \delta^3(\mathbf{r} - \mathbf{r}_\alpha). \tag{9.18}$$

Then the scalar potential is

$$\phi(\mathbf{r}, t) = \int d^3r' \frac{\rho(\mathbf{r}', t)}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|}. \tag{9.19}$$

So, we can now evaluate the integral in the last term of Eq. (9.17),

$$\int d^3r\, \rho\phi = \int d^3r \int d^3r' \frac{\rho(\mathbf{r}, t)\rho(\mathbf{r}', t)}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{r}'|} = 2 \sum_{\alpha > \beta} \frac{q_\alpha q_\beta}{4\pi\epsilon_0 |\mathbf{r}_\alpha - \mathbf{r}_\beta|}, \tag{9.20}$$

so that we can write the total Hamiltonian (9.14) as

$$H = \sum_{\alpha} \frac{[\mathbf{p}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\alpha)]^2}{2m_\alpha} + \frac{1}{4\pi\epsilon_0} \sum_{\alpha > \beta} \frac{q_\alpha q_\beta}{|\mathbf{r}_\alpha - \mathbf{r}_\beta|} + \frac{\epsilon_0}{2} \int d^3r \left(E^{\perp 2} + c^2 B^2 \right). \tag{9.21}$$

Thus, we see that we can associate the longitudinal field $\mathbf{E}^{\parallel}$ with the fields due to the charged particles.

Now suppose that we take all but one of the particles to be fixed, with the moveable atom an electron of charge $q = -e$, as appropriate for a one-electron atom (or an atom where one electron has the predominant interaction with the field). Then the Hamiltonian becomes

$$H = \frac{[\mathbf{p}_\mathrm{e} + e\mathbf{A}(\mathbf{r}_\mathrm{e})]^2}{2m_\mathrm{e}} - e\phi(\mathbf{r}_\mathrm{e}) + \frac{\epsilon_0}{2} \int d^3r \left(E^{\perp 2} + c^2 B^2 \right), \tag{9.22}$$

where $\mathbf{p}_\mathrm{e}$ and $\mathbf{r}_\mathrm{e}$ are the canonical coordinates of the electron. In an atom, we interpret the potential ϕ due to the other charged particles to give the binding potential $V(\mathbf{r}_\mathrm{e})$. We can also see then that when quantizing the field, it is only necessary to quantize the *transverse* field. To describe the atom–field interaction, we can associate the longitudinal field with the atom itself. This is true of the magnetic field, since the magnetic field is *already* transverse (since there are no magnetic monopoles).

Thus, the quantization of the total Hamiltonian, including the atomic coupling to the electromagnetic field, proceeds as follows. Take the quantized Hamiltonian for the atom and the field, in the uncoupled limit, which we already know:

$$H = \frac{p_{\rm e}^2}{2m_{\rm e}} + V(\mathbf{r}_{\rm e}) + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right). \tag{9.23}$$

Now, to include the atom–field interaction, we make the **minimal-coupling replacement** $\mathbf{p}_{\rm e} \longrightarrow \mathbf{p}_{\rm e} + e\mathbf{A}$ in the above Hamiltonian, to obtain the **minimal-coupling Hamiltonian**

$$H = \frac{[\mathbf{p}_{\rm e} + e\mathbf{A}(\mathbf{r}_{\rm e})]^2}{2m_{\rm e}} + V(\mathbf{r}_{\rm e}) + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right)$$

(minimal-coupling Hamiltonian) (9.24)

describing the coupled atom-field system within quantum electrodynamics.

9.3 Dipole Interaction

Now we will move towards recovering the usual dipole-interaction Hamiltonian.[3] Consider the first (kinetic) term in the minimal-coupling Hamiltonian (9.24):

$$\frac{[\mathbf{p}_{\rm e} + e\mathbf{A}(\mathbf{r}_{\rm e})]^2}{2m_{\rm e}} = \frac{p_{\rm e}^2}{2m_{\rm e}} + \frac{e}{m_{\rm e}}\mathbf{A}\cdot\mathbf{p}_{\rm e} + \frac{e^2}{2m_{\rm e}}A^2. \tag{9.25}$$

Note that in general $\mathbf{p}_{\rm e}$ and $\mathbf{A}$ do not commute, since $\mathbf{p}_{\rm e} = -i\hbar\nabla_{\rm e}$, and $\mathbf{A} = \mathbf{A}(\mathbf{r}_{\rm e})$. However, they *do* commute here, since we are in the Coulomb gauge where $\nabla\cdot\mathbf{A} = 0$. Within the electric-dipole approximation, we take the vector potential $\mathbf{A}$ to be independent of position, evaluating $\mathbf{A}$ at the nuclear position (and taking $\mathbf{r}_{\rm nuc} = 0$). That is, we take the variation of $\mathbf{A}$ to be negligible over the scale of the atomic size. This approximation is also called the **long-wavelength approximation**. Then the minimal-coupling Hamiltonian becomes

$$H = \frac{\mathbf{p}_{\rm e}^2}{2m_{\rm e}} + V(\mathbf{r}_{\rm e}) + \frac{e}{m_{\rm e}}\mathbf{p}_{\rm e}\cdot\mathbf{A}(0) + \frac{e^2}{2m_{\rm e}}A^2(0) + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right)$$

(minimal-coupling Hamiltonian, long-wavelength approximation) (9.26)

in the long-wavelength approximation. Comparison to the uncoupled Hamiltonian (9.23) gives

$$H_{\rm AF} = \frac{e}{m_{\rm e}}\mathbf{p}_{\rm e}\cdot\mathbf{A}(0) + \frac{e^2}{2m_{\rm e}}A^2(0)$$

(minimal-coupling interaction, long-wavelength approximation) (9.27)

as the interaction Hamiltonian in terms of the vector potential. Here, the $\mathbf{p}_{\rm e}\cdot\mathbf{A}$ term plays a role similar to the familiar $\mathbf{d}\cdot\mathbf{E}$ Hamiltonian, as we will discuss in more detail below, while the A^2 term is atomic-level-independent and for many purposes may be ignored.

9.3.1 Power–Zienau Transformation

The atom–field interaction here is still in terms of the vector potential, and so we would like to see the connection to the usual interaction with the electric field. We thus use the unitary **Power–Zienau transformation**[4] (again, in the long-wavelength approximation)

$$U = e^{i e\mathbf{r}_{\rm e}\cdot\mathbf{A}(0)/\hbar}$$

(Power–Zienau transformation, long-wavelength approximation) (9.28)

[3] For further reading, see Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993), Sections 4.3-4.4, pp. 119-125; and J. R. Ackerhalt and P. W. Milonni, ''Interaction Hamiltonian of quantum optics,'' *Journal of the Optical Society of America B* **1**, 116 (1984).

[4] E. A. Power and S. Zienau, ''Coulomb Gauge in Non-Relativistic Quantum Electro-Dynamics and the Shape of Spectral Lines,'' *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **251**, 427 (1959); R. G. Woolley, ''Molecular Quantum Electrodynamics,'' *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **321**, 557 (1971).

to transform the Hamiltonian. The unitary transformation here amounts to a *gauge transformation*, and we will refer to the situations before and after the unitary transformation as being different gauges. The new Hamiltonian is

$$\tilde{H} = UHU^\dagger. \tag{9.29}$$

Using the identity

$$e^A B e^{-A} = B + [A,B] + \frac{1}{2!}[A,[A,B]] + \cdots, \tag{9.30}$$

we can see that the new momentum is

$$\tilde{\mathbf{p}}_\mathrm{e} = U\mathbf{p}_\mathrm{e}U^\dagger = \mathbf{p}_\mathrm{e} + [ie\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(0)/\hbar, \mathbf{p}_\mathrm{e}] = \mathbf{p}_\mathrm{e} - e\mathbf{A}(0), \tag{9.31}$$

so that we can take care of the transformation of the kinetic energy by writing it in terms of the *untransformed* momentum (which becomes both the canonical and the kinematic momentum in the Power–Zienau transformation):

$$U(\mathbf{p}_\mathrm{e} + e\mathbf{A})^2 U^\dagger = \mathbf{p}_\mathrm{e}^2. \tag{9.32}$$

Also, the electric-field components transform as

$$\begin{aligned}\tilde{E}^\perp_\beta(\mathbf{r}) &= U(\mathbf{r}_\mathrm{e})E^\perp_\beta(\mathbf{r})U^\dagger(\mathbf{r}_\mathrm{e}) \\ &= E^\perp_\beta(\mathbf{r}) + \left[ie\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(0)/\hbar, E^\perp_\beta(\mathbf{r})\right] \\ &= E^\perp_\beta(\mathbf{r}) + \frac{e}{\epsilon_0} r_{\mathrm{e},\alpha}\delta^\perp_{\alpha\beta}(\mathbf{r}),\end{aligned} \tag{9.33}$$

where we used the commutation relation between the vector potential and electric field in free space from Eq. (8.207). Thus, the transformation of the electric-field energy gives

$$\begin{aligned}U(\mathbf{r}_\mathrm{e})\left[\int d^3r \left[\mathbf{E}^\perp(\mathbf{r})\right]^2\right]U^\dagger(\mathbf{r}_\mathrm{e}) &= \int d^3r \left[U(\mathbf{r}_\mathrm{e})\mathbf{E}^\perp(\mathbf{r})U^\dagger(\mathbf{r}_\mathrm{e})\right]^2 \\ &= \int d^3r \left[\mathbf{E}^\perp(\mathbf{r})\right]^2 + \frac{2e}{\epsilon_0}\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp(0) + \frac{1}{\epsilon_0^2}\int d^3r \left[\mathbf{P}^\perp(\mathbf{r})\right]^2,\end{aligned} \tag{9.34}$$

where

$$\mathbf{P}(\mathbf{r}) := -e\mathbf{r}_\mathrm{e}\delta^3(\mathbf{r}) \tag{9.35}$$

(atomic polarization density)

is the polarization density for the atom, and the transverse polarization is the same but with the delta function replaced by the transverse delta function:

$$P^\perp_\beta(\mathbf{r}) := -er_{\mathrm{e},\alpha}\delta^\perp_{\alpha\beta}(\mathbf{r}). \tag{9.36}$$

Thus, the transformed Hamiltonian is

$$\tilde{H} = \frac{\mathbf{p}_\mathrm{e}^2}{2m_\mathrm{e}} + V(\mathbf{r}_\mathrm{e}) + e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2B^2\right) + \frac{1}{2\epsilon_0}\int d^3r \left[\mathbf{P}^\perp(\mathbf{r})\right]^2,$$

(transformed Hamiltonian) (9.37)

which again is written only in terms of the untransformed coordinates $\mathbf{p}_\mathrm{e}$ and $\mathbf{E}(\mathbf{r})$. Comparison to the uncoupled Hamiltonian (9.23) gives an interaction Hamiltonian of

$$\tilde{H}_\mathrm{AF} = e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp(0) + \frac{1}{2\epsilon_0}\int d^3r \left[\mathbf{P}^\perp(\mathbf{r})\right]^2 = -\mathbf{d}\cdot\mathbf{E}^\perp(0) + \frac{1}{2\epsilon_0}\int d^3r \left[\mathbf{P}^\perp(\mathbf{r})\right]^2,$$

(dipole interaction Hamiltonian) (9.38)

where the atomic dipole moment is $\mathbf{d} = -e\mathbf{r}_\mathrm{e}$, and the atomic center is located at $\mathbf{r} = 0$. The second term, representing a (divergent) dipole self-energy, is commonly dropped, although sometimes it makes an

important explicit contribution (e.g., in the calculation of the Lamb shift). We thus recover the familiar form for the electric-dipole Hamiltonian

$$\tilde{H}_{\text{AF}} \approx -\mathbf{d}\cdot\mathbf{E}^{\perp} \tag{9.39}$$
(dipole interaction Hamiltonian)

in the long-wavelength approximation.

9.3.1.1 Electric Displacement

Let's once again examine the transformed electric field.[5] From Eq. (9.33), we have

$$\begin{aligned}\tilde{E}^{\perp}_{\alpha}(\mathbf{r}) &= E^{\perp}_{\alpha}(\mathbf{r}) + \frac{e}{\epsilon_0} r_{\text{e},\alpha}\delta^{\perp}_{\alpha\beta}(\mathbf{r}) \\ &= E^{\perp}_{\alpha}(\mathbf{r}) - \frac{1}{\epsilon_0}P^{\perp}_{\alpha}(\mathbf{r}),\end{aligned} \tag{9.40}$$

so that

$$\epsilon_0\mathbf{E}^{\perp}(\mathbf{r}) = \epsilon_0\tilde{\mathbf{E}}^{\perp}(\mathbf{r}) + \mathbf{P}^{\perp}(\mathbf{r}) = \tilde{\mathbf{D}}^{\perp}(\mathbf{r}). \tag{9.41}$$

Thus, we see that the electric field in the original gauge ("A gauge"), or more precisely $\epsilon_0\mathbf{E}^{\perp}$, which is what couples to the atom, corresponds to the *dielectric displacement* in the new gauge ("E gauge"). Since the polarization density is localized, this is in fact the same as $\epsilon_0\tilde{\mathbf{E}}^{\perp}$ away from the origin.

9.3.1.2 Active and Passive Viewpoints

The viewpoint of the Power–Zienau transformation that we presented above—that the electric-dipole Hamiltonian arises from a unitary transformation of the Hamiltonian—is the "active" view of the transformation. We can alternately use a "passive" view, where we can get the same form of the interaction Hamiltonian (9.37) without transforming it, if we use the *coordinate* transformation[6]

$$U' = e^{-ie\mathbf{r}_{\text{e}}\cdot\mathbf{A}/\hbar}, \tag{9.42}$$

which gives new coordinates

$$\begin{aligned}\mathbf{p}'_{\text{e}} &= U'\mathbf{p}_{\text{e}}U'^{\dagger} = \mathbf{p}_{\text{e}} + e\mathbf{A} \\ \mathbf{A}' &= \mathbf{A} \\ \mathbf{r}'_{\text{e}} &= \mathbf{r}_{\text{e}} \\ \mathbf{B}' &= \mathbf{B} \\ \mathbf{E}'^{\perp} &= \mathbf{E}^{\perp} + \frac{1}{\epsilon_0}\mathbf{P}^{\perp}.\end{aligned} \tag{9.43}$$

Then the *untransformed* Hamiltonian is

$$\begin{aligned}H &= \frac{[\mathbf{p}_{\text{e}} + e\mathbf{A}(\mathbf{r}_{\text{e}})]^2}{2m_{\text{e}}} + V(\mathbf{r}_{\text{e}}) + \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right) \\ &= \frac{\mathbf{p}'^{\,2}_{\text{e}}}{2m_{\text{e}}} + V(\mathbf{r}'_{\text{e}}) + e\mathbf{r}'_{\text{e}}\cdot\mathbf{E}'^{\perp} + \frac{\epsilon_0}{2}\int d^3r\left(E'^{\perp 2} + c^2B'^2\right) + \frac{1}{2\epsilon_0}\int d^3r\left[\mathbf{P}^{\perp}(\mathbf{r})\right]^2,\end{aligned} \tag{9.44}$$

which has the same form as (9.37), but in *transformed coordinates*. (Before, we wrote the *transformed* Hamiltonian in *untransformed* coordinates). Note here that $\epsilon_0\mathbf{E}'$ is the dielectric displacement in the old variables. In both the active and passive viewpoints, the field that couples to the atom is in fact a displacement, not an electric field, although it is conventional to write it as an electric field (since they are the same outside the atom).

[5] For further discussion, see J. R. Ackerhalt and P. W. Milonni, "Interaction Hamiltonian of quantum optics," *Journal of the Optical Society of America B* **1**, 116 (1984).

[6] For further discussion, see J. R. Ackerhalt and P. W. Milonni, *op. cit.*

9.3.1.3 Göppert-Mayer Transformation

The Power–Zienau transformation was actually preceded by an equivalent canonical transformation derived by Göppert-Mayer.[7] Recall that the *classical* equations of motion are unchanged if we add a total time derivative, say

$$\frac{d}{dt}S(q,t), \tag{9.45}$$

to the Lagrangian $L(q,\dot{q};t)$ in terms of the generalized coordinate q. Doing so induces a **canonical transformation**, and we can choose the **generating function of the canonical transformation** $S(q,t)$ to be $S(\mathbf{r}_\mathrm{e},t) = e\mathbf{r}_\mathrm{e}\cdot\mathbf{A}$. The Lagrangian (9.7) is

$$L = \frac{1}{2}m_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e}^2 - e\dot{\mathbf{r}}_\mathrm{e}\cdot\mathbf{A}(\mathbf{r}_\mathrm{e}) + e\phi(\mathbf{r}_\mathrm{e}), \tag{9.46}$$

which thus transforms to

$$\begin{aligned}\tilde{L} &= L + \frac{d}{dt}e\mathbf{r}_\mathrm{e}\cdot\mathbf{A}\\ &= L + e\dot{\mathbf{r}}_\mathrm{e}\cdot\mathbf{A} + e\mathbf{r}_\mathrm{e}\cdot\dot{\mathbf{A}}\\ &= \frac{1}{2}m_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e}^2 + e\phi(\mathbf{r}_\mathrm{e}) + e\mathbf{r}_\mathrm{e}\cdot\dot{\mathbf{E}}^\perp.\end{aligned} \tag{9.47}$$

We can see that the generating function exactly cancels the $\mathbf{p}_\mathrm{e}\cdot\mathbf{A}$ term and adds in the $\mathbf{r}_\mathrm{e}\cdot\mathbf{E}$ term. Thus, this canonical transformation classically effects the transformation from H to $\tilde{H}$.

9.3.2 $\mathbf{p}\cdot\mathbf{A}$ vs. $\mathbf{r}\cdot\mathbf{E}$

Now it appears that, depending on the choice of gauge, we have two possible interaction Hamiltonians in the long-wavelength approximation. In the E gauge, we have from Eq. (9.38)

$$H_\mathrm{AF}^{(E)} = e\mathbf{r}_\mathrm{e}\cdot\mathbf{E} + \frac{1}{2\epsilon_0}\int d^3r\left[\mathbf{P}^\perp(\mathbf{r})\right]^2 \approx e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}, \tag{9.48}$$

where again the polarization term is typically negligible. The A gauge, on the other hand, gives from Eq. (9.27)

$$H_\mathrm{AF}^{(A)} = \frac{e}{m_\mathrm{e}}\mathbf{p}_\mathrm{e}\cdot\mathbf{A} + \frac{e^2}{2m_\mathrm{e}}A^2 \approx \frac{e}{m_\mathrm{e}}\mathbf{p}_\mathrm{e}\cdot\mathbf{A}, \tag{9.49}$$

where we have assumed that the A^2 term is negligible, which is typically the case. Comparing these two Hamiltonians amounts to comparing $\mathbf{p}_\mathrm{e}\cdot\mathbf{A}$ to $m_\mathrm{e}\mathbf{r}_\mathrm{e}\cdot\mathbf{E}$. This seems reasonable, as $\mathbf{p}_\mathrm{e} = m_\mathrm{e}\partial_t\mathbf{r}_\mathrm{e}$, and $\mathbf{E}^\perp = -\partial_t\mathbf{A}$, so the two Hamiltonians seem to differ by moving a time derivative from one factor to the other, as in some sort of integration by parts. However, the *matrix elements* of these two Hamiltonians differ, as we will now show.

First, we must derive a relation between matrix elements of $\mathbf{r}$ and $\mathbf{p}$. Consider the commutator

$$[\mathbf{r}_\mathrm{e}, H_\mathrm{A}] = i\hbar\frac{\mathbf{p}_\mathrm{e}}{m_\mathrm{e}}, \tag{9.50}$$

where the atomic Hamiltonian is as usual $H_\mathrm{A} = \mathbf{p}_\mathrm{e}^2/2m_\mathrm{e} + V(\mathbf{r}_\mathrm{e})$. Then the momentum operator becomes

$$\mathbf{p}_\mathrm{e} = -i\frac{m_\mathrm{e}}{\hbar}[\mathbf{r}_\mathrm{e}, H_\mathrm{A}], \tag{9.51}$$

or in matrix elements in the basis of eigenstates $|j\rangle$ of H_A,

$$\begin{aligned}\langle j|\mathbf{p}_\mathrm{e}|j'\rangle &= -i\frac{m_\mathrm{e}}{\hbar}\langle j|[\mathbf{r}_\mathrm{e}, H_\mathrm{A}]|j'\rangle\\ &= im_\mathrm{e}\omega_{jj'}\langle j|\mathbf{r}_\mathrm{e}|j'\rangle,\end{aligned} \tag{9.52}$$

[7] M. Göppert-Mayer, "Über Elementarakte mit zwei Quantensprüngen," *Annalen der Physik* **9**, 273 (1931).

where

$$\omega_{jj'} := \frac{E_j - E_{j'}}{\hbar} \tag{9.53}$$

is the transition frequency (and could be positive or negative depending on the ordering of the states).

Thus, for the matrix elements of the two Hamiltonians, we may write

$$\left| \frac{\langle j|H_{\rm AF}^{(E)}|j'\rangle}{\langle j|H_{\rm AF}^{(A)}|j'\rangle} \right| = \left| \frac{m_{\rm e}\langle j|\mathbf{r}_{\rm e}\cdot\mathbf{E}|j'\rangle}{\langle j|\mathbf{p}_{\rm e}\cdot\mathbf{A}|j'\rangle} \right| = \frac{\omega}{|\omega_{jj'}|}. \tag{9.54}$$

The matrix elements for the interaction Hamiltonians are different! This would seem to give different physical predictions, depending on which Hamiltonian we use. What gives?

This problem has generated much discussion, and its resolution is somewhat tricky. One ''resolution'' states that a unitary transformation generated the second Hamiltonian from the first. Thus, to get the same answers in both gauges, one must also apply the same transformation to the states, and then trivially the matrix elements must be the same (matrix elements and expectation values are always invariant under unitary transformations). But that still doesn't help much: given a particular pair of states, say atomic energy eigenstates, which is the appropriate interaction Hamiltonian to use? Using the passive viewpoint, which avoids the unitary transformation of the Hamiltonian—the Hamiltonian is the same in either gauge, just expressed in different coordinates—doesn't actually help, because we would still need to find the action of the new variables on the old state, which is equivalent to making the unitary transformation.

The point is, that *physically measureable* quantities are gauge-invariant, and thus should be calculable with either Hamiltonian.[8] One ''resolution'' of this ''paradox'' asserts that because the E-gauge atomic energy operator is equivalent to the unperturbed Hamiltonian (and in particular, the kinematic and canonical momenta are equivalent), the usual energy eigenstates are associated with the E gauge, and the computation of matrix elements is most straightforward here.[9] This interpretation is a bit tricky, since even in the A gauge, the kinematic and canonical momenta are equivalent in the absence of a perturbing field, and we have already counted the longitudinal binding field as part of the background, not the perturbation. The interpretation that we will prefer here is that the energy eigenstates can appropriately be used for *either* gauge, but only when we ask physical questions.[10] But then what about the different matrix elements? Broadly speaking, there are two situations that have slightly different resolutions.

1. A matrix element between two *stable* states is physical only for an *energy-conserving* process (at least in the case of a time-independent Hamiltonian). In this case, the laser and transition frequencies are equivalent ($\omega = \omega_0$) to enforce energy conservation. This is, for example, the case when dealing with the Hamiltonian treatment of spontaneous decay.

2. In cases where $\omega \neq \omega_0$, as can happen for a homogeneously broadened line or an inelastic process, the matrix element represents an *intermediate* transition in a larger, multiphoton process that conserves energy. We then regard the overall process as the physical one, and the combination of the matrix elements, *summed over all intermediate states*, is gauge-invariant. We will see examples of this in the Casimir–Polder effect and Lamb shift.

The main idea here is that when the matrix elements differ between the gauges, then a physically relevant sum over the states is gauge-invariant. It may be the case that the sum has faster *convergence* in one gauge compared to another, so that for a specific calculation there may be a more convenient gauge. However, the final answer should always be the same.

It's possible that summing over all intermediate states and restricting yourself to physical results *still* doesn't produce the same answer in both gauges, in which case the next step is to keep the extra self-energy

[8] For a good discussion of this point, see Marlan O. Scully and M. Suhail Zubairy, *Quantum Optics* (Cambridge, 1997), Appendix 5.A, p. 178.

[9] Marlan O. Scully and M. Suhail Zubairy, *op. cit.*

[10] Edwin A. Power, ''A Review of Canonical Transformations as they Affect Multiphoton Processes,'' in *Multiphoton Processes: Proceedings of an International Conference at the University of Rochester, Rochester, N.Y., June 6-9, 1977*, Joseph H. Eberly and Peter Lambropoulos, Eds. (Wiley, 1978); Claude Cohen–Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Photons & Atoms* (Wiley, 1989), Complement $B_{\rm IV}$, p. 316; Zoltan Fried, ''Vector Potential Versus Field Intensity,'' *Physical Review A* **8**, 2835 (1973) (doi: 10.1103/PhysRevA.8.2835).

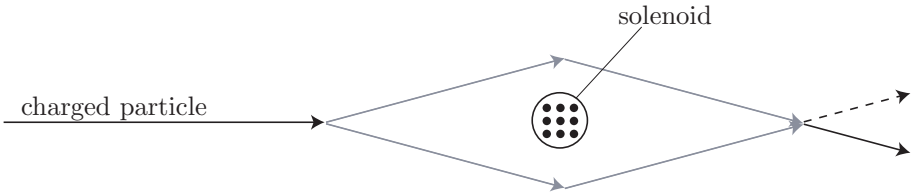
terms we ignored in the two interaction Hamiltonians (9.48) and (9.49). These extra terms are important in getting gauge-independent results, for example, in the Lamb shift, as we will see later in Section 13.12.

9.4 Why the Vector Potential?

We have seen in the canonical quantization of the field, that the vector potential plays a central role. We have now also just seen that it plays a central role in the coupling of an atom to the electromagnetic field. But why the potential, and not the fields themselves? After all, the electromagnetic fields are far more intuitive, and the vector potential is somewhat ambiguous due to gauge freedom.

In physics, it helps our intuition to have *local* interactions. For example, in classical electrodynamics, we can do away with electric and magnetic fields and regard electrodynamics as a theory of interacting charged particles. However, some strange things happen: the force between two moving, charged particles is *not* a central force. That is, it appears to violate Newton's third law, that the electromagnetic force on one particle is not necessarily equal and oppose to the force on the other. Momentum conservation is saved by attributing some of the momentum to the electromagnetic fields.[11] Another example comes again from considering the force between two initially stationary particles. You observe that when you start to wiggle one of them, the other doesn't respond to the wiggle until a time r/c later, and thus the retarded time is important in electrodynamics. However, consider the *direction* of the force on a stationary particle due to one moving at constant velocity: you might be tempted to conclude that it points to the *retarded* location of the moving particle, when in fact it points to the *instantaneous* location.[12] (If the motion is not of constant velocity, then the direction of the force is different still.) In classical physics, then, one function of introducing the electromagnetic fields is to avoid such counterintuitive, nonlocal interactions: one particle generates a field, which propagates to the other particle and thus influences it.

In quantum mechanics, to preserve the same sort of locality of interactions, we are forced to include the vector potential at a fundamental level.[13] The most striking example where this is the case is the **Aharonov–Bohm effect**,[14] which deals with a charged particle moving in the exterior of a solenoid. In particular, suppose we set up an interference experiment, where the charged particle, after being split, travels on either side of the solenoid before being recombined.



If the ideal solenoid is of radius R, is oriented along the z-axis, and has a linear density of turns N with current I, the magnetic field is[15]

$$\mathbf{B} = \begin{cases} \mu_0 N I \hat{z}, & r < R \\ 0, & r > R, \end{cases} \tag{9.55}$$

[11] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice-Hall, 2013), Section 8.2, p. 360.

[12] Richard Feynman, Robert B. Leighton, and Matthew L. Sands, *The Feynman Lectures on Physics*, Vol. II, Chapter 21 (Addison-Wesley, 1989).

[13] I learned this analogy to locality in classical physics from Tanmoy Bhattacharya. For more notes and a good discussion of the Aharonov–Bohm effect, see J. J. Sakurai, *Modern Quantum Mechanics* (Addison-Wesley, 1994), pp. 136-9.

[14] Y. Aharonov and D. Bohm, "Significance of Electromagnetic Potentials in the Quantum Theory," *Physical Review* **115**, 485 (1959) (doi: 10.1103/PhysRev.115.485). This effect was discussed earlier by W. Ehrenberg and R. E. Siday, "The Refractive Index in Electron Optics and the Principles of Dynamics," *Proceedings of the Physical Society. Section B* **62**, 8 (1949) (doi: 10.1088/0370-1301/62/1/303).

[15] David J. Griffiths, *op. cit.*, p. 237.

and thus vanishes outside the solenoid. The vector potential, on the other hand, is[16]

$$\mathbf{A} = \begin{cases} \dfrac{\mu_0 NI}{2} r\hat{\phi}, & r < R \\ \dfrac{\mu_0 NI}{2}\dfrac{R^2}{r}\hat{\phi}, & r > R, \end{cases} \tag{9.56}$$

so that the vector potential does *not* vanish outside the solenoid.

Now consider the Schrödinger equation, including the minimal-coupling replacement, describing the particle motion in the presence of the magnetic field:

$$i\hbar\partial_t\psi = \frac{1}{2m}\left[\frac{\hbar}{i}\nabla - q\mathbf{A}\right]^2\psi. \tag{9.57}$$

We will now effect the gauge transformation that we noted above as follows. Under the replacement

$$\psi \longrightarrow \psi e^{-iq\chi(\mathbf{r})/\hbar}, \tag{9.58}$$

where $\chi(\mathbf{r})$ is some function (that defines the gauge transformation), we see that

$$\nabla\psi \longrightarrow \left[\nabla\psi - \frac{iq\nabla\chi(\mathbf{r})}{\hbar}\psi\right] e^{-iq\chi(\mathbf{r})/\hbar}, \tag{9.59}$$

and thus the Schrödinger equation is invariant if we also let

$$\mathbf{A} \longrightarrow \mathbf{A} - \nabla\chi(\mathbf{r}). \tag{9.60}$$

In the region outside the solenoid, $\mathbf{B} = \nabla\times\mathbf{A} = 0$, so that we may choose $\chi(\mathbf{r})$ such that $\mathbf{A} = \nabla\chi$, and in particular,

$$\chi(\mathbf{r}) = \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}\cdot d\mathbf{s}. \tag{9.61}$$

With this choice of χ, the vector potential goes away, and the Schrödinger equation becomes that of the free particle. That is, assuming a wave function

$$\psi = \psi_0 \exp\left[\frac{iq}{\hbar}\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}\cdot d\mathbf{s}\right], \tag{9.62}$$

then ψ_0 is a solution to the free-particle wave equation. Thus, the phase shift accumulated by a moving particle due to the presence of the field is

$$\phi = \frac{q}{\hbar}\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}\cdot d\mathbf{s}, \tag{9.63}$$

where the integral is along the particle's path, and in the interferometer above, we can write the phase *difference* of the two arms as a closed-path integral

$$\Delta\phi = \frac{q}{\hbar}\oint \mathbf{A}\cdot d\mathbf{s}, \tag{9.64}$$

where the contour is around the total path of the interferometer. Note that we can now write

$$\Delta\phi = \frac{q}{\hbar}\int(\nabla\times\mathbf{A})\cdot d\mathbf{a} = \frac{q}{\hbar}\int \mathbf{B}\cdot d\mathbf{a} = \frac{q}{\hbar}\Phi_B, \tag{9.65}$$

where Φ_B is the enclosed magnetic flux. For an electron with $q = -e$, the phase shift becomes

$$\Delta\phi = -\frac{e}{\hbar}\Phi_B = -2\pi\frac{\Phi_B}{\Phi_0}, \tag{9.66}$$

(Aharonov–Bohm phase)

[16] David J. Griffiths, *op. cit.*, p. 247.

where

$$\Phi_0 = \frac{h}{e} \approx 4.14 \times 10^{-7} \text{ Wb} \tag{9.67}$$

is a fundamental unit of magnetic flux, and 1 Wb = 1 T m^2 = 1 V s.

The whole point is this: we can observe interference fringes due to the magnetic field, *even though the particle stays in regions of zero magnetic field*. Even though this remarkable result can be explained in terms of flux of the magnetic field, it motivates the fundamental nature of the vector potential if we are to maintain a *local* interaction between particles and fields: evidently the quantum interaction of a particle and the magnetic field is *nonlocal*.

9.5 Multipole Interactions

To generalize the above results for the dipole Hamiltonian (9.37) in the long-wavelength approximation, we will now consider the more general Power–Zienau transformation *without* making the long-wavelength approximation. In this way, we will derive general expressions for the atomic interaction with the electric and magnetic fields, and then we will expand these to generate the higher-order multipole couplings.[17]

9.5.1 Atomic Polarization Field

We will start by making the approximation of a heavy nucleus, $m_{\text{nuc}} \gg m_{\text{e}}$, so we identify the reduced electron mass with the normal electron mass and we will assume that the nuclear position $\mathbf{r}_{\text{nuc}} = 0$ defines the center of mass for the system. Then we can write the polarization field for the atom—here, a singly charged nucleus at $\mathbf{r}_{\text{nuc}} = 0$ and an electron at $\mathbf{r}_{\text{e}}$—as the line integral

$$\mathbf{P}(\mathbf{r}) = -e\mathbf{r}_{\text{e}} \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_{\text{e}}). \tag{9.68}$$
(atomic polarization density)

To see that this is correct, recall that the polarization field corresponding to a charge density ρ satisfies[18]

$$\nabla \cdot \mathbf{P} = -\rho, \tag{9.69}$$

or since the transverse polarization does not contribute, this is really only a constraint on the *longitudinal* polarization.

$$\nabla \cdot \mathbf{P}^{\parallel} = -\rho. \tag{9.70}$$

Computing the divergence of the atomic polarization (9.68),

$$\begin{aligned} \nabla \cdot \mathbf{P} &= -e\mathbf{r}_{\text{e}} \cdot \nabla \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_{\text{e}}) \\ &= e \int_0^1 ds\, \frac{\partial}{\partial s} \delta^3(\mathbf{r} - s\mathbf{r}_{\text{e}}) \\ &= e\delta^3(\mathbf{r} - \mathbf{r}_{\text{e}}) - e\delta^3(\mathbf{r}) \\ &= -\rho, \end{aligned} \tag{9.71}$$

where the last equality holds if we identify the charge distribution

$$\rho = e\delta^3(\mathbf{r}) - e\delta^3(\mathbf{r} - \mathbf{r}_{\text{e}}). \tag{9.72}$$

[17] E. A. Power and S. Zienau, *op. cit.*; R. G. Woolley, *op. cit.* See also Claude Cohen–Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Photons & Atoms* (Wiley, 1989), Section IV.C, p. 280; and Werner Vogel and Dirk-Gunnar Welsch, *Quantum Optics*, 3rd ed. (Wiley, 2006).

[18] See, for example, David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice-Hall, 2013), p. 174.

The first term is obviously the nuclear charge density, while the second term represents the electron charge density. Note that despite the expression here, we are *not* necessarily assuming a localized charge density for the electron, as $\mathbf{r}_\mathrm{e}$ is an operator and thus is still subject to uncertainty and quantum fluctuations.

You can visualize the above result (9.68) for the polarization as follows. The atom, in our simplified model, consists of two opposite and separated charges. The polarization is the dipole moment per unit volume, and here we represent it by a continuum of delta-function-localized dipoles, forming a line between the two charges. Each dipole moment is an idealized charge pair, and the charges for the successive dipole moments exactly cancel each other, except at the endpoints of the line. Of course we don't want the endpoints of the line to cancel, since those are the atomic charges. The line of dipoles isn't unique, since any path connecting the nucleus to the electron will do. That freedom is implicit since we did not constrain the *curl* of $\mathbf{P}$, merely the divergence. However, we have chosen the simplest path, and it is consistent with the requirement (9.69).

9.5.2 Atomic Magnetization Field

We can also define a magnetization field for the electron: while the polarization related to the atomic charge distribution, the magnetization (magnetic dipole moment per unit volume) summarizes the magnetic properties of the atom due to *motion* of the charge distribution. To motivate this field, we can differentiate Eq. (9.69) to obtain

$$\nabla \cdot \partial_t \mathbf{P} = -\partial_t \rho. \tag{9.73}$$

Comparing this to the continuity equation,

$$\nabla \cdot \mathbf{j} = -\partial_t \rho, \tag{9.74}$$

we can see that we can identify $\mathbf{j} - \partial_t \mathbf{P}$ as an irrotational vector field:

$$\nabla \cdot (\mathbf{j} - \partial_t \mathbf{P}) = 0. \tag{9.75}$$

We can thus write this field as the curl of some other vector field. Recalling also that the curl of the magnetization $\mathbf{M}$ behaves as an effective current density,[19]

$$\mathbf{j}_\mathrm{m} = \nabla \times \mathbf{M}, \tag{9.76}$$

we can conveniently interpret $\mathbf{j} - \partial_t \mathbf{P}$ as being the curl of the atomic magnetization:

$$\mathbf{j} - \partial_t \mathbf{P} = \nabla \times \mathbf{M}. \tag{9.77}$$

The longitudinal magnetization does not contribute here, so this is really only a constraint on the *transverse* magnetization:

$$\mathbf{j} - \partial_t \mathbf{P} = \nabla \times \mathbf{M}^\perp. \tag{9.78}$$

If we identify the atomic current density

$$\mathbf{j} = -e\dot{\mathbf{r}}_\mathrm{e}\delta^3(\mathbf{r} - \mathbf{r}_\mathrm{e}), \tag{9.79}$$

and differentiate the atomic polarization (9.68),

$$\partial_t \mathbf{P}(\mathbf{r}) = -e\dot{\mathbf{r}}_\mathrm{e} \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}) - e\mathbf{r}_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e} \cdot \int_0^1 ds\, \nabla_\mathrm{e}\delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}), \tag{9.80}$$

we can verify directly that the constraint (9.78) is satisfied by the expression

$$\mathbf{M}(\mathbf{r}) = -e\mathbf{r}_\mathrm{e} \times \dot{\mathbf{r}}_\mathrm{e} \int_0^1 ds\, s\, \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}), \tag{9.81}$$

(atomic magnetization density)

[19] David J. Griffiths, *op. cit.*, p. 275.

which we do as follows, by employing the "bac-cab rule" $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B})$:

$$\begin{aligned}\nabla\times\mathbf{M}(\mathbf{r}) &= e\,(\mathbf{r}_\mathrm{e}\times\dot{\mathbf{r}}_\mathrm{e})\times\int_0^1 ds\,s\,\nabla\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\ &= -e\dot{\mathbf{r}}_\mathrm{e}\mathbf{r}_\mathrm{e}\cdot\int_0^1 ds\,s\,\nabla\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e}) - e\mathbf{r}_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e}\cdot\int_0^1 ds\,s\,\nabla\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\ &= -e\dot{\mathbf{r}}_\mathrm{e}\int_0^1 ds\,s\,\frac{\partial}{\partial s}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e}) + e\mathbf{r}_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e}\cdot\int_0^1 ds\,\nabla_\mathrm{e}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\ &= -e\dot{\mathbf{r}}_\mathrm{e}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e}) - e\dot{\mathbf{r}}_\mathrm{e}\int_0^1 ds\,\frac{\partial}{\partial s}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e}) + e\mathbf{r}_\mathrm{e}\dot{\mathbf{r}}_\mathrm{e}\cdot\int_0^1 ds\,\nabla_\mathrm{e}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\ &= \mathbf{j}(\mathbf{r})-\partial_t\mathbf{P}(\mathbf{r}).\end{aligned} \tag{9.82}$$

Here, we used expressions (9.79) and (9.80) for the atomic current density and derivative of the polarization, respectively. Again, the choice of magnetization here is not unique, but is a simple choice that satisfies the constraint (9.78).

9.5.3 Power–Zienau Transformation

The more general Power–Zienau transformation is then given by the unitary operator

$$U=\exp\left[-\frac{i}{\hbar}\int d^3r\;\mathbf{P}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r})\right]. \tag{9.83}$$
(Power–Zienau transformation)

We will see later [in Eq. (9.111)] that the dipole polarization is the first term in a multipole expansion of $\mathbf{P}(\mathbf{r})$, and so in the dipole approximation the Power–Zienau operator reduces to

$$U=\exp\left[-\frac{i}{\hbar}\int d^3r\;\mathbf{P}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r})\right]\approx\exp\left[\frac{ie}{\hbar}\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(0)\right], \tag{9.84}$$

which is precisely the operator (9.28) we used in the long-wavelength approximation. Clearly, the electron position and vector potential (hence, magnetic field) are still invariant under this transformation, so it remains to transform the electric field and canonical electron momentum, and of course, the minimal-coupling Hamiltonian (9.24)

$$H=\frac{[\mathbf{p}_\mathrm{e}+e\mathbf{A}(\mathbf{r}_\mathrm{e})]^2}{2m_\mathrm{e}}+V(\mathbf{r}_\mathrm{e})+\frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2}+c^2B^2\right), \tag{9.85}$$

in order to obtain the multipole interactions in terms of the electric and magnetic fields.

9.5.3.1 Electric Field

Using the transformation (9.30) and the Jordan–Pauli commutation relation between the vector potential and electric field in free space from Eq. (8.207),

$$[A_\alpha(\mathbf{r},t),E_\beta(\mathbf{r}',t)]=-\frac{i\hbar}{\epsilon_0}\delta^\perp_{\alpha\beta}(\mathbf{r}-\mathbf{r}'), \tag{9.86}$$

the electric field transforms as

$$\begin{aligned}\tilde{E}^\perp_\beta(\mathbf{r}) &= U(\mathbf{r}_\mathrm{e})E^\perp_\beta(\mathbf{r})U^\dagger(\mathbf{r}_\mathrm{e})\\ &= E^\perp_\beta(\mathbf{r})+\left[-\frac{i}{\hbar}\int d^3r'\,\mathbf{P}(\mathbf{r}')\cdot\mathbf{A}(\mathbf{r}'),E^\perp_\beta(\mathbf{r})\right]\\ &= E^\perp_\beta(\mathbf{r})-\frac{i}{\hbar}\int d^3r'\,P_\alpha(\mathbf{r}')\left[A_\alpha(\mathbf{r}'),E^\perp_\beta(\mathbf{r})\right]\\ &= E^\perp_\beta(\mathbf{r})-\frac{1}{\epsilon_0}\int d^3r'\,P_\alpha(\mathbf{r}')\,\delta^\perp_{\alpha\beta}(\mathbf{r}'-\mathbf{r}),\end{aligned} \tag{9.87}$$

so that

$$\tilde{E}^{\perp}_{\beta}(\mathbf{r}) = E^{\perp}_{\beta}(\mathbf{r}) - \frac{1}{\epsilon_0} P^{\perp}_{\beta}(\mathbf{r}). \tag{9.88}$$
(transformed field operator)

We see here again that the transformed electric field still corresponds to the *untransformed dielectric displacement*. Thus, the electric-field part of the Hamiltonian becomes

$$\begin{aligned} U(\mathbf{r}_\mathrm{e}) \left[\frac{\epsilon_0}{2} \int d^3r \left[\mathbf{E}^{\perp}(\mathbf{r})\right]^2 \right] U^\dagger(\mathbf{r}_\mathrm{e}) &= \frac{\epsilon_0}{2} \int d^3r \left[U(\mathbf{r}_\mathrm{e})\mathbf{E}^{\perp}(\mathbf{r})U^\dagger(\mathbf{r}_\mathrm{e})\right]^2 \\ &= \frac{\epsilon_0}{2} \int d^3r \left[\mathbf{E}^{\perp}(\mathbf{r}) - \frac{1}{\epsilon_0}\mathbf{P}^{\perp}(\mathbf{r})\right]^2 \\ &= \frac{\epsilon_0}{2} \int d^3r \left[\mathbf{E}^{\perp}(\mathbf{r})\right]^2 - \int d^3r\, \mathbf{P}^{\perp}(\mathbf{r}) \cdot \mathbf{E}^{\perp}(\mathbf{r}) + \frac{1}{2\epsilon_0} \int d^3r \left[\mathbf{P}^{\perp}(\mathbf{r})\right]^2. \end{aligned} \tag{9.89}$$

The first term is simply the energy of the transformed field, the second is the atomic interaction (via its polarization density) with the electric field, and the last term is again a dipole self-energy due to the dipole field.

9.5.3.2 Canonical Electron Momentum

Now to carry out the transformation of the momentum. Using

$$[\mathbf{p}_\mathrm{e}, f(\mathbf{r}_\mathrm{e})] = -i\hbar \nabla_\mathrm{e} f(\mathbf{r}_\mathrm{e}), \tag{9.90}$$

for an arbitrary function f, the momentum transforms as

$$\begin{aligned} \tilde{\mathbf{p}}_\mathrm{e} &= U(\mathbf{r}_\mathrm{e})\mathbf{p}_\mathrm{e}U^\dagger(\mathbf{r}_\mathrm{e}) \\ &= \mathbf{p}_\mathrm{e} + \left[-\frac{i}{\hbar} \int d^3r\, \mathbf{P}(\mathbf{r}) \cdot \mathbf{A}(\mathbf{r}), \mathbf{p}_\mathrm{e} \right] \\ &= \mathbf{p}_\mathrm{e} + \int d^3r\, \nabla_\mathrm{e} \left[\mathbf{P}(\mathbf{r}) \cdot \mathbf{A}(\mathbf{r})\right] \\ &= \mathbf{p}_\mathrm{e} - e \int d^3r\, \nabla_\mathrm{e} \left[\mathbf{r}_\mathrm{e} \cdot \mathbf{A}(\mathbf{r}) \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}) \right] \\ &= \mathbf{p}_\mathrm{e} - e \int d^3r\, \mathbf{A}(\mathbf{r}) \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}) - e \int d^3r\, \mathbf{r}_\mathrm{e} \cdot \mathbf{A}(\mathbf{r}) \int_0^1 ds\, \nabla_\mathrm{e} \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}). \end{aligned} \tag{9.91}$$

In the last step, we used the fact that $\nabla(\mathbf{r} \cdot \mathbf{A}) = \mathbf{A}$ if $\mathbf{A}$ is independent of $\mathbf{r}$. The second term is

$$-e \int d^3r\, \mathbf{A}(\mathbf{r}) \int_0^1 ds\, \delta^3(\mathbf{r} - s\mathbf{r}_\mathrm{e}) = -e \int_0^1 ds\, \mathbf{A}(s\mathbf{r}_\mathrm{e}), \tag{9.92}$$

and the third term is

$$\begin{aligned}
-e\int d^3r\,\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(\mathbf{r})\int_0^1 ds\,\nabla_\mathrm{e}\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e}) &= e\int d^3r\,\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(\mathbf{r})\int_0^1 ds\,s\,\nabla\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&= -e\int d^3r\,\nabla\left[\mathbf{r}_\mathrm{e}\cdot\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&= -e\int d^3r\,\left[\mathbf{r}_\mathrm{e}\times\left[\nabla\times\mathbf{A}(\mathbf{r})\right]+\left(\mathbf{r}_\mathrm{e}\cdot\nabla\right)\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&= -e\int d^3r\,\mathbf{r}_\mathrm{e}\times\left[\nabla\times\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&\quad -e\int_0^1 ds\,s\,\left(\mathbf{r}_\mathrm{e}\cdot\nabla\right)\mathbf{A}(s\mathbf{r}_\mathrm{e})\\
&= -e\int d^3r\,\mathbf{r}_\mathrm{e}\times\left[\nabla\times\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&\quad -e\int_0^1 ds\,s\,\frac{\partial}{\partial s}\mathbf{A}(s\mathbf{r}_\mathrm{e})\\
&= -e\int d^3r\,\mathbf{r}_\mathrm{e}\times\left[\nabla\times\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&\quad -e\mathbf{A}(\mathbf{r}_\mathrm{e})+e\int_0^1 ds\,\mathbf{A}(s\mathbf{r}_\mathrm{e}).
\end{aligned} \tag{9.93}$$

Collecting terms, we find

$$\begin{aligned}
\tilde{\mathbf{p}}_\mathrm{e} &= \mathbf{p}_\mathrm{e}-e\mathbf{A}(\mathbf{r}_\mathrm{e})-e\int d^3r\,\mathbf{r}_\mathrm{e}\times\left[\nabla\times\mathbf{A}(\mathbf{r})\right]\int_0^1 ds\,s\,\delta^3(\mathbf{r}-s\mathbf{r}_\mathrm{e})\\
&= \mathbf{p}_\mathrm{e}-e\mathbf{A}(\mathbf{r}_\mathrm{e})-e\mathbf{r}_\mathrm{e}\times\int_0^1 ds\,s\,\left[\nabla\times\mathbf{A}(s\mathbf{r}_\mathrm{e})\right],
\end{aligned} \tag{9.94}$$

and finally the new momentum is

$$\tilde{\mathbf{p}}_\mathrm{e} = \mathbf{p}_\mathrm{e}-e\mathbf{A}(\mathbf{r}_\mathrm{e})-e\mathbf{r}_\mathrm{e}\times\int_0^1 ds\,s\,\mathbf{B}(s\mathbf{r}_\mathrm{e}). \tag{9.95}$$

(transformed momentum)

Thus, the atomic part of the minimal-coupling Hamiltonian transforms as

$$U(\mathbf{r}_\mathrm{e})\frac{\left[\mathbf{p}_\mathrm{e}+e\mathbf{A}(\mathbf{r}_\mathrm{e})\right]^2}{2m_\mathrm{e}}U^\dagger(\mathbf{r}_\mathrm{e}) = \frac{1}{2m_\mathrm{e}}\left[\mathbf{p}_\mathrm{e}-e\mathbf{r}_\mathrm{e}\times\int_0^1 ds\,s\,\mathbf{B}(s\mathbf{r}_\mathrm{e})\right]^2. \tag{9.96}$$

We can see that this part of the transformed Hamiltonian describes the coupling of the atom to the *magnetic* field.

9.5.3.3 Hamiltonian

Collecting pieces from the last two sections, the minimal-coupling Hamiltonian (9.85) becomes

$$\begin{aligned}
\tilde{H} &= U(\mathbf{r}_\mathrm{e})HU^\dagger(\mathbf{r}_\mathrm{e})\\
&= \frac{1}{2m_\mathrm{e}}\left[\mathbf{p}_\mathrm{e}-e\mathbf{r}_\mathrm{e}\times\int_0^1 ds\,s\,\mathbf{B}(s\mathbf{r}_\mathrm{e})\right]^2+V(\mathbf{r}_\mathrm{e})\\
&\quad +\frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2}+c^2B^2\right)-\int d^3r\,\mathbf{P}^\perp(\mathbf{r})\cdot\mathbf{E}^\perp(\mathbf{r})+\frac{1}{2\epsilon_0}\int d^3r\,\left[\mathbf{P}^\perp(\mathbf{r})\right]^2.
\end{aligned} \tag{9.97}$$

(multipole Hamiltonian)

Note again that the canonical variables are *old* canonical variables; it is equally possible to obtain this result by writing the *old* Hamiltonian in terms of new variables. From its form, we can deduce (see Problem 9.3) that the classical canonical momentum for this Hamiltonian is

$$\mathbf{p}_{\rm e} = m_{\rm e}\dot{\mathbf{r}}_{\rm e} + e\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e}). \tag{9.98}$$
(multipole canonical momentum)

That is, it differs from the *kinematic* momentum only by a term proportional to the local magnetic field. Although this argument is classical, we can also interpret this expression for the canonical momentum as the Heisenberg equation

$$\begin{aligned}\dot{\mathbf{r}}_{\rm e} &= -\frac{i}{\hbar}[\tilde{H}, \mathbf{r}_{\rm e}] \\ &= \mathbf{p}_{\rm e} - \frac{e}{m_{\rm e}}\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e}),\end{aligned} \tag{9.99}$$

giving the velocity operator in terms of the canonical momentum.

It is convenient to multiply out the kinetic term in the above Hamiltonian and separate the terms in the Hamiltonian according to their ''function:''

$$\tilde{H} = H_{\rm A} + H_{\rm F} + H_{\rm AE} + H_{\rm AM}. \tag{9.100}$$
(multipole Hamiltonian)

The isolated atomic Hamiltonian is as usual

$$H_{\rm A} = \frac{\mathbf{p}_{\rm e}^2}{2m_{\rm e}} + V(\mathbf{r}_{\rm e}), \tag{9.101}$$
(free-atom Hamiltonian)

and the field Hamiltonian also has its usual form:

$$H_{\rm F} = \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right). \tag{9.102}$$
(free field Hamiltonian)

Recall here that the *longitudinal* electric field is already included in the Coulomb binding potential $V(\mathbf{r}_{\rm e})$.

The coupling of the atom to the electric field is given by the interaction Hamiltonian

$$H_{\rm AE} = -\int d^3r\, \mathbf{P}^{\perp}(\mathbf{r}) \cdot \mathbf{E}^{\perp}(\mathbf{r}) + \frac{1}{2\epsilon_0}\int d^3r\, \left[\mathbf{P}^{\perp}(\mathbf{r})\right]^2. \tag{9.103}$$
(atom–E-field coupling)

The first term gives the atom–field coupling via the atomic polarization density, while the second term represents an atomic self-energy from the coupling of the polarization to its own field. Note that the Coulomb binding potential may also be written a similar form in terms of $\mathbf{P}^{\parallel}(\mathbf{r})$ and thus combined with the transverse self energy (just as it can be regarded as the energy of the longitudinal electric field), but we will separate the contributions here.

Finally, the coupling to the magnetic field has the most complicated form:

$$H_{\rm AM} = -\frac{e}{2m_{\rm e}}\left[\mathbf{p}_{\rm e} \cdot \left(\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e})\right) + \left(\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e})\right) \cdot \mathbf{p}_{\rm e}\right] + \frac{e^2}{2m_{\rm e}}\left[\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e})\right]^2. \tag{9.104}$$

Using the identity $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A})$, while being careful with the order of operators (and noting $\mathbf{r}_{\rm e} \times \mathbf{p}_{\rm e} = -\mathbf{p}_{\rm e} \times \mathbf{r}_{\rm e}$), we can rewrite the first term of the interaction to obtain the symmetrized form

$$H_{\rm AM} = \frac{e}{2m_{\rm e}}\int_0^1 ds\, s\left[(\mathbf{r}_{\rm e} \times \mathbf{p}_{\rm e}) \cdot \mathbf{B}(s\mathbf{r}_{\rm e}) + \mathbf{B}(s\mathbf{r}_{\rm e}) \cdot (\mathbf{r}_{\rm e} \times \mathbf{p}_{\rm e})\right] + \frac{e^2}{2m_{\rm e}}\left[\mathbf{r}_{\rm e} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\rm e})\right]^2. \tag{9.105}$$

To further understand the first term, note that we can rewrite it as

$$H_{\rm AM} = -\frac{1}{2}\int d^3r \Big[\mathbf{M}^{\leftarrow}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r}) + \mathbf{B}(\mathbf{r})\cdot\mathbf{M}^{\rightarrow}(\mathbf{r})\Big] + \frac{e^2}{2m_{\rm e}}\left[\mathbf{r}_{\rm e}\times\int_0^1 ds\, s\,\mathbf{B}(s\mathbf{r}_{\rm e})\right]^2,$$

(atom–B-field coupling) (9.106)

where the two quantum magnetizations are

$$\begin{aligned}\mathbf{M}^{\leftarrow}(\mathbf{r}) &= -e\left(\mathbf{r}_{\rm e}\times\frac{\mathbf{p}_{\rm e}}{m_{\rm e}}\right)\int_0^1 ds\, s\,\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e})\\ \mathbf{M}^{\rightarrow}(\mathbf{r}) &= -e\int_0^1 ds\, s\,\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e})\left(\mathbf{r}_{\rm e}\times\frac{\mathbf{p}_{\rm e}}{m_{\rm e}}\right).\end{aligned}$$

(quantum atomic magnetizations) (9.107)

This brings the first part of the magnetic interaction into the same form as the electric interaction (9.103). Note that if we identify $\mathbf{p}_{\rm e} \longrightarrow m_{\rm e}\dot{\mathbf{r}}_{\rm e}$, both of these reduce to the classical magnetization,

$$\mathbf{M}(\mathbf{r}) = -e\mathbf{r}_{\rm e}\times\dot{\mathbf{r}}_{\rm e}\int_0^1 ds\, s\,\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e}), \tag{9.108}$$

from Eq. (9.81) above. Of course, as we mentioned above, $\mathbf{p}_{\rm e}$ is *not* the kinematic momentum, but rather is given by Eq. (9.98). Thus, the magnetic field couples to an atomic quantity that is close to, but not exactly, the classical momentum. Further, because the magnetic field couples to the atomic momentum (or rather, the angular momentum), the symmetric ordering is important in the above interaction.

The second term in the magnetic interaction Hamiltonian (9.105) is quadratic in the magnetic field, and we can interpret it as a diamagnetic energy of the atom in the magnetic field.

9.5.4 Electric Multipole Expansion

Now we will effect the expansion of the electric-field interaction Hamiltonian (9.103) into multipole moments. The self-interaction term can also be expanded, but often it is dropped and we will do this here, expanding only the remaining atom–field interaction:

$$H_{\rm AE} = -\int d^3r\,\mathbf{P}^{\perp}(\mathbf{r})\cdot\mathbf{E}^{\perp}(\mathbf{r}). \tag{9.109}$$

We begin by expanding the delta function in s, since we assume that the variation of the fields are slow over the length scale $\mathbf{r}_{\rm e}$:

$$\begin{aligned}\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e}) &= \delta^3(\mathbf{r}) - s\left[\mathbf{r}_{\rm e}\cdot\nabla\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e})\right]_{s=0} + \frac{s^2}{2}\left[(\mathbf{r}_{\rm e}\cdot\nabla)^2\,\delta^3(\mathbf{r}-s\mathbf{r}_{\rm e})\right]_{s=0} + \cdots\\ &= \delta^3(\mathbf{r}) - s\,(\mathbf{r}_{\rm e}\cdot\nabla)\,\delta^3(\mathbf{r}) + \frac{s^2}{2}\,(\mathbf{r}_{\rm e}\cdot\nabla)^2\,\delta^3(\mathbf{r}) + \cdots.\end{aligned} \tag{9.110}$$

This odd expression is sensible since we are in a sense *not* expanding the delta function, but rather the test function on which the delta function acts. Multiplying by an arbitrary test function $\mathbf{f}(\mathbf{r})$ and integrating leads to the usual series expansion for $\mathbf{f}(\mathbf{r}-s\mathbf{r}_{\rm e})$ about $s=0$. The polarization field then expands as

$$\mathbf{P}(\mathbf{r}) = -e\mathbf{r}_{\rm e}\delta^3(\mathbf{r}) + \frac{1}{2}e\mathbf{r}_{\rm e}\,(\mathbf{r}_{\rm e}\cdot\nabla)\,\delta^3(\mathbf{r}) - \frac{1}{6}e\mathbf{r}_{\rm e}\,(\mathbf{r}_{\rm e}\cdot\nabla)^2\,\delta^3(\mathbf{r}) + \cdots. \tag{9.111}$$

Thus, the interaction Hamiltonian becomes

$$H_{\rm AE} = e\mathbf{r}_{\rm e}\cdot\mathbf{E}^{\perp}(0) - \frac{e}{2}\,(\mathbf{r}_{\rm e}\cdot\nabla)\,\mathbf{r}_{\rm e}\cdot\mathbf{E}^{\perp}(0) + \frac{e}{6}\,(\mathbf{r}_{\rm e}\cdot\nabla)^2\,\mathbf{r}_{\rm e}\cdot\mathbf{E}^{\perp}(0) + \cdots, \tag{9.112}$$

where note that the gradients operate on the electric fields. Then we can write the Hamiltonian in terms of the multipole moments as

$$H_{\rm AE} = -d_\alpha E^{\perp}_\alpha(0) + Q_{\alpha\beta}\partial_\alpha E^{\perp}_\beta(0) - O_{\alpha\beta\gamma}\partial_\alpha\partial_\beta E^{\perp}_\gamma(0) + \cdots,$$

(electric multipole expansion) (9.113)

where note that there are implied summations over repeated indices. Here, the electric dipole moment is as expected

$$d_\alpha := -e r_{\mathrm{e},\alpha}, \tag{9.114}$$
(electric dipole moment)

while the **electric quadrupole moment** is

$$Q_{\alpha\beta} := -\frac{1}{2} e \left(r_{\mathrm{e},\alpha} r_{\mathrm{e},\beta} - \frac{r_\mathrm{e}^2}{3} \delta_{\alpha\beta} \right). \tag{9.115}$$
(electric quadrupole moment)

The first term in the quadrupole operator follows directly from (9.112), while the Kronecker-delta term is added to remove the traces of the moments, $Q_{\alpha\alpha} = 0$. The trace vanishes since $\nabla \cdot \mathbf{E}^\perp = 0$, which in index notation is $\partial_\alpha E_\alpha^\perp = 0$, and thus implies $\delta_{\alpha\beta} \partial_\alpha E_\beta^\perp = 0$. Note that the quadrupole moment as written here is the irreducible, rank-2 part of the symmetric, Cartesian tensor $(-e r_{\mathrm{e},\alpha} r_{\mathrm{e},\beta}/2)$ [see Eq. (7.206)]. Finally, the **electric octupole moment** is

$$O_{\alpha\beta\gamma} := -\frac{1}{6} e \left[r_{\mathrm{e},\alpha} r_{\mathrm{e},\beta} r_{\mathrm{e},\gamma} - \frac{r_\mathrm{e}^2}{5} \left(r_{\mathrm{e},\alpha} \delta_{\beta\gamma} + r_{\mathrm{e},\beta} \delta_{\gamma\alpha} + r_{\mathrm{e},\gamma} \delta_{\alpha\beta} \right) \right].$$
(electric octupole moment) (9.116)

Again, the first term is the physically important part, while the Kronecker deltas ensure that the tensor is traceless, giving the irreducible, rank-3 part[20] of the symmetric, Cartesian tensor $(-e r_{\mathrm{e},\alpha} r_{\mathrm{e},\beta} r_{\mathrm{e},\gamma}/6)$. The first two traces vanish, $O_{\alpha\beta\alpha} = O_{\beta\alpha\alpha} = 0$, in the same way as for the quadrupole moment. Note that for an octupole interaction with a *longitudinal field* of the form $O_{\alpha\beta\gamma} \partial_\alpha \partial_\beta \partial_\gamma \phi$ in terms of the scalar potential, the remaining trace $O_{\alpha\alpha\beta}$ also vanishes by permutation symmetry of the indices. However, this is not the case for the interaction with the *transverse* field, where in general $\partial_\alpha \partial_\alpha E_\gamma^\perp(0)$ is nonvanishing. However, for a monochromatic interaction (as appropriate near resonance, when driving a narrow octupole transition), the electric field obeys the Helmholtz equation, and thus $\partial_\alpha \partial_\alpha E_\gamma^\perp(0) = -k^2 E_\gamma^\perp(0)$. Thus, this trace of the octupole moment leads to an interaction of the form $-O_{\alpha\alpha\beta} \partial_\alpha \partial_\alpha E_\beta^\perp(0) = -e(k^2 r_\mathrm{e}^2/6) r_{\mathrm{e},\beta} E_\beta^\perp(0)$. This is of the same form as the *dipole* interaction, so we should also remove this trace from the octupole moment, and regard it as a correction to the dipole operator, which should thus have the form $d_\alpha = -e(1 + k^2 r_\mathrm{e}^2/6) r_{\mathrm{e},\alpha}$, which now has a small correction at the level of only a part in 10^6.

Due to the presence of additional factors of $\mathbf{r}_\mathrm{e}$ with derivative operators, the quadrupole interaction is weaker than the dipole interaction by a factor $k r_\mathrm{e}$, where $k = \omega/c$ is the optical wave number. Generally this factor is small for optical transitions ($k r_\mathrm{e} \ll 1$): for the D_2 transtion of ^{87}Rb, for example, with $\lambda = 780$ nm and $r_\mathrm{e} \sim 2a_0$ (a_0 is the Bohr radius), $k r_\mathrm{e} \sim 0.0051$. The octupole interaction is a factor of $k r_\mathrm{e}$ weaker yet than the quadrupole term. However, octupole transitions have been driven in experiments.[21]

9.5.5 Magnetic Multipole Expansion

Now expanding the magnetic field to lowest order in $\mathbf{r}_\mathrm{e}$, we find that the magnetic-field interaction Hamiltonian (9.106) becomes

$$H_{\mathrm{AM}} = \frac{e}{2m_\mathrm{e}} (\mathbf{r}_\mathrm{e} \times \mathbf{p}_\mathrm{e}) \cdot \mathbf{B}(0) + \frac{e^2}{8m_\mathrm{e}} \left[\mathbf{r}_\mathrm{e} \times \mathbf{B}(0) \right]^2. \tag{9.117}$$

Defining the magnetic dipole moment

$$\mathbf{m} := -\frac{e}{2m_\mathrm{e}} (\mathbf{r}_\mathrm{e} \times \mathbf{p}_\mathrm{e}), \tag{9.118}$$
(magnetic dipole moment)

[20] J. Jerphagnon, "Invariants of the Third-Rank Cartesian Tensor: Optical Nonlinear Susceptibilities," *Physical Review B* **2**, 1091 (1970) (doi: 10.1103/PhysRevB.2.1091).

[21] M. Roberts, P. Taylor, G. P. Barwood, P. Gill, H. A. Klein, and W. R. C. Rowley, "Observation of an Electric Octupole Transition in a Single Ion," *Physical Review Letters* **78**, 1876 (1997) (doi: 10.1103/PhysRevLett.78.1876).

we can then write the magnetic-interaction Hamiltonian in the dipole approximation as

$$H_{\text{AM}} = -\mathbf{m}\cdot\mathbf{B}(0) + \frac{1}{8m_{\text{e}}}\left[\mathbf{d}\times\mathbf{B}(0)\right]^2.$$

(atom–B-field interaction, magnetic dipole approximation) (9.119)

The first term is then the usual interaction of the magnetic dipole with the magnetic field, while the diamagnetic term appears as a coupling of the *electric* dipole moment with the magnetic field.

To compare the magnitude of the interaction with the electric multipoles, note that we can identify the magnitudes $B \sim E/c$ and $p_{\text{e}} \sim m_{\text{e}}ckr_{\text{e}}$, so that the magnetic-dipole interaction is of the order $er_{\text{e}}(kr_{\text{e}})E$, which is the same as the order of the electric quadrupole interaction.

The diamagnetic term, however, depends on r_{e}^2, but not on the wave number k, and thus its comparison to other terms depends on the strength of the field. Making the same order-of-magnitude replacements, the diamagnetic term is of order $e^2r_{\text{e}}^2E^2/m_{\text{e}}c^2$. This is of the same order as the magnetic-dipole term for a field strength satisfying $e\lambda E \sim m_{\text{e}}c^2$, which would result in a magnetic-dipole interaction energy $\sim(kr_{\text{e}})^2m_{\text{e}}c^2$, which is a very high (relativistic) energy. Hence for moderate (perturbative) field strengths, the diamagnetic term is negligible compared to the magnetic dipole term.

9.6 Center-of-Mass Röntgen Interaction

Thus far, we have ignored the center-of-mass motion of the atom, assuming the nucleus to be fixed at $\mathbf{r} = 0$. Motion of the center of mass generates additional multipole terms,[22] and we will now consider them here. Consistently accounting for the center-of-mass velocity is important, for example, in obtaining physical results for the angular distribution of photons radiated by a moving atom.[23]

9.6.1 Polarization

To include the center-of-mass motion in the Power–Zienau transformation, we generalize the atomic polarization field to the polarization due to an arbitrary system of point charges with respect to an arbitrary "atomic location" $\mathbf{r}_{\text{A}}$ (which we will take below to be the center of mass):

$$\mathbf{P}(\mathbf{r}) = \sum_\alpha q_\alpha(\mathbf{r}_\alpha - \mathbf{r}_{\text{A}})\int_0^1 ds\,\delta^3[\mathbf{r} - \mathbf{r}_{\text{A}} - s(\mathbf{r}_\alpha - \mathbf{r}_{\text{A}})]. \qquad (9.120)$$

(atomic polarization)

Here, q_α is the charge of the αth particle located at $\mathbf{r}_\alpha$. We will then use the unitary operator

$$U = \exp\left[-\frac{i}{\hbar}\int d^3r\,\mathbf{P}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r})\right] \qquad (9.121)$$

with this polarization to transform the suitably generalized minimal-coupling Hamiltonian

$$H = \sum_\alpha \frac{\left[\mathbf{p}_\alpha - q_\alpha\mathbf{A}(\mathbf{r}_\alpha)\right]^2}{2m_\alpha} + V(\mathbf{r}_\alpha) + \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right),$$

(minimal-coupling Hamiltonian, many particles) (9.122)

where m_α is the mass of particle α.

[22] E. A. Power and T. Thirunamachandran, "The Multipolar Hamiltonian in Radiation Theory," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **372**, 265 (1980).

[23] Martin Wilkens, "Spurious velocity dependence of free-space spontaneous emission," *Physical Review A* **47**, 671 (1993) (doi: 10.1103/PhysRevA.47.671); Martin Wilkens, "Significance of Röntgen current in quantum optics: Spontaneous emission of moving atoms," *Physical Review A* **49**, 570 (1994) (doi: 10.1103/PhysRevA.49.570).

9.6.2 Center-of-Mass Coordinates

We now introduce the usual center-of-mass coordinates as follows. The total atomic mass is

$$m_{\rm A} := \sum_\alpha m_\alpha, \tag{9.123}$$

and the center-of-mass coordinates are

$$\begin{aligned} \mathbf{r}_{\rm A} &:= \frac{1}{m_{\rm A}} \sum_\alpha m_\alpha \mathbf{r}_\alpha \\ \mathbf{p}_{\rm A} &:= \sum_\alpha \mathbf{p}_\alpha. \end{aligned} \tag{9.124}$$

Then the relative coordinates are

$$\begin{aligned} \bar{\mathbf{r}}_\alpha &:= \mathbf{r}_\alpha - \mathbf{r}_{\rm A} \\ \bar{\mathbf{p}}_\alpha &:= \mathbf{p}_\alpha - \frac{m_\alpha}{m_{\rm A}} \mathbf{p}_{\rm A}, \end{aligned} \tag{9.125}$$

so that

$$\sum_\alpha \bar{\mathbf{r}}_\alpha = 0, \qquad \sum_\alpha \bar{\mathbf{p}}_\alpha = 0. \tag{9.126}$$

Note that since the Coulomb binding potential is entirely composed of internal forces,

$$V(\mathbf{r}_\alpha) \equiv V(\bar{\mathbf{r}}_\alpha). \tag{9.127}$$

From the commutation relation for the standard coordinates,

$$[r_{\alpha j}, p_{\beta k}] = i\hbar \delta_{\alpha\beta} \delta_{jk}, \tag{9.128}$$

we can see that the relative positions commute with the center-of-mass momentum,

$$[\bar{r}_{\alpha j}, p_{{\rm A}k}] = \left[r_{\alpha j} - \frac{1}{m_{\rm A}} \sum_\gamma m_\gamma r_{\gamma j},\ \sum_\beta p_{\beta k} \right] = i\hbar\delta_{jk} - \sum_\alpha \frac{m_\alpha}{m_{\rm A}} i\hbar\delta_{jk} = 0, \tag{9.129}$$

and similarly the center-of-mass position commutes with the relative momenta,

$$[r_{{\rm A}j}, \bar{p}_{\alpha k}] = 0. \tag{9.130}$$

Thus, the center-of-mass coordinates act as an independent degree of freedom from the relative coordinates (which are themselves obey the above constraint equations which reduce the dimension of the relative-coordinate space). Furthermore, the commutation relation for the center-of-mass coordinates is

$$[r_{{\rm A}j}, p_{{\rm A}k}] = \left[\frac{1}{m_{\rm A}} \sum_\alpha m_\alpha r_{\alpha j}, \sum_\beta p_{\beta k} \right] = \frac{1}{m_{\rm A}} \sum_{\alpha,\beta} [m_\alpha r_{\alpha j}, p_{\beta k}] = \frac{1}{m_{\rm A}} \sum_{\alpha,\beta} m_\alpha i\hbar\delta_{\alpha\beta}\delta_{jk} = i\hbar\delta_{jk}, \tag{9.131}$$

as we would expect. Furthermore, it is interesting to note that the commutator of the relative coordinates reads

$$[\bar{r}_{\alpha j}, \bar{p}_{\beta k}] = [r_{\alpha j}, p_{\beta k}] - \frac{m_\alpha}{m_{\rm A}} [r_{{\rm A}j}, p_{{\rm A}k}] = i\hbar\delta_{jk} \left(\delta_{\alpha\beta} - \frac{m_\alpha}{m_{\rm A}} \right), \tag{9.132}$$

and thus the relative coordinates themselves are *not* canonical.

In these center-of-mass coordinates, the polarization (9.133) becomes

$$\mathbf{P}(\mathbf{r}) = \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \int_0^1 ds\, \delta^3(\mathbf{r} - \mathbf{r}_{\rm A} - s\bar{\mathbf{r}}_\alpha). \tag{9.133}$$

We can see that both the relative and center-of-mass positions are present here, so that moth the relative and center-of-mass momenta will become modified under the Power–Zienau transformation. Further, the minimal-coupling Hamiltonian in center-of-mass coordinates becomes simply

$$H = \sum_\alpha \frac{1}{2m_\alpha}\left[\frac{m_\alpha}{m_{\rm A}}\mathbf{p}_{\rm A} + \bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\alpha)\right]^2 + V(\bar{\mathbf{r}}_\alpha) + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right). \tag{9.134}$$

Partially multiplying out the momentum term gives

$$H = \frac{{p_{\rm A}}^2}{2m_{\rm A}} + \sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\alpha)\right]^2 - \sum_\alpha \frac{q_\alpha}{2m_{\rm A}}\left[\mathbf{p}_{\rm A}\cdot\mathbf{A}(\mathbf{r}_\alpha) + \mathbf{A}(\mathbf{r}_\alpha)\cdot\mathbf{p}_{\rm A}\right] + V(\bar{\mathbf{r}}_\alpha) + \frac{\epsilon_0}{2}\int d^3r \left(E^{\perp 2} + c^2 B^2\right), \tag{9.135}$$

so that the center-of-mass and relative components are separated, with an interaction term between the center-of-mass momentum and the vector potential evaluated at the particle locations.

9.6.3 Transformation: Electric Dipole Approximation

We must now transform the canonical momenta for the atom and field to obtain the center-of-mass multipole Hamiltonian. The electric field still transforms under the unitary Power–Zienau transformation (9.121) as in Eqs. (9.88), so that

$$U E^\perp_\beta(\mathbf{r}) U^\dagger = E^\perp_\beta(\mathbf{r}) - \frac{1}{\epsilon_0}P^\perp_\beta(\mathbf{r}). \tag{9.136}$$

Thus, the electric-field part of the multipole interaction Hamiltonian is exactly the same as before. The relative momentum also transforms essentially as we worked out before in Eqs (9.95):

$$U\bar{\mathbf{p}}_\alpha U^\dagger = \bar{\mathbf{p}}_\alpha + q_\alpha \mathbf{A}(\mathbf{r}_\alpha) + q_\alpha \bar{\mathbf{r}}_\alpha \times \int_0^1 ds\, s\, \mathbf{B}(\mathbf{r}_{\rm A} + s\bar{\mathbf{r}}_\alpha). \tag{9.137}$$

We must also transform the center-of-mass momentum, with the result (see Problem 9.5)

$$U\mathbf{p}_{\rm A} U^\dagger = \mathbf{p}_{\rm A} - q_{\rm A}\mathbf{A}(\mathbf{r}_{\rm A}) + \sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha) + \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \times \int_0^1 ds\, \mathbf{B}(\mathbf{r}_{\rm A} + s\bar{\mathbf{r}}_\alpha). \tag{9.138}$$

Here,

$$q_{\rm A} := \sum_\alpha q_\alpha \tag{9.139}$$

is the total atomic charge (which vanished for a neutral atom).

The transformation of the atomic part of the minimal-coupling Hamiltonian is thus rather involved. To simplify our discussion here, we will make the dipole approximation and neglect the variation of the fields over the scale of the atom. In particular, the momenta now transform as

$$U\bar{\mathbf{p}}_\alpha U^\dagger = \bar{\mathbf{p}}_\alpha + q_\alpha \mathbf{A}(\mathbf{r}_{\rm A}) + \frac{q_\alpha}{2}\bar{\mathbf{r}}_\alpha \times \mathbf{B}(\mathbf{r}_{\rm A}) \tag{9.140}$$

and

$$U\mathbf{p}_{\rm A} U^\dagger = \mathbf{p}_{\rm A} + \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \times \mathbf{B}(\mathbf{r}_{\rm A}) = \mathbf{p}_{\rm A} + \mathbf{d}\times\mathbf{B}(\mathbf{r}_{\rm A}), \tag{9.141}$$

where the dipole operator is defined by

$$\mathbf{d} := \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha, \tag{9.142}$$

(electric dipole moment)

as is consistent with our previous definitions.

In the dipole approximation, we also make the replacement $\mathbf{A}(\mathbf{r}_\alpha) \longrightarrow \mathbf{A}(\mathbf{r}_\mathrm{A})$ in the Hamiltonian, so that the minimal coupling Hamiltonian (9.135) becomes

$$H = \frac{{p_\mathrm{A}}^2}{2m_\mathrm{A}} + \sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\mathrm{A})\right]^2 - \frac{q_\mathrm{A}}{2m_\mathrm{A}}\left[\mathbf{p}_\mathrm{A}\cdot\mathbf{A}(\mathbf{r}_\mathrm{A}) + \mathbf{A}(\mathbf{r}_\mathrm{A})\cdot\mathbf{p}_\mathrm{A}\right] + V(\bar{\mathbf{r}}_\alpha) + \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right). \tag{9.143}$$

For a *neutral* atom, $q_\mathrm{A} = 0$, and thus we can ignore the $\mathbf{p}_\mathrm{A}\cdot\mathbf{A}(\mathbf{r}_\mathrm{A})$ terms to obtain

$$H = \frac{{p_\mathrm{A}}^2}{2m_\mathrm{A}} + \sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\mathrm{A})\right]^2 + V(\bar{\mathbf{r}}_\alpha) + \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right). \tag{9.144}$$

Thus, we may consider the transformations of the center-of-mass and relative momenta separately. Using Eq. (9.141), the transformation of the center-of-mass kinetic-energy part of the Hamiltonian is thus

$$\begin{aligned} U\frac{{p_\mathrm{A}}^2}{2m_\mathrm{A}}U^\dagger &= \frac{1}{2m_\mathrm{A}}\left[\mathbf{p}_\mathrm{A} + \mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})\right]^2 \\ &= \frac{{p_\mathrm{A}}^2}{2m_\mathrm{A}} + \frac{1}{2m_\mathrm{A}}\Big[\mathbf{p}_\mathrm{A}\cdot[\mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})] + [\mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})]\cdot\mathbf{p}_\mathrm{A}\Big] + \frac{1}{2m_\mathrm{A}}\left[\mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})\right]^2. \end{aligned} \tag{9.145}$$

Using Eq. (9.140), the relative-momentum part of the Hamiltonian transforms as

$$\begin{aligned} U\left[\sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha - q_\alpha\mathbf{A}(\mathbf{r}_\alpha)\right]^2\right]U^\dagger &= \sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha + \frac{q_\alpha}{2}\bar{\mathbf{r}}_\alpha\times\mathbf{B}(\mathbf{r}_\mathrm{A})\right]^2 \\ &= \sum_\alpha \frac{\bar{p}_\alpha^2}{2m_\alpha} - \mathbf{m}\cdot\mathbf{B}(\mathbf{r}_\mathrm{A}) + \sum_\alpha \frac{q_\alpha^2}{8m_\alpha}\left[\bar{\mathbf{r}}_\alpha\times\mathbf{B}(\mathbf{r}_\mathrm{A})\right]^2, \end{aligned} \tag{9.146}$$

where we have defined the magnetic dipole moment as

$$\mathbf{m} := \sum_\alpha \frac{q_\alpha}{2m_\alpha}(\bar{\mathbf{r}}_\alpha\times\bar{\mathbf{p}}_\alpha). \tag{9.147}$$
(magnetic dipole moment)

We can thus recognize the last two terms of Eq. (9.145) as the magnetic dipole and diamagnetic terms that we already discussed. We concluded that these were small compared to the electric-dipole interaction, so we will make the *electric* dipole approximation and drop these. We will similarly drop the last term of Eq. (9.145), which we can identify as a center-of-mass diamagnetic term, since it is quadratic in the atomic dipole moment and thus of higher order than the electric-dipole interaction.

From what remains, we can write the full, transformed Hamltonian as

$$\tilde{H} = H_\mathrm{A} + H_\mathrm{F} + H_\mathrm{AE} + H_\mathrm{R}, \tag{9.148}$$
(multipole Hamiltonian)

where the isolated atomic Hamiltonian is

$$H_\mathrm{A} = \frac{{p_\mathrm{A}}^2}{2m_\mathrm{a}} + \sum\frac{p_\alpha^2}{2m_\alpha} + V(\bar{\mathbf{r}}_\alpha), \tag{9.149}$$
(free-atom Hamiltonian)

the field Hamiltonian has its usual form,

$$H_\mathrm{F} = \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right), \tag{9.150}$$
(free-field Hamiltonian)

and we have the usual form for the electric-dipole interaction,

$$H_\mathrm{AE} = -\mathbf{d}\cdot\mathbf{E}^\perp(r_\mathrm{A}) + \frac{1}{2\epsilon_0}\int d^3r\,\left[\mathbf{P}^\perp(\mathbf{r})\right]^2, \tag{9.151}$$
(electric-dipole interaction)

where the polarization (9.133) becomes

$$\mathbf{P}(\mathbf{r}) = \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \delta^3(\mathbf{r} - \mathbf{r}_\mathrm{A}) = \mathbf{d}\,\delta^3(\mathbf{r} - \mathbf{r}_\mathrm{A}) \tag{9.152}$$

in the dipole approximation.

The new interaction, corresponding to the remaining term in (9.145) that we did not drop, is the **Röntgen interaction**,[24]

$$H_\mathrm{R} = \frac{1}{2m_\mathrm{A}}\Big[\mathbf{p}_\mathrm{A}\cdot[\mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})] + [\mathbf{d}\times\mathbf{B}(\mathbf{r}_\mathrm{A})]\cdot\mathbf{p}_\mathrm{A}\Big], \tag{9.153}$$

(Röntgen interaction, electric dipole approximation)

which gives the coupling energy of the *electric* dipole and the *magnetic* field: recall that under a Lorentz boost, electric and magnetic fields interchange to some extent, so this energy can be interpreted as the electric dipole interaction with the transformed magnetic field. Naturally, the *internal* motion of the charges should likewise induce a Röntgen-type coupling to the magnetic field; this is what we have already identified as the magnetic-dipole interaction, and in fact we can obtain the magnetic-dipole Hamiltonian from the Röntgen Hamiltonian by making the replacements $m_\mathrm{A} \longrightarrow m_\alpha$, $\mathbf{p}_\mathrm{A} \longrightarrow \mathbf{p}_\alpha$, both inside the summation implicit in the definition of $\mathbf{d}$.

Thus, even in the electric dipole approximation, an extra interaction Hamiltonian must be considered if the atom is in motion for consistency with the minimal-coupling Hamiltonian. This comes up, for example, in atom optics. Consider the usual problem of radiation pressure, where a plane, traveling wave impinges on an atom, and causes a net force in the propagation direction of the field to to absorption. If the atom is moving, the Röntgen interaction can add another component to the radiation-pressure force that is along the *polarization vector* for the field, instead of along its propagation direction.[25] Note, however, that this component is rather weak compared to the usual radiation pressure, and only occurs if the atomic dipole is not parallel to the field polarization (which is not possible for an S ground state of the resonant transition). The Röntgen interaction is generally weak, even though necessary for the consistency in the radiation of a moving atom as we mentioned above.

9.6.4 Full Transformation

So then what does the completely general multipole Hamiltonian look like, if we account for center-of-mass motion but *don't* make the electric-dipole approximation? You might regret that you asked that, but it is certainly possible to write down the answer.[26]

9.6.4.1 Effecting the General Transformation

Our goal is to perform the Power–Zienau transformation on the minimal-coupling Hamiltonian in the form (9.135). The part we should concentrate on here is the atomic-momentum part,

$$H = \frac{{p_\mathrm{A}}^2}{2m_\mathrm{A}} + \sum_\alpha \frac{1}{2m_\alpha}\left[\bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\alpha)\right]^2 - \sum_\alpha \frac{q_\alpha}{2m_\mathrm{A}}\left[\mathbf{p}_\mathrm{A}\cdot\mathbf{A}(\mathbf{r}_\alpha) + \mathbf{A}(\mathbf{r}_\alpha)\cdot\mathbf{p}_\mathrm{A}\right], \tag{9.154}$$

since our transformation for the field as above is still valid. The transformation (9.137) for the relative momentum leads to the following transformation for the relative-momentum term, which we can derive

[24] Named for Wilhelm Conrad Röntgen, who figured out that charges moving in a magnetic field see an effective electric field $\mathbf{E}_\mathrm{M} = \dot{\mathbf{r}}\times\mathbf{B}$. See W. C. Röntgen, "Ueber die durch Bewegung eines im homogenen electrischen Felde befindlichen Dielectricums hervorgerufene electrodynamische Kraft," *Annalen der Physik und Chemie* **35**, 264 (1888).

[25] V. E. Lembessis, M. Babiker, C. Baxter, and R. Loudon, "Theory of radiation forces and momenta for mobile atoms in light fields," *Physical Review A* **48** 1594 (1993) (doi: 10.1103/PhysRevA.48.1594).

[26] C. Baxter, M. Babiker and R. Loudon, "Canonical Approach to Photon Pressure," *Physical Review A* **47**, 1278 (1993) (doi: 10.1103/PhysRevA.47.1278).

simply by adapting our treatment from Section 9.5.3.3:

$$U\left[\sum_\alpha [\bar{\mathbf{p}}_\alpha - q_\alpha \mathbf{A}(\mathbf{r}_\alpha)]^2\right] U^\dagger = \sum_\alpha \frac{\bar{\mathbf{p}}_\alpha^{\,2}}{2m_\alpha} - \frac{1}{2}\int d^3r \Big[\mathbf{M}^{\leftarrow}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r}) + \mathbf{M}^{\rightarrow}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r})\Big] + \sum_\alpha \frac{q_\alpha^2}{2m_\alpha}\left[\bar{\mathbf{r}}_\alpha \times \int_0^1 ds\, s\, \mathbf{B}(\mathbf{r}_{\rm A} + s\bar{\mathbf{r}}_\alpha)\right]^2 . \tag{9.155}$$

The two quantum magnetizations are now suitably modified to include the center-of-mass coordinate to read

$$\begin{aligned}\mathbf{M}^{\leftarrow}(\mathbf{r}) &= \sum_\alpha q_\alpha \left(\bar{\mathbf{r}}_\alpha \times \frac{\bar{\mathbf{p}}_\alpha}{m_\alpha}\right)\int_0^1 ds\, s\, \delta^3(\mathbf{r} - \mathbf{r}_{\rm A} - s\bar{\mathbf{r}}_\alpha)\\ \mathbf{M}^{\rightarrow}(\mathbf{r}) &= \sum_\alpha q_\alpha \int_0^1 ds\, s\, \delta^3(\mathbf{r} - \mathbf{r}_{\rm A} - s\bar{\mathbf{r}}_\alpha)\left(\bar{\mathbf{r}}_\alpha \times \frac{\bar{\mathbf{p}}_\alpha}{m_\alpha}\right).\end{aligned} \tag{9.156}$$

We can recognize the terms in Eq. (9.155) as ones we have seen before: atomic kinetic energy, magnetic-field interaction with the atomic magnetization, and the diamagnetic energy.

On the other hand, the transformation (9.138) for the center-of-mass momentum leads to a slightly more complicated transformation for the Hamiltonian. First, we can consider the transformation of the center-of-mass kinetic energy:

$$U\frac{{p_{\rm A}}^2}{2m_{\rm A}}U^\dagger = \frac{1}{2m_{\rm A}}\left[\mathbf{p}_{\rm A} - q_{\rm A}\mathbf{A}(\mathbf{r}_{\rm A}) + \sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha) + \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \times \int_0^1 ds\, \mathbf{B}(\mathbf{r}_{\rm A} + s\bar{\mathbf{r}}_\alpha)\right]^2 . \tag{9.157}$$

Before multiplying this out, it simplifies things to consider the transformation of the remaining term in the Hamiltonian (9.154):

$$\begin{aligned}U\left[-\sum_\alpha \frac{q_\alpha}{2m_{\rm A}}\left[\mathbf{p}_{\rm A}\cdot\mathbf{A}(\mathbf{r}_\alpha) + \mathbf{A}(\mathbf{r}_\alpha)\cdot\mathbf{p}_{\rm A}\right]\right]U^\dagger = &-\frac{1}{2m_{\rm A}}\sum_\alpha\left[\mathbf{p}_{\rm A}\cdot\mathbf{A}(\mathbf{r}_\alpha) + \mathbf{A}(\mathbf{r}_\alpha)\cdot\mathbf{p}_{\rm A}\right]\\ &+\sum_\alpha \frac{q_{\rm A}q_\alpha}{m_{\rm A}}\mathbf{A}(\mathbf{r}_{\rm A})\cdot\mathbf{A}(\mathbf{r}_\alpha) - \frac{1}{m_{\rm A}}\left[\sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha)\right]^2\\ &-\frac{1}{m_{\rm A}}\sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha)\cdot\sum_\beta q_\beta\bar{\mathbf{r}}_\beta \times \int_0^1 ds\, \mathbf{B}(\mathbf{r}_{\rm A} + s\bar{\mathbf{r}}_\beta).\end{aligned} \tag{9.158}$$

These four terms exactly cancel the four terms in the expansion of (9.157) that involve the factor $\sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha)$.

Combining the center-of-mass parts of the transformed Hamiltonian thus gives

$$
\begin{aligned}
U &\left[\frac{{p_{\mathrm{A}}}^2}{2m_{\mathrm{A}}} - \sum_\alpha \frac{q_\alpha}{2m_{\mathrm{A}}}\left[\mathbf{p}_{\mathrm{A}}\cdot\mathbf{A}(\mathbf{r}_\alpha) + \mathbf{A}(\mathbf{r}_\alpha)\cdot\mathbf{p}_{\mathrm{A}}\right]\right]U^\dagger \\
&= \frac{{p_{\mathrm{A}}}^2}{2m_{\mathrm{A}}} - \frac{q_{\mathrm{A}}}{m_{\mathrm{A}}}\mathbf{p}_{\mathrm{A}}\cdot\mathbf{A}(\mathbf{r}_{\mathrm{A}}) + \frac{{q_{\mathrm{A}}}^2}{2m_{\mathrm{A}}}A^2(\mathbf{r}_{\mathrm{A}}) \\
&\quad + \frac{1}{2m_{\mathrm{A}}}\left[\mathbf{p}_{\mathrm{A}}\cdot\left(\sum_\alpha q_\alpha\bar{\mathbf{r}}_\alpha\times\int_0^1 ds\,\mathbf{B}(\mathbf{r}_{\mathrm{A}}+s\bar{\mathbf{r}}_\alpha)\right) + \left(\sum_\alpha q_\alpha\bar{\mathbf{r}}_\alpha\times\int_0^1 ds\,\mathbf{B}(\mathbf{r}_{\mathrm{A}}+s\bar{\mathbf{r}}_\alpha)\right)\cdot\mathbf{p}_{\mathrm{A}}\right] \\
&\quad - \frac{q_{\mathrm{A}}}{m_{\mathrm{A}}}\sum_\alpha q_\alpha\mathbf{A}(\mathbf{r}_{\mathrm{A}})\cdot\left[\bar{\mathbf{r}}_\alpha\times\int_0^1 ds\,\mathbf{B}(\mathbf{r}_{\mathrm{A}}+s\bar{\mathbf{r}}_\alpha)\right] + \frac{1}{2m_{\mathrm{A}}}\left[\sum_\alpha q_\alpha\bar{\mathbf{r}}_\alpha\times\int_0^1 ds\,\mathbf{B}(\mathbf{r}_{\mathrm{A}}+s\bar{\mathbf{r}}_\alpha)\right]^2 \\
&= \frac{{p_{\mathrm{A}}}^2}{2m_{\mathrm{A}}} - \frac{q_{\mathrm{A}}}{m_{\mathrm{A}}}\mathbf{p}_{\mathrm{A}}\cdot\mathbf{A}(\mathbf{r}_{\mathrm{A}}) + \frac{{q_{\mathrm{A}}}^2}{2m_{\mathrm{A}}}A^2(\mathbf{r}_{\mathrm{A}}) \\
&\quad + \frac{1}{2m_{\mathrm{A}}}\int d^3r\left[\mathbf{p}_{\mathrm{A}}\cdot\Big(\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\Big) + \Big(\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\Big)\cdot\mathbf{p}_{\mathrm{A}}\right] \\
&\quad - \frac{q_{\mathrm{A}}}{m_{\mathrm{A}}}\mathbf{A}(\mathbf{r}_{\mathrm{A}})\cdot\int d^3r\,\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r}) + \frac{1}{2m_{\mathrm{A}}}\left[\int d^3r\,\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\right]^2,
\end{aligned}
\tag{9.159}
$$

where we have used the polarization field in the form (9.133). We can identify the first three terms here as residual interaction of the center of mass with the vector potential, and the rest of the terms represent the generalized Röntgen interaction.

9.6.4.2 Final Result

Collecting all terms after the transformation, we have the new Hamiltonian

$$\tilde{H} = H_{\mathrm{A}} + H_{\mathrm{F}} + H_{\mathrm{AE}} + H_{\mathrm{AM}} + H_{\mathrm{R}}.$$

(general multipole Hamiltonian, with center-of-mass motion) (9.160)

The isolated atomic Hamiltonian

$$H_{\mathrm{A}} = \frac{{p_{\mathrm{A}}}^2}{2m_{\mathrm{A}}} + \sum_\alpha \frac{\bar{p}_\alpha^2}{2m_\alpha} + V(\bar{\mathbf{r}}_\alpha), \tag{9.161}$$

(free-atom Hamiltonian)

is the sum of the external and internal kinetic energies as well as the internal binding potential. The field Hamiltonian has its usual form,

$$H_{\mathrm{F}} = \frac{\epsilon_0}{2}\int d^3r\left(E^{\perp 2} + c^2B^2\right), \tag{9.162}$$

(free-field Hamiltonian)

while we have the same interaction Hamiltonian for the coupling of the atom to the electric field via the atomic polarization,

$$H_{\mathrm{AE}} = -\int d^3r\,\mathbf{P}^\perp(\mathbf{r})\cdot\mathbf{E}^\perp(\mathbf{r}) + \frac{1}{2\epsilon_0}\int d^3r\,\left[\mathbf{P}^\perp(\mathbf{r})\right]^2,$$

(atom–E-field interaction Hamiltonian) (9.163)

where the polarization $\mathbf{P}(\mathbf{r})$ is again given by Eq. (9.133). This Hamiltonian is again the generalization of the usual dipole interaction Hamiltonian, along with the electric self energy of the atom. The interaction Hamiltonian for the coupling of the internal degrees of freedom of the atom to the magnetic field is

$$H_{\mathrm{AM}} = -\frac{1}{2}\int d^3r\Big[\mathbf{M}^{\leftarrow}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r}) + \mathbf{M}^{\rightarrow}(\mathbf{r})\cdot\mathbf{B}(\mathbf{r})\Big] + \sum_\alpha\frac{q_\alpha^2}{2m_\alpha}\left[\bar{\mathbf{r}}_\alpha\times\int_0^1 ds\,s\,\mathbf{B}(\mathbf{r}_{\mathrm{A}}+s\bar{\mathbf{r}}_\alpha)\right]^2,$$

(atom–B-field interaction Hamiltonian) (9.164)

where the quantum magnetizations are given by Eqs. (9.156). The Hamiltonian here is the generalization of the magnetic-dipole interaction plus the diamagnetic energy of the atom in the magnetic field. Finally, the generalized Röntgen interaction coupling the center-of-mass motion of the atomic system to the field is

$$\begin{aligned} H_{\rm R} = & -\frac{q_{\rm A}}{m_{\rm A}}\mathbf{p}_{\rm A}\cdot\mathbf{A}(\mathbf{r}_{\rm A}) + \frac{{q_{\rm A}}^2}{2m_{\rm A}}A^2(\mathbf{r}_{\rm A}) \\ & + \frac{1}{2m_{\rm A}}\int d^3r\left[\mathbf{p}_{\rm A}\cdot\Big(\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\Big) + \Big(\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\Big)\cdot\mathbf{p}_{\rm A}\right] \\ & - \frac{q_{\rm A}}{m_{\rm A}}\mathbf{A}(\mathbf{r}_{\rm A})\cdot\int d^3r\,\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r}) + \frac{1}{2m_{\rm A}}\left[\int d^3r\,\mathbf{P}(\mathbf{r})\times\mathbf{B}(\mathbf{r})\right]^2, \end{aligned}$$

(center-of-mass (Röntgen) interaction Hamiltonian) (9.165)

which is quite a complicated interaction involving the atomic center-of-mass momentum, the vector potential, and the magnetic field. In our treatment above of the Röntgen interaction in the electric dipole approximation, we dropped all but the third term in the above expression for $H_{\rm R}$. Recalling that the total charge vanishes ($q_{\rm A} = 0$) for a neutral atom, we can see that the terms proportional to $q_{\rm A}$ vanish in the neutral-atom case, and the first two terms of $H_{\rm R}$ clearly represent a minimal-coupling-type interaction of the net atomic charge with the vector potential.

9.7 Exercises

Problem 9.1

Prove the relation

$$\int d^3r\, \mathbf{E}^{\perp} \cdot \mathbf{E}^{\parallel} = 0, \tag{9.166}$$

which we used in analyzing the minimal-coupling Hamiltonian.

Problem 9.2

Consider the Lagrangian

$$L = \frac{1}{2} m \dot{\mathbf{r}}^2 + q \dot{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r}) - q\phi(\mathbf{r}) \tag{9.167}$$

for a particle of charge q in electric and magnetic fields

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \qquad \mathbf{B} = \nabla \times \mathbf{A}, \tag{9.168}$$

written here in terms of the scalar potential ϕ and the vector potential $\mathbf{A}$.

Show that the Euler–Lagrange equation, together with this Lagrangian, is equivalent to Newton's Second Law with the Lorentz force.

Problem 9.3

Show that a (classical) Hamiltonian of the form

$$H = \frac{[\mathbf{p} + \mathbf{a}(\mathbf{q})]^2}{2m} + V(\mathbf{q}) \tag{9.169}$$

has canonical momentum given by

$$\mathbf{p} = m\dot{\mathbf{q}} - \mathbf{a}(\mathbf{q}) \tag{9.170}$$

through correspondence with the Lagrangian

$$L = \frac{1}{2} m \dot{q}^2 - \dot{\mathbf{q}} \cdot \mathbf{a}(\mathbf{q}) - V(\mathbf{q}). \tag{9.171}$$

Problem 9.4

Consider a single-electron atom in the heavy-nucleus approximation. Go through the derivation of the coupling of the atom to the electric field,

$$H_{\text{AE}} = -\int d^3r\, \mathbf{P}^{\perp}(\mathbf{r}) \cdot \mathbf{E}^{\perp}(\mathbf{r}) + \frac{1}{2\epsilon_0} \int d^3r\, \left[\mathbf{P}^{\perp}(\mathbf{r})\right]^2, \tag{9.172}$$

and the coupling of the atom to the magnetic field,

$$H_{\text{AM}} = -\frac{1}{2} \int d^3r \left[\mathbf{M}^{\leftarrow}(\mathbf{r}) \cdot \mathbf{B}(\mathbf{r}) + \mathbf{B}(\mathbf{r}) \cdot \mathbf{M}^{\rightarrow}(\mathbf{r})\right] + \frac{e^2}{2m_{\text{e}}} \left[\mathbf{r}_{\text{e}} \times \int_0^1 ds\, s\, \mathbf{B}(s\mathbf{r}_{\text{e}})\right]^2, \tag{9.173}$$

without making the long-wavelength approximation. You should also go through the derivation of the atomic polarization field $\mathbf{P}(\mathbf{r})$ and (classical) magnetization density $\mathbf{M}(\mathbf{r})$, paying special attention to their physical interpretation. These results generalize the results we derived in class for the dipole interaction Hamiltonian to include all multipole orders (but they still neglect center-of-mass motion of the atom and the presence of more than one electron.)

Problem 9.5

Carry out the Power–Zienau transformation of the center-of-mass momentum $\mathbf{p}_{\rm A}$, with transformation operator

$$U = \exp\left[-\frac{i}{\hbar}\int d^3r\, \mathbf{P}(\mathbf{r})\cdot\mathbf{A}(\mathbf{r})\right], \tag{9.174}$$

and polarization given by

$$\mathbf{P}(\mathbf{r}) = \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \int_0^1 ds\, \delta^3(\mathbf{r}-\mathbf{r}_{\rm A}-s\bar{\mathbf{r}}_\alpha), \tag{9.175}$$

to obtain the result

$$U\mathbf{p}_{\rm A}U^\dagger = \mathbf{p}_{\rm A} - q_{\rm A}\mathbf{A}(\mathbf{r}_{\rm A}) + \sum_\alpha q_\alpha \mathbf{A}(\mathbf{r}_\alpha) + \sum_\alpha q_\alpha \bar{\mathbf{r}}_\alpha \times \int_0^1 ds\, \mathbf{B}(\mathbf{r}_{\rm A}+s\bar{\mathbf{r}}_\alpha) \tag{9.176}$$

where $q_{\rm A}$ is the total atomic charge.

Problem 9.6

Consider a dipole-forbidden transition $|a\rangle \longrightarrow |b\rangle$ in an atom (i.e., $\langle a|\mathbf{d}|b\rangle = 0$).

(a) Assuming the transition may be driven by a monochromatic plane wave via the quadrupole interaction, show that the quadrupole interaction has the same form as the dipole interaction, with an effective dipole moment $d_{{\rm eff},\alpha} = -ik_\beta Q_{\alpha\beta}$. (Here, $\mathbf{k}$ is the wave vector of the plane wave.) Thus, the transition will be excited with an effective, quadrupole Rabi frequency

$$\Omega_Q = \frac{k_\alpha \langle a|Q_{\alpha\beta}|b\rangle (E_0)_\beta}{\hbar}, \tag{9.177}$$

(up to an arbitrary, overall phase that can be absorbed into the field), in analogy with the dipole Rabi frequency $\Omega = -\langle a|d_\alpha|b\rangle (E_0)_\alpha/\hbar$.

(b) Assuming that the plane wave propagates along the x-direction and is polarized along the z-direction, give an explicit expression for the relevant component of the effective dipole moment, in terms of a sum over products of dipole matrix elements (i.e., matrix elements like $\langle a|d_x|c\rangle$, involving some auxiliary state $|c\rangle$).

Problem 9.7

Go through the steps in the derivation [Eq. (9.71)] to show that for an atom of charge density

$$\rho = e\delta^3(\mathbf{r}) - e\delta^3(\mathbf{r}-\mathbf{r}_{\rm e}), \tag{9.178}$$

the polarization density

$$\mathbf{P}(\mathbf{r}) = -e\mathbf{r}_{\rm e}\int_0^1 ds\, \delta^3(\mathbf{r}-s\mathbf{r}_{\rm e}) \tag{9.179}$$

is consistent with the constraint $\nabla\cdot\mathbf{P}^{\parallel} = -\rho$.

Chapter 10

Cavity QED and the Jaynes–Cummings Model

Now we consider the simplest fully quantum model for the atom–field interaction: a two-level atom and a *single* mode of the electromagnetic field. As we will discuss, this model applies to an atom interacting with the field of an optical cavity in the ‘‘good-cavity’’ limit.

10.1 Single Cavity Mode

The uncoupled Hamiltonian for a two-level atom and a single mode of the optical field is

$$H_{\mathrm{A}} + H_{\mathrm{F}} = \hbar\omega_0 |\mathrm{e}\rangle\langle\mathrm{e}| + \hbar\omega\left(a^\dagger a + \frac{1}{2}\right), \tag{10.1}$$

where the ground state has zero energy, ω_0 is the atomic transition frequency, and ω is the cavity resonance frequency corresponding to the field mode. The dipole form of the atom–field interaction Hamiltonian is

$$H_{\mathrm{AF}} = -\mathbf{d}\cdot\mathbf{E}, \tag{10.2}$$

where the atomic dipole operator is

$$\mathbf{d} = \langle\mathrm{g}|\mathbf{d}|\mathrm{e}\rangle\Big(|\mathrm{g}\rangle\langle\mathrm{e}| + |\mathrm{e}\rangle\langle\mathrm{g}|\Big) =: \mathbf{d}_{\mathrm{ge}}\left(\sigma + \sigma^\dagger\right). \tag{10.3}$$

Here, $\sigma := |\mathrm{g}\rangle\langle\mathrm{e}|$ is the atomic lowering operator, and $\mathbf{d}_{\mathrm{ge}} := \langle\mathrm{g}|\mathbf{d}|\mathrm{e}\rangle$ is the dipole matrix element for the atomic transition. The (Heisenberg-picture) electric field mode of the cavity from Eq. (8.56) is

$$\mathbf{E}(\mathbf{r},t) = -\sqrt{\frac{\hbar\omega}{2\epsilon_0}}\left[\mathbf{f}(\mathbf{r})\,a(t) + \mathbf{f}^*(\mathbf{r})\,a^\dagger(t)\right], \tag{10.4}$$

where $\mathbf{f}(\mathbf{r})$ is the normalized spatial mode profile, and a is the mode annihilation operator. Thus, the interaction Hamiltonian becomes

$$H_{\mathrm{AF}} = -\sqrt{\frac{\hbar\omega}{2\epsilon_0}}\left(\sigma + \sigma^\dagger\right)\mathbf{d}_{\mathrm{ge}}\cdot\left[\mathbf{f}(\mathbf{r})\,a + \mathbf{f}^*(\mathbf{r})\,a^\dagger\right]. \tag{10.5}$$

Then defining the atom–field coupling energy

$$\hbar g(\mathbf{r}) := -\sqrt{\frac{\hbar\omega}{2\epsilon_0}}\mathbf{d}_{\mathrm{ge}}\cdot\mathbf{f}(\mathbf{r}) \tag{10.6}$$

(cavity QED coupling constant)

(g is called the **cavity QED coupling constant**, and $2g$ is called the **one-photon Rabi frequency**, as we will see below), the interaction Hamiltonian becomes

$$H_{\text{AF}} = \hbar \left(\sigma + \sigma^\dagger\right) \left[g(\mathbf{r})\, a + g^*(\mathbf{r})\, a^\dagger\right]. \tag{10.7}$$

At any given location $\mathbf{r}$, we may choose the phase of the atomic dipole such that g is a real and positive. In this case, the atom–field Hamiltonian becomes

$$H_{\text{AF}} = \hbar g \left(\sigma + \sigma^\dagger\right) \left(a + a^\dagger\right). \tag{10.8}$$

In the rotating-wave approximation, we drop the energy nonconserving terms (corresponding to fast-rotating terms in the Heisenberg picture), as we discussed before in Section 5.1.5.2, so that

$$H_{\text{AF}} = \hbar g \left(\sigma a^\dagger + \sigma^\dagger a\right), \tag{10.9}$$

thus keeping only terms corresponding to photon annihilation with atomic excitation, and atomic lowering with photon creation. Note that in view of the normalization

$$\int d^3r\, |\mathbf{f}(\mathbf{r})|^2 = 1, \tag{10.10}$$

if the mode is uniform over an optical cavity volume V (also called the **quantization volume**), then $|f(\mathbf{r})| = 1/\sqrt{V}$. In this case we can write the coupling constant (10.6) in terms of the mode volume as

$$g = -\hat{\varepsilon} \cdot \mathbf{d}_{\text{ge}} \sqrt{\frac{\omega}{2\epsilon_0 \hbar V}}, \tag{10.11}$$

where $\hat{\varepsilon}$ is the polarization vector of the field mode at the location of the atom. Thus, in general, it is common to define the coupling constant

$$g_0 = -\hat{\varepsilon} \cdot \mathbf{d}_{\text{ge}} \sqrt{\frac{\omega}{2\epsilon_0 \hbar V}}, \tag{10.12}$$

and then write the spatial dependence as

$$g(\mathbf{r}) = g_0 f'(\mathbf{r}), \tag{10.13}$$

where $f'(\mathbf{r})$ is a *dimensionless* mode profile, with maximum modulus of order unity. The coupling constant increases with decreasing cavity volume: this reflects the fact that locally, the electric field *for a single photon* increases as the confinement of the photon increases.

The total Hamiltonian is, dropping the vacuum-field energy, and regarding the atom to be fixed at position $\mathbf{r}$,

$$H = H_{\text{A}} + H_{\text{F}} + H_{\text{AF}} = \hbar\omega_0 \sigma^\dagger \sigma + \hbar\omega a^\dagger a + \hbar g \left(\sigma a^\dagger + \sigma^\dagger a\right). \tag{10.14}$$

(Jaynes–Cummings model)

This Hamiltonian defines the **Jaynes–Cummings model**:[1] it is the model for an atom interacting with a single, nearly resonant cavity mode within the RWA, ignoring any dissipation process such as spontaneous emission or any input or output from the cavity.

10.2 Dynamics

To investigate the dynamics of the Jaynes–Cummings model, we will decompose the state in terms of the joint eigenstates of H_{A} and H_{AF}:

$$|\psi\rangle = \sum_{n=0}^{\infty} \Big[c_{\text{g},n} |\text{g}, n\rangle + c_{\text{e},n} |\text{e}, n\rangle\Big]. \tag{10.15}$$

[1] E. T. Jaynes and F. W. Cummings, "Comparison of quantum and semiclassical radiation theories with application to the beam maser," *Proceedings of the IEEE* **51**, 89 (1963).

Putting this form of the state vector into the Schrödinger equation

$$\partial_t|\psi\rangle = -\frac{i}{\hbar}\left(H_{\rm A} + H_{\rm F} + H_{\rm AF}\right)|\psi\rangle \tag{10.16}$$

and projecting with $\langle \mathrm{e}, n|$ and $\langle \mathrm{g}, n+1|$ gives the coupled pairs of equations

$$\begin{aligned}\partial_t c_{\mathrm{e},n} &= -i(\omega_0 + n\omega)c_{\mathrm{e},n} - i\sqrt{n+1}\, g\, c_{\mathrm{g},n+1}\\ \partial_t c_{\mathrm{g},n+1} &= -i(n+1)\omega\, c_{\mathrm{g},n+1} - i\sqrt{n+1}\, g\, c_{\mathrm{e},n},\end{aligned} \tag{10.17}$$

where we have used the following form of the field annihilation operator:

$$a = \sum_{n=0}^{\infty} \sqrt{n+1}\,|n\rangle\langle n+1|. \tag{10.18}$$

The above structure of the Jaynes–Cummings model is important: only *pairs* of eigenstates are coupled, and thus the Hamiltonian is block diagonal, in 2×2 blocks, making it simple to diagonalize analytically.[2] Physically, the meaning here is that an excited atom can emit a photon into the cavity and reabsorb it, but that is the extent of the energy transfer. (Of course, then the vacuum amplitude $c_{\mathrm{g},0}$ is not coupled to any other amplitude, since no absorption can occur in the absence of light quanta.)

The above pair of equations (10.17) is formally equivalent to the *semiclassical* equations of motion for the atom–field interaction (the Rabi-flopping problem in the rotating frame), with Rabi frequency $2g\sqrt{n+1}$ and detuning

$$(n+1)\omega - (\omega_0 + n\omega) = \omega - \omega_0 = \Delta, \tag{10.19}$$

which is just the usual field detuning from the atomic resonance. Thus, we have already solved this problem. For example, an atom initially in the state $|\mathrm{g}, n+1\rangle$, coupled to a resonant field, has the populations

$$\begin{aligned}P_{\mathrm{e},n} &= \sin^2\left(\sqrt{n+1}\, gt\right) = \frac{1}{2}\left[1 - \cos\left(2\sqrt{n+1}\, gt\right)\right]\\ P_{\mathrm{g},n+1} &= \cos^2\left(\sqrt{n+1}\, gt\right) = \frac{1}{2}\left[1 + \cos\left(2\sqrt{n+1}\, gt\right)\right].\end{aligned} \tag{10.20}$$

(Rabi flopping)

Thus, the Rabi oscillations for $n+1$ energy quanta occur at angular frequency $2\sqrt{n+1}\, g$.[3] In particular, for a single photon, the Rabi oscillations occur at frequency $2g$: hence g is referred to as the single-photon Rabi frequency (though note the difference in convention of a factor of 2). For an off-resonant field, the Rabi oscillation proceed at the generalized Rabi frequency

$$\tilde{\Omega}_n = \sqrt{4(n+1)g^2 + \Delta^2}, \tag{10.21}$$

(quantized generalized Rabi frequency)

as in the semiclassical Rabi problem.

We can also consider this to be a crude model for spontaneous emission. The atom, initially in the state $|\mathrm{e}, 0\rangle$, oscillates to the state $|\mathrm{g}, 1\rangle$. This model has the rather odd result that a spontaneously emitted photon will be reabsorbed, because only a single frequency is present. For a general superposition of states, the Rabi oscillations occur for each pair of states at their respective Rabi frequencies. This can lead to complicated beating behavior between the many frequencies involved. For example, if we consider an initially excited

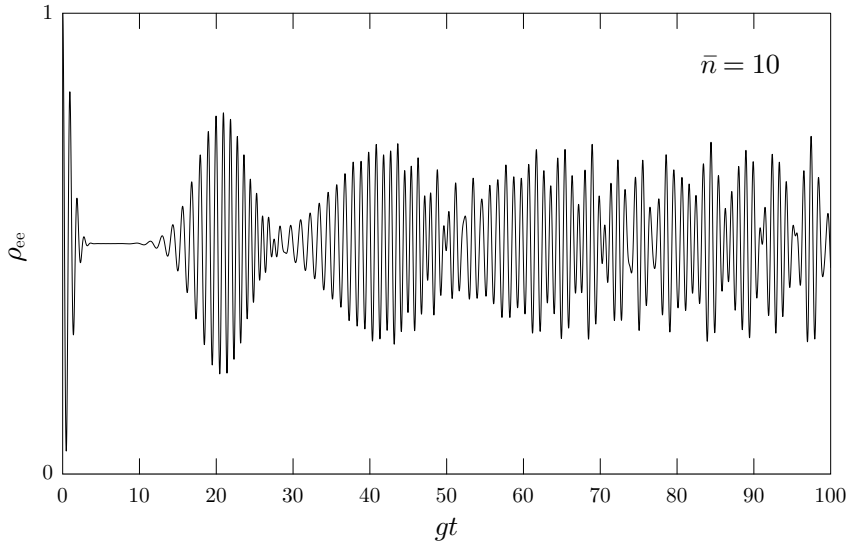
[2] It turns out to be possible to diagonalize the Hamiltonian also *without* the rotating-wave approximation, though the solution is considerably more complicated; see D. Braak, "Integrability of the Rabi Model," *Physical Review Letters* **107**, 100401 (2011) (doi: 10.1103/PhysRevLett.107.100401).

[3] The quantization and $\sqrt{n+1}$ dependence of the Rabi frequency has been observed with Rydberg atoms by M. Brune, F. Schmidt-Kaler, A. Maali, J. Dreyer, E. Hagley, J. M. Raimond, and S. Haroche, "Quantum Rabi Oscillation: A Direct Test of Field Quantization in a Cavity," *Physical Review Letters* **76**, 1800 (1996) (doi: 10.1103/PhysRevLett.76.1800).

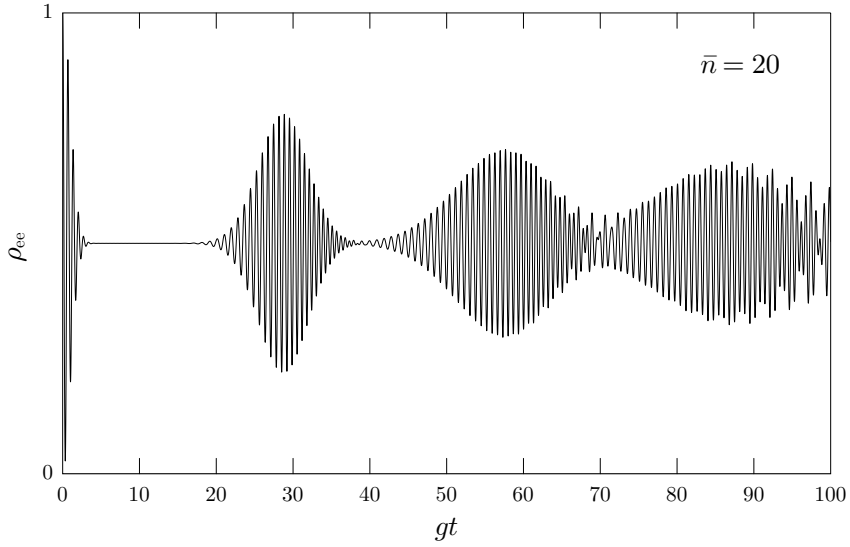
atom, where the field is in a superposition of states, then we no longer expect perfectly harmonic dynamics. For example, if the initial field state is the coherent state

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} e^{-|\alpha|^2/2} |n\rangle, \tag{10.22}$$

then we find **collapses and revivals** of the atomic population.[4] This phenomenon is illustrated in the plot below, which shows the excited-state population for an initially excited atom, and a cavity initially in a coherent state with mean photon number $\bar{n} = 10$ ($\alpha = \sqrt{10}$).



We see the Rabi oscillations quickly collapse, but then recur later. If the mean photon number is larger, $\bar{n} = 20$ ($\alpha = \sqrt{20}$), the recurrences occur at later times, and more coherent recurrences are visible.



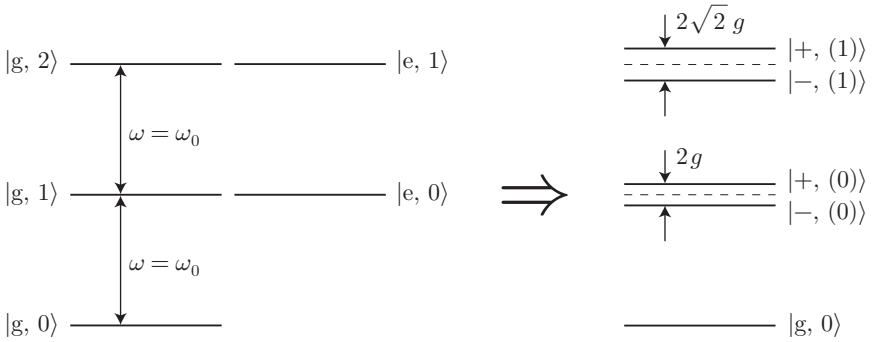
The decay of the Rabi oscillations still is not exponential, as we expect from the optical Bloch equations. The collapses and revivals are characteristic of a discrete spectrum, and in fact anytime we have a discrete spectrum, we have almost quasiperiodic behavior (when a finite number of states are significantly populated),

[4] The collapses and revivals were observed in a single-atom maser. See Gerhard Rempe, Herbert Walther, and Norbert Klein, "Observation of quantum collapse and revival in a one-atom maser," *Physical Review Letters* **58**, 353 (1987) (doi: 10.1103/PhysRevLett.58.353).

and we thus expect the state to eventually recur arbitrarily close to the initial condition. In this sense, spontaneous emission into a single field mode is *reversible*. The *irreversible*, exponential decay of the Rabi oscillations only occurs when the atom is coupled to a continuum of states, and we will show that this is the case later in Chapter 11.

10.3 Dressed States and the Vacuum Rabi Doublet

We can also take advantage of the block-diagonal structure that occurs due to the RWA in the Jaynes–Cummings Hamiltonian to define dressed states for the coupled atom–quantized field system. Each of the 2×2 blocks is again formally equivalent to the semiclassical Rabi problem, and thus the previous dressed-state results of Section 5.3 apply here. For an exactly resonant field mode ($\omega = \omega_0$), the bare states $|\mathrm{e}, n\rangle$ and $|\mathrm{g}, n+1\rangle$ are degenerate and coupled by a Rabi frequency $2g\sqrt{n+1}$. Thus, after diagonalizing the 2×2 blocks, the eigenstates are the dressed states $|+, (n)\rangle$ and $|-, (n)\rangle$, which have energies $(n+1)\omega \pm g\sqrt{n+1}$ (i.e., the splittings are $2g\sqrt{n+1}$).



For general detunings, the bare states are split in energy by the detuning $\hbar\Delta = \hbar\omega - \hbar\omega_0$ (with $|\mathrm{e}, n\rangle$ having higher energy than $|\mathrm{g}, n+1\rangle$ for $\Delta < 0$, and $|\mathrm{g}, n+1\rangle$ having the higher energy for $\Delta > 0$). The pairs of bare states are repeated every $\hbar\omega$ in energy, corresponding to having an additional photon around. Also, as before, in the general case the dressed states are defined by the rotation-type superposition

$$\begin{aligned} |+, (n)\rangle &= \sin\theta_n |\mathrm{g}, n+1\rangle + \cos\theta_n |\mathrm{e}, n\rangle \\ |-, (n)\rangle &= \cos\theta_n |\mathrm{g}, n+1\rangle - \sin\theta_n |\mathrm{e}, n\rangle, \end{aligned} \qquad \begin{matrix} (10.23) \\ \text{(dressed states)} \end{matrix}$$

where the Stückelberg angles θ_n are defined as before by

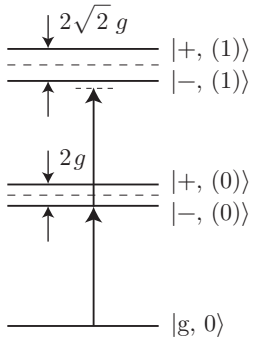
$$\tan 2\theta_n = -\frac{2g\sqrt{n+1}}{\Delta} \quad \left(0 \le \theta_n < \frac{\pi}{2}\right). \qquad \begin{matrix} (10.24) \\ \text{(Stückelberg angles)} \end{matrix}$$

Notice that within the rotating-wave approximation, the ground state $|\mathrm{g}, 0\rangle$ is completely uncoupled and does not experience a shift in the dressed-state basis. Recalling that in the classical limit ($n \gg 1$), the splittings for adjacent pairs of dressed states are nearly equal, giving rise to the Mollow triplet in resonance fluorescence, as we discussed in Section 5.7.4.2. However, this picture breaks down for small photon numbers where the splittings depend strongly on n. In particular, if a single photon interacts with an unexcited atom, there are only two possible transitions in the dressed-state basis, giving rise to the *vacuum Rabi doublet*. This becomes manifest, for example, for spectroscopy of a cavity tuned to the atomic resonance, as a doublet in the cavity transmission spectrum as a function of the frequency of a weak input field.[5]

[5]The vacuum Rabi splitting or ''normal-mode structure'' was observed for a single atom (the same atom throughout the entire spectral measurement) in a high-finesse cavity by A. Boca, R. Miller, K. M. Birnbaum, A. D. Boozer, J. McKeever, and H. J. Kimble, ''Observation of the Vacuum Rabi Spectrum for One Trapped Atom,'' *Physical Review Letters* **93**, 233603 (2004) (doi: 10.1103/PhysRevLett.93.233603); and P. Maunz, T. Puppe, I. Schuster, N. Syassen, P. W. H. Pinkse, and G. Rempe, ''Normal-Mode Spectroscopy of a Single-Bound-Atom–Cavity System,'' *Physical Review Letters* **94**, 033002 (2005) (doi: 10.1103/PhysRevLett.94.033002).

10.3.1 Photon Blockade

Due to the $\sqrt{n+1}$-dependence of the energy-level splittings, the atom-cavity system can act effectively as a highly nonlinear optical medium. One of the most striking demonstrations of this is the **photon blockade** effect.[6] The idea here is that under the right conditions, only a single photon can occupy the cavity at once. Suppose that the incident field is tuned to one of the sidebands of the vacuum Rabi doublet (say the red sideband), for the case when the cavity and atomic resonances coincide. This corresponds to a probe frequency $\omega_p = \omega - g$.



In this case, it is possible for a photon to enter the cavity, when the cavity state makes the transition $|g,0\rangle \longrightarrow |-,(0)\rangle$. However, if another photon is to enter the cavity, the atom must make the transition to the state $|-,(1)\rangle$. But the splitting for the two-photon manifold is $\sqrt{2}$ larger than for the one-photon manifold, and thus this second transition has a resonant frequency $(2\omega - \sqrt{2}\,g) - (\omega - g) = \omega - (\sqrt{2}-1)g$. Thus, the probe field is detuned by $(\sqrt{2}-2)\,g$ from the second transition. If g is large, much larger than the widths of these transitions (when we include dissipative effect), then the second transition is suppressed, and at most one photon occupies the cavity at once. The cavity output is thus antibunched, even for a coherent input light.

10.3.2 Atom-Photon "Molecule"

Of course, the splittings of the dressed states are space-dependent, and this can lead to dipole forces when the cavity photon number is nonzero in steady state. In fact, for large g, the dipole forces can be appreciable even for a cavity photon number of around unity. The dipole forces are sufficiently strong that they can trap an atom inside the cavity. This is something really remarkable: a single atom is bound by the dipole force to the field *due to a single photon*. This is called the "atom-photon molecule," since it represents a mechanically bound state of an atom and a photon.[7] The key to getting such high coupling strengths is to make the cavity mode volume V very small, which from Eq. (10.12) we see makes g large: the cavities in these experiments were spherical-mirror Fabry–Perot cavities with mirror separations of only 10 μm. (They also had very high reflectance coatings, and achieved near-record finesses in the 10^5 range.)

How does the atom get in there in the first place? Cesium atoms were cooled and trapped in a usual MOT, placed above the cavity, and then dropped. The density is low enough that only one atom at a time crosses through the cavity. You can actually *see* the atoms crossing through the cavity in real time, by tuning a weak probe laser to one of the vacuum-Rabi sidebands (normal modes or dressed-state transitions).

[6] L. Tian and H. J. Carmichael, "Quantum trajectory simulations of two-state behavior in an optical cavity containing one atom," *Physical Review A* **46**, R6801 (1992) (doi: 10.1103/PhysRevA.46.R6801); the term "photon blockade" was introduced and the effect clarified by A. Imamoğlu, H. Schmidt, G. Woods, and M. Deutsch, "Strongly interacting photons in a nonlinear cavity," *Physical Review Letters* **79**, 1467 (1997) (doi: 10.1103/PhysRevLett.79.1467). This effect was observed using a single atom trapped in a high-finesse microcavity by K. M. Birnbaum, A. Boca, R. Miller, A. D. Boozer, T. E. Northup, and H. J. Kimble, "Photon blockade in an optical cavity with one trapped atom," *Nature* **436**, 87 (2005) (doi: 10.1038/nature03804). The name "photon blockade" is in analogy to the **Coulomb blockade**, where charge transport through small devices happens one electron at a time, since only one electron can occupy an intermediate structure.

[7] C. J. Hood, T. W. Lynn, A. C. Doherty, A. S. Parkins, and H. J. Kimble, "The Atom-Cavity Microscope: Single Atoms Bound in Orbit by Single Photons," *Science* **287**, 1447 (2000) (doi: 10.1126/science.287.5457.1447). In these experiments and others by the same group, the coupling rate is in the range of $g/2\pi \sim 120$ MHz, compared to a cavity decay rate of $\kappa/2\pi \sim 40$ MHz and a spontaneous emission rate of $\Gamma/2\pi \sim 5$ MHz.

The splitting is zero when there is no atom in the cavity, and the probe is thus not transmitted. An atom crossing event is marked by a spike in the cavity transmission, at which time the probe intensity can simply be increased to trap the atom.

10.4 Refinements of the Model

For completeness, we will give some modifications to the Jaynes–Cummings model that make it more realistic. For the moment, we will just state them without proof, since they are quite reasonable, and we will defer the derivations of these results until Chapter 12.

The first modification that we can put in is atomic spontaneous emission, or the interaction of the atom with all the *other* modes, particularly outside the cavity. From the semiclassical treatment, we can simply tack on a Lindblad term to the master equation to obtain

$$\partial_t \rho = -\frac{i}{\hbar}[H, \rho] + \Gamma \mathcal{D}[\sigma]\rho, \tag{10.25}$$

where the Hamiltonian H is given in Eq. (10.14). This is pretty straightforward, although since the cavity modifies the local vacuum modes, the decay rate Γ may not be the same as the free-space value. The *enhancement* of the spontaneous-emission rate by a resonant cavity is called the **Purcell effect**,[8] but the spontaneous-emission rate can also be *suppressed* by an off-resonant cavity.

Along the same lines, the cavity intensity also decays. The cavity is just a harmonic oscillator, and we can use a decay term of the same form as for spontaneous emission. The total master equation then becomes

$$\partial_t \rho = -\frac{i}{\hbar}[H, \rho] + \Gamma \mathcal{D}[\sigma]\rho + \kappa \mathcal{D}[a]\rho, \tag{10.26}$$

where κ is the decay rate of the cavity energy (it is also common to use the convention where κ is the *field* decay rate, which would be smaller by a factor of 2). Recall that we already analyzed the harmonic oscillator with this type of damping in Section 5.6.1.2, where we saw that it is consistent with the classical damped oscillator (here, the classical damped oscillator being the classical damped cavity).

These damping rates can obviously have a strong effect on the cavity dynamics, particularly the Hamiltonian dynamics that we analyzed above. For the Jaynes–Cummings model to give a good approximation to the true dynamics, the atom–cavity system must be in the regime of **strong coupling**, where $g \gg \kappa, \Gamma$. In this case the dissipation is relatively slow, and the dynamics are Hamiltonian for short times. Note that "strong coupling" is sometimes also used for $g \gg \Gamma$, where the atomic emission is primarily into the cavity mode and not into other vacuum modes.

Finally, a source of energy is typically necessary for interesting atom–field interactions. Of course, the cavity can in principle be prepared in an arbitrary initial state, but damping will drive the system towards the vacuum steady state. It is also common to pump the cavity with an external classical field, to provide photons in steady state to drive the atom–field interaction. We can model this by adding a drive term to the Hamiltonian

$$H = H_{\mathrm{A}} + H_{\mathrm{F}} + H_{\mathrm{AF}} + H_{\mathrm{drive}} = \hbar\omega_0 \sigma^\dagger\sigma + \hbar\omega a^\dagger a + \hbar g\left(\sigma a^\dagger + \sigma^\dagger a\right) + \hbar\mathscr{E}\left(a e^{i\omega_{\mathrm{L}} t} + a^\dagger e^{-i\omega_{\mathrm{L}} t}\right), \tag{10.27}$$

where ω_{L} is the frequency of the classical field, $\mathscr{E} = \sqrt{\kappa P/\hbar\omega_{\mathrm{L}}}$, and P is the power of the driving laser. Note that this term has the same form as the semiclassical atom–field interaction in the optical Bloch equations. As we discussed before in Section 5.6.1.2, in the absence of the atom–cavity coupling, the field would settle down to a coherent state whose amplitude depends on $\mathscr{E}$ and the light–cavity detuning.

[8] E. M. Purcell, "Spontaneous Emission Probabilities at Radio Frequencies," *Physical Review* **69**, 681 (1946) (doi: 10.1103/PhysRev.69.674.2). This was first observed with Rydberg atoms in a superconducting cavity by P. Goy, J. M. Raimond, M. Gross, and S. Haroche, "Observation of Cavity-Enhanced Single-Atom Spontaneous Emission," *Physical Review Letters* **50**, 1903 (1983) (doi: 10.1103/PhysRevLett.50.1903).

10.5 Exercises

Problem 10.1

A simple model of damping of an optical cavity is via the interaction of the cavity mode with a beam of ground-state two-level atoms,[9] assuming that only one atom interacts with the cavity mode at a time, that the atoms interact briefly with the cavity mode, and that the rate at which atoms arrive is much faster than the cavity evolution. In this case, you will show that the reduced density operator for the cavity obeys the master equation

$$\partial_t \rho_{\rm c}(t) = -i\omega[a^\dagger a, \rho_{\rm c}] + \kappa \mathcal{D}[a]\rho_{\rm c}, \tag{10.28}$$

where $\rho_{\rm c}(t)$ is the reduced density operator for the cavity.

(a) Assume that the atom–cavity interaction occurs via the Jaynes–Cummings Hamiltonian. Assuming that the density operator at time t is of the form $\rho(t) = \rho_{\rm c}(t)|{\rm g}\rangle\langle{\rm g}|$, where $\rho_{\rm c}(t)$ is the reduced density operator for the cavity, compute an expression for $\rho(t+\tau)$ to first order in τ, and to *second* order in $g\tau$, since we will eventually take the limit as g becomes large. Assume that the cavity QED coupling rate g is constant during the interaction time τ. *Note:* there are fewer terms to work out in the interaction picture, but the Schrödinger picture will work as well.

(b) Then trace over the atomic state to obtain $\rho_{\rm c}(t+\tau)$.

(c) Finally, assume that τ is small, and that immediately after one atom leaves the cavity (after the interaction of strength g and time τ), the next one immediately follows and does the same thing. Then take the limit $\tau \longrightarrow 0$ (with $g \longrightarrow \infty$ in some sensible way) and write down a differential equation for $\rho_{\rm c}(t)$. What is κ?

Problem 10.2

A two-level atom initially in the superposition

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left[|{\rm g}\rangle + |{\rm e}\rangle\right] \tag{10.29}$$

passes slowly across the mode volume of an optical cavity. Assume the cavity to contain photons in only one field mode, which is tuned far off the atomic resonance, and ignore things like spontaneous emission and cavity decay. Explain qualitatively why a measurement of the phase of the atom after the interaction acts as a measurement of the cavity photon number. Indicate how the phase shift should scale with the photon number, assuming a small atom–field coupling rate.

Problem 10.3

In Problem 10.1, you worked out a simple model for the damping of a cavity due to a beam of two-level atoms. There, the resonant atom-cavity systems lead to energy dissipation in the *absorptive* regime. In this problem you will work out an example of cavity damping in the *dispersive* regime. Consider the same cavity, with a beam of two-level atoms crossing the cavity one at a time, treating the atom–cavity coupling via the Jaynes–Cummings model, but now where the atom-cavity detuning is large ($|\Delta| \gg g$).

(a) Recall that for very large detunings, we can identify the bare atom-cavity states $|{\rm e}, n\rangle$ $|{\rm g}, n+1\rangle$ with their dressed counterparts. Use what you know about the semiclassical and Jaynes–Cummings dressed states to write down the ac Stark shifts for these states due to an atom–cavity coupling rate g. Show that your answer agrees with the ac Stark shifts we derived for the two-level atom interacting with a classical field.

(b) As the atoms cross the cavity, the coupling rate $g(t)$ varies with time (going from zero to a maximum and back to zero). Assume that $g(t)$ changes slowly with time so that the dressed states are adiabatic

[9] For a literal realization of this setup for monitoring the state of a cavity field using a beam of Rydberg atoms, see Christine Guerlin, Julien Bernu, Samuel Deléglise, Clément Sayrin, Sébastien Gleyzes, Stefan Kuhr, Michel Brune, Jean-Michel Raimond, and Serge Haroche, "Progressive field-state collapse and quantum non-demolition photon counting," *Nature* **448**, 889 (2007) (doi: 10.1038/nature06057).

eigenstates, and then use the results of part (a) to argue that the atom–field interaction Hamiltonian can be replaced by the *effective* interaction

$$H_{\rm int} = \frac{\hbar g^2(t)}{\Delta}\left(a^\dagger a \sigma\sigma^\dagger - a a^\dagger \sigma^\dagger \sigma\right). \tag{10.30}$$

(c) Suppose that each atom starts in the initial state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|{\rm g}\rangle + e^{i\theta}|{\rm e}\rangle\right). \tag{10.31}$$

Show that under the effective interaction of part (b), the relative phase θ changes according to the cavity state, and thus realizes a (nondemolition) measurement of the cavity photon number. Write down an expression for $\delta\theta$, assuming the cavity to be in state $|n\rangle$.

(d) Finally, suppose one atom crosses the cavity in each time interval of duration τ, with the above initial state, but with θ random (what is the atomic density operator?). Assume τ to be long enough that $g(t)$ varies slowly, but fast compared to the cavity dynamics. Thus, you may trace over the atomic states and then formally take the limit $\tau \longrightarrow 0$ (keeping the lowest-order terms in the atom–field coupling necessary to obtain the simplest nontrivial result) to derive a master equation for the cavity state of the form

$$\partial_t \rho_{\rm c} = -\frac{i}{\hbar}[H_0, \rho_{\rm c}] + \gamma \mathcal{D}[a^\dagger a]\rho_{\rm c}. \tag{10.32}$$

Give a physical interpretation to the damping equations you find (especially regarding energy dissipation).

Chapter 11

Spontaneous Emission

11.1 Atom–Field Coupling

Here we will consider the spontaneous decay of an atomic excited level due to coupling to the vacuum field, according to the treatment of Weisskopf and Wigner[1]. We will also consider corrections and extensions to this result as well as the implications of this treatment for the spontaneous-emission master equation. The uncoupled Hamiltonian for a two-level atom and the field (including a sum over all field modes) is

$$H_0 = \hbar\omega_0 |\mathrm{e}\rangle\langle\mathrm{e}| + \sum_{\mathbf{k},\zeta} \hbar\omega_{\mathbf{k}} \left(a^\dagger_{\mathbf{k},\zeta} a_{\mathbf{k},\zeta} + \frac{1}{2} \right), \tag{11.1}$$

where the ground state has zero energy, ω_0 is the atomic transition frequency, the wave vector $\mathbf{k}$ labels the field modes of different frequency and orientation, the index ζ labels the two independent polarizations, and $a_{\mathbf{k},\zeta}$ is the annihilation operator for the $(\mathbf{k},\zeta)$ mode. We will write the eigenstates that we need of the free Hamiltonian in the form $|\alpha, n_{\mathbf{k},\zeta}\rangle$, which means that the atom is in state $|\alpha\rangle$, while the field mode $(\mathbf{k},\zeta)$ has n photons (other modes not explicitly labeled are in the vacuum state). We use the dipole form of the atom–field interaction Hamiltonian

$$H_{\mathrm{AF}} = -\mathbf{d}\cdot\mathbf{E}, \tag{11.2}$$

where as before the dipole operator is

$$\mathbf{d} = \langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle \Big(|\mathrm{g}\rangle\langle\mathrm{e}| + |\mathrm{e}\rangle\langle\mathrm{g}| \Big) =: \mathbf{d}_{\mathrm{ge}} \left(\sigma + \sigma^\dagger \right). \tag{11.3}$$

Here, $\sigma := |\mathrm{g}\rangle\langle\mathrm{e}|$ is the atomic lowering operator, and $\mathbf{d}_{\mathrm{ge}} := \langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle$ is the dipole matrix element for the atomic transition. We can write the electric field modes as

$$\mathbf{E}_{\mathbf{k},\zeta}(\mathbf{r}) = \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\, \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\, a_{\mathbf{k},\zeta} + \mathrm{H.c.}, \tag{11.4}$$

where the $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$ are normalized (classical) mode functions. Thus, the interaction Hamiltonian becomes

$$H_{\mathrm{AF}} = -\sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}} \left(\sigma + \sigma^\dagger \right) \mathbf{d}_{\mathrm{ge}} \cdot \left(\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) a_{\mathbf{k},\zeta} + \mathbf{f}^*_{\mathbf{k},\zeta}(\mathbf{r}) a^\dagger_{\mathbf{k},\zeta} \right). \tag{11.5}$$

[1] V. Weisskopf and E. Wigner, "Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie," *Zeitschrift für Physik* **63**, 54 (1930) (doi: 10.1007/BF01336768). See also Marlan O. Scully and M. Suhail Zubairy, *Quantum Optics* (Cambridge, 1997) (ISBN: 9780511813993) (doi: 10.1017/CBO9780511813993), p. 206; or Peter W. Milonni, *The Quantum Vacuum* (Academic Press, 1993) (ISBN: 0124980805) (doi: 10.1016/C2009-0-21295-5), p. 204.

In the rotating-wave approximation, we drop the energy nonconserving terms, so that

$$\begin{aligned} H_{\text{AF}} &= -\sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\, \mathbf{d}_{\text{ge}} \cdot \left(\sigma^\dagger a_{\mathbf{k},\zeta} \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) + \sigma a^\dagger_{\mathbf{k},\zeta} \mathbf{f}^*_{\mathbf{k},\zeta}(\mathbf{r})\right) \\ &= \sum_{\mathbf{k},\zeta} \hbar \left(g_{\mathbf{k},\zeta} \sigma^\dagger a_{\mathbf{k},\zeta} + g^*_{\mathbf{k},\zeta} \sigma a^\dagger_{\mathbf{k},\zeta}\right), \end{aligned} \tag{11.6}$$

where the coupling factor (one-photon Rabi frequency) is defined as

$$g_{\mathbf{k},\zeta}(\mathbf{r}) := -\sqrt{\frac{\omega_{\mathbf{k}}}{2\epsilon_0 \hbar}}\, \mathbf{d}_{\text{ge}} \cdot \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) \tag{11.7}$$

for each mode.

11.2 Evolution

Again, we will write the eigenstates that we need of the free Hamiltonian in the form $|\alpha, n_{\mathbf{k},\zeta}\rangle$, which means that the atom is in state $|\alpha\rangle$, while the field mode $(\mathbf{k}, \zeta)$ has n photons (other modes not explicitly labeled are in the vacuum state). We will assume that the atom is initially excited, and the field is in the vacuum state. According to the interaction (11.6), the only states coupled to the initial state are where the atom is in the ground state and one photon is present. Thus, we may write the state of the atom and field as

$$|\psi\rangle = c_{\text{e}}|\text{e}\rangle + \sum_{\mathbf{k},\zeta} c_{\mathbf{k},\zeta}|\text{g}, 1_{\mathbf{k},\zeta}\rangle. \tag{11.8}$$

The evolution is given by the Schrödinger equation,

$$\partial_t|\psi\rangle = -\frac{i}{\hbar}(H_0 + H_{\text{AF}})|\psi\rangle, \tag{11.9}$$

which gives the following coupled equations for the amplitudes:

$$\begin{aligned} \partial_t c_{\text{e}} &= -i\omega_0 c_{\text{e}} - i\sum_{\mathbf{k},\zeta} g_{\mathbf{k},\zeta} c_{\mathbf{k},\zeta} \\ \partial_t c_{\mathbf{k},\zeta} &= -i\omega_{\mathbf{k}} c_{\mathbf{k},\zeta} - i g^*_{\mathbf{k},\zeta} c_{\text{e}}. \end{aligned} \tag{11.10}$$

Now we will define slowly varying amplitudes

$$\tilde{c}_{\text{e}} := c_{\text{e}} e^{i\omega_0 t}, \qquad \tilde{c}_{\mathbf{k},\zeta} := c_{\mathbf{k},\zeta} e^{i\omega_{\mathbf{k}} t}, \tag{11.11}$$

in terms of which the equations of motion (11.10) become

$$\begin{aligned} \partial_t \tilde{c}_{\text{e}} &= -i\sum_{\mathbf{k},\zeta} g_{\mathbf{k},\zeta} \tilde{c}_{\mathbf{k},\zeta} e^{-i(\omega_{\mathbf{k}} - \omega_0)t} \\ \partial_t \tilde{c}_{\mathbf{k},\zeta} &= -i g^*_{\mathbf{k},\zeta} \tilde{c}_{\text{e}} e^{i(\omega_{\mathbf{k}} - \omega_0)t}. \end{aligned} \tag{11.12}$$

Integrating the second equation, we find

$$\tilde{c}_{\mathbf{k},\zeta}(t) = -i g^*_{\mathbf{k},\zeta} \int_0^t dt'\, \tilde{c}_{\text{e}}(t')\, e^{i(\omega_{\mathbf{k}} - \omega_0)t'}. \tag{11.13}$$

We can use this in the first of Eqs. (11.12) to decouple the equations,

$$\partial_t \tilde{c}_{\text{e}} = -\sum_{\mathbf{k},\zeta} |g_{\mathbf{k},\zeta}|^2 \int_0^t dt'\, \tilde{c}_{\text{e}}(t')\, e^{-i(\omega_{\mathbf{k}} - \omega_0)(t - t')}, \tag{11.14}$$

so that now we must simply evaluate this expression to uncover the decay process.

11.3 Large-Box Limit

Now we can put in the explicit form of the coupling factor from Eq. (11.7):

$$\partial_t \tilde{c}_{\rm e} = -\frac{1}{2\epsilon_0 \hbar} \sum_{\mathbf{k},\zeta} |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \mathbf{d}_{\rm ge}|^2 \omega_{\mathbf{k}} |f_{\mathbf{k},\zeta}(\mathbf{r})|^2 \int_0^t dt'\, \tilde{c}_{\rm e}(t')\, e^{-i(\omega_{\mathbf{k}}-\omega_0)(t-t')}, \tag{11.15}$$

where we used

$$|\mathbf{d}_{\rm ge} \cdot \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2 = |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \mathbf{d}_{\rm ge}|^2 |f_{\mathbf{k},\zeta}(\mathbf{r})|^2 \tag{11.16}$$

and where $\hat{\varepsilon}_{\mathbf{k},\zeta}$ is the local polarization unit vector for the field mode. Also, note that for the spherically symmetric atom,

$$|\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \mathbf{d}_{\rm ge}|^2 = |\hat{z} \cdot \mathbf{d}_{\rm ge}|^2 = \frac{1}{3} d_{\rm ge}^2, \tag{11.17}$$

since $d^2 = e^2 r^2 = e^2(x^2 + y^2 + z^2)$. Note that this replacement holds even if the assumption of a spherically symmetric atom is *not* valid. Because we are summing the quantities $|\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \mathbf{d}_{\rm ge}|^2$ over all field modes, and the electromagnetic vacuum is isotropic, the final result can't be changed by a termwise average over the relative orientation of atom and field polarization.

Now we can put in the explicit $(\mathbf{k}, \zeta)$ modes. In free space, from Section 8.4.2, the mode functions are

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \frac{1}{\sqrt{V}} \hat{\varepsilon}_{\mathbf{k},\zeta} e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{11.18}$$

and so

$$|\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2 = \frac{1}{V}, \tag{11.19}$$

where V is the quantization volume, and summing over both polarizations simply gives an extra factor of 2:

$$\partial_t \tilde{c}_{\rm e} = -\frac{d_{\rm ge}^2}{3\epsilon_0 \hbar V} \sum_{\mathbf{k}} \omega_{\mathbf{k}} \int_0^t dt'\, \tilde{c}_{\rm e}(t')\, e^{-i(\omega_{\mathbf{k}}-\omega_0)(t-t')}. \tag{11.20}$$

The next step is to evaluate the wave-vector sum in the continuum limit.

When the box becomes large ($V \longrightarrow \infty$), the spacing between the modes becomes small. (We covered this procedure when working out the free-space commutators in Section 8.6.1.2, but we'll review it to keep this self-contained.) In this limit, an integral of a function is equivalent to a sum weighted by the mode spacings. Since the modes exist only for positive and negative k_α, we can write

$$\sum_{\mathbf{k}} f(\mathbf{k})\, \Delta k_x\, \Delta k_y\, \Delta k_z \longrightarrow \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z\, f(\mathbf{k}) \tag{11.21}$$

for an arbitrary function $f(\mathbf{k})$. Since

$$\Delta k_\alpha = \frac{2\pi}{L}, \tag{11.22}$$

we can thus make the formal replacement

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{(2\pi)^3} \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z, \tag{11.23}$$

where $V = L^3$. Thus, we can write the equation of motion as

$$\begin{aligned} \partial_t \tilde{c}_{\rm e} &= -\frac{d_{\rm ge}^2}{3\epsilon_0 \hbar (2\pi)^3} \int d^3k\, \omega_{\mathbf{k}} \int_0^t dt'\, \tilde{c}_{\rm e}(t')\, e^{-i(\omega_{\mathbf{k}}-\omega_0)(t-t')} \\ &= -\frac{d_{\rm ge}^2 c}{6\pi^2 \epsilon_0 \hbar} \int_0^{\infty} dk\, k^3 \int_0^t dt'\, \tilde{c}_{\rm e}(t')\, e^{-i(\omega_{\mathbf{k}}-\omega_0)(t-t')} \\ &= -\frac{d_{\rm ge}^2}{6\pi^2 \epsilon_0 \hbar c^3} \int_0^{\infty} d\omega\, \omega^3 \int_0^t dt'\, \tilde{c}_{\rm e}(t')\, e^{-i(\omega-\omega_0)(t-t')}, \end{aligned} \tag{11.24}$$

where we have now carried out the angular integrals in spherical coordinates, and we are using $\omega = \omega_{\mathbf{k}} = ck$. Note here the characteristic ω^3 dependence, due partially to the frequency dependence of the vacuum density of states and also due to the explicit frequency dependence of H_{AF}, which will become manifest as an ω_0^3 dependence of the spontaneous decay rate (see Section 11.6.1 for an explicit calculation of the spontaneous decay rate in terms of the density of states).

11.4 Decay Rate

To proceed, we can now note that $\tilde{c}_{\text{e}}(t')$ varies slowly on optical time scales. Also, ω^3 is slowly varying compared to the exponential factor in Eq. (11.24), which oscillates rapidly (at least for large times t) about zero except when $t \approx t'$ and $\omega \approx \omega_0$. Thus, we will get a negligible contribution from the ω integral away from $\omega = \omega_0$. Thus, we will make the replacement $\omega^3 \longrightarrow \omega_0^3$:

$$\partial_t \tilde{c}_{\text{e}} = -\frac{\omega_0^3 d_{\text{ge}}^2}{6\pi^2 \epsilon_0 \hbar c^3} \int_0^\infty d\omega \int_0^t dt' \, \tilde{c}_{\text{e}}(t') \, e^{-i(\omega-\omega_0)(t-t')}. \tag{11.25}$$

The same argument gives

$$\int_0^\infty d\omega \, e^{-i(\omega-\omega_0)(t-t')} \approx \int_{-\infty}^\infty d\omega \, e^{-i(\omega-\omega_0)(t-t')} = 2\pi\delta(t-t'). \tag{11.26}$$

We can see from this that our argument here about the exponential factor is equivalent to the Markovian approximation. Thus,

$$\begin{aligned} \partial_t \tilde{c}_{\text{e}} &= -\frac{\omega_0^3 d_{\text{ge}}^2}{3\pi\epsilon_0 \hbar c^3} \int_0^t dt' \, \tilde{c}_{\text{e}}(t') \, \delta(t-t') \\ &= -\frac{\omega_0^3 d_{\text{ge}}^2}{3\pi\epsilon_0 \hbar c^3} \frac{\tilde{c}_{\text{e}}(t)}{2}. \end{aligned} \tag{11.27}$$

Here, we have split the δ-function since the upper limit of the t' integral was t, in view of the original form (11.25) for the t' integral, where the integration limit is centered at the peak of the exponential factor. We can rewrite the final result as

$$\partial_t \tilde{c}_{\text{e}} = -\frac{\Gamma}{2} \tilde{c}_{\text{e}}, \tag{11.28}$$

where the spontaneous decay rate is given by

$$\Gamma := \frac{\omega_0^3 d_{\text{ge}}^2}{3\pi\epsilon_0 \hbar c^3}. \tag{11.29}$$

(spontaneous decay rate)

This decay rate is of course defined so that the *probability* decays exponentially at the rate Γ:

$$\tilde{c}_{\text{e}}(t) = \tilde{c}_{\text{e}}(0) \, e^{-\Gamma t/2}, \qquad |\tilde{c}_{\text{e}}(t)|^2 = |\tilde{c}_{\text{e}}(0)|^2 e^{-\Gamma t}. \tag{11.30}$$

Also, note that

$$\partial_t c_{\text{e}} = \left(-i\omega_0 - \frac{\Gamma}{2}\right) c_{\text{e}} \tag{11.31}$$

after transforming out of the slow variables. Thus, we now have the an expression for the spontaneous decay rate in terms of the atomic parameters, which comes from a fully quantum treatment of the atom–field interaction. Recall that we derived this same expression earlier in Eq. (5.242) by comparing the optical Bloch equations to the rate equations; that result was correct because the rate equations are valid for the thermal quantum state of the field.

The above argument, using Eq. (11.26) is actually a bit sloppy because it misses another effect. A more careful argument makes the Markovian approximation directly on Eq. (11.25), before making the $\omega \approx \omega_0$ replacement:

$$\partial_t \tilde{c}_{\text{e}} \approx -\frac{d_{\text{ge}}^2}{6\pi^2 \epsilon_0 \hbar c^3} \tilde{c}_{\text{e}}(t) \int_0^\infty d\omega \, \omega^3 \int_0^t dt' \, e^{-i(\omega-\omega_0)(t-t')}, \tag{11.32}$$

and then considers the integral

$$\int_0^t dt'\, e^{-i(\omega-\omega_0)(t-t')} = -i\frac{1-\cos[(\omega-\omega_0)t]}{\omega-\omega_0} + \frac{\sin[(\omega-\omega_0)t]}{\omega-\omega_0}. \tag{11.33}$$

The integral of the sin term over all frequencies is π for $t > 0$, and for long times this term is arbitrarily narrowly peaked in frequency as t increases, as in our previous argument. The imaginary term is zero for $\omega = \omega_0$, and for large t is well approximated by $-i/(\omega-\omega_0)$ for ω away from ω_0. Thus, within a frequency integral, we can make the replacement

$$\int_0^t dt'\, e^{-i(\omega-\omega_0)(t-t')} \longrightarrow -i\mathscr{P}\frac{1}{\omega-\omega_0} + \pi\delta(\omega-\omega_0), \tag{11.34}$$

where $\mathscr{P}$ denotes that the ω integral should handle the singularity at $\omega = \omega_0$ in the sense of a Cauchy principle value (Section 14.1.4.2). Making this replacement in Eq. (11.32), we have

$$\partial_t \tilde{c}_\mathrm{e} \approx -\frac{d_\mathrm{ge}^2}{6\pi^2\epsilon_0\hbar c^3}\tilde{c}_\mathrm{e}(t)\left[\pi\omega_0^3 - i\int_0^\infty d\omega\,\frac{\omega^3}{\omega-\omega_0}\right]. \tag{11.35}$$

This can be written

$$\partial_t \tilde{c}_\mathrm{e} \approx (-\Gamma/2 - i\delta\omega_0)\,\tilde{c}_\mathrm{e}(t) \tag{11.36}$$

where we have the constants

$$\Gamma = \frac{\omega_0^3 d_\mathrm{ge}^2}{3\pi\epsilon_0\hbar c^3}, \qquad \delta\omega_0 = -\frac{d_\mathrm{ge}^2}{6\pi^2\epsilon_0\hbar c^3}\int_0^\infty d\omega\,\frac{\omega^3}{\omega-\omega_0} = -\frac{\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3}{\omega-\omega_0}.$$

(decay rate and Lamb shift) (11.37)

The decay rate is the same rate (11.29) that we obtained before, but the frequency shift $\delta\omega_0$ of the transition frequency is the effect we missed (this also pops up in the course of solving Problem 11.4). This is the **Lamb shift**, which is an ac Stark shift due to the coupling of the excited state to the vacuum electromagnetic field modes. This quantity involves a divergent integral that must be handled carefully, and we will defer this procedure to Section 13.12. The Lamb shift can justifiably be ignored (despite being divergent here!), because ω_0 is typically assumed to be the *observed* frequency, which already includes effects like the Lamb shift. Also, note that the ground state also experiences a Lamb shift, but here we have ignored it by declining to work out the equation of motion for the ground-state amplitude (we have also ignored couplings of the ground state to other important states besides $|\mathrm{e}\rangle$).

So what *is* the decay rate, typically? If we assume an optical transition, $\omega_0/2\pi = 400$ THz, and a dipole matrix element of order $d_\mathrm{ge} \sim ea_0$, where the Bohr radius $a_0 \approx 0.5$ Å, then we get a decay rate of $\Gamma \sim 5\times10^6\ \mathrm{s}^{-1}$. This is a reasonably good estimate, although for the larger alkalis, the decay rate is slightly larger (around $30\times10^6\ \mathrm{s}^{-1}$), since the dipole matrix elements have larger magnitudes for these large atoms. However, this justifies our assertion that $\tilde{c}_\mathrm{e}$ is a slowly varying amplitude (slow compared to the optical frequency).

Finally, note that we are only treating the decay of two-level atoms here, whereas real atoms are more complicated. We already treated the case of degeneracy due to angular momentum in Section (7.3.7.4). There we considered the decay of the $J_\mathrm{g} \longrightarrow J_\mathrm{e}$ fine-structure transition (with J_e being the excited state as usual). Then the decay rate from sublevel $|J_\mathrm{e}\ m_\mathrm{e}\rangle \longrightarrow |J_\mathrm{g}\ m_\mathrm{g}\rangle$ is just given by Eq. (11.29) with the appropriate matrix element:

$$\Gamma_{J_\mathrm{g},m_\mathrm{g};J_\mathrm{e},m_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}|\langle J_\mathrm{g}\ m_\mathrm{g}|\mathbf{d}|J_\mathrm{e}\ m_\mathrm{e}\rangle|^2. \tag{11.38}$$

With the normalization conventions assumed for the reduced matrix element, we found

$$\Gamma_{J_\mathrm{g}J_\mathrm{e}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}\frac{2J_\mathrm{g}+1}{2J_\mathrm{e}+1}|\langle J_\mathrm{g}\|\mathbf{d}\|J_\mathrm{e}\rangle|^2$$

(spontaneous decay rate, fine-structure transition) (11.39)

for the decay rate in terms of the reduced dipole matrix element. This same formula also applies to hyperfine transitions.

11.5 Master Equation for Spontaneous Emission

In order to arrive at the usual damping part of the master equation for the atom due to spontaneous emission (i.e., the damping part of the optical Bloch equations), we want to consider the *reduced* density operator for the evolution of the atomic state, tracing over the state of the field. Here we will compute the individual matrix elements

$$\rho_{\alpha\beta} := \langle\alpha|\rho|\beta\rangle \tag{11.40}$$

for the atomic state.

The easiest matrix element to treat is the excited-level population,

$$\rho_{\rm ee} = c_{\rm e} c_{\rm e}^*. \tag{11.41}$$

Differentiating this equation and using (11.31) gives

$$\partial_t \rho_{\rm ee} = -\Gamma \rho_{\rm ee}. \tag{11.42}$$

The matrix element for the ground-state population follows from summing over all the other states:

$$\rho_{\rm gg} := \sum_\zeta \int d\mathbf{k}\, \tilde{c}_{\mathbf{k},\zeta} \tilde{c}^*_{\mathbf{k},\zeta}. \tag{11.43}$$

Notice that the states $|{\rm e}\rangle$ and $|{\rm g}\rangle$ are effectively degenerate, but when we eliminate the field, we want $|{\rm e}\rangle$ to have $\hbar\omega_0$ more energy than the ground state. The shortcut for doing this is to realize that the latter situation corresponds to the ''interaction picture'' with respect to the field, where we use the slowly varying ground-state amplitudes $\tilde{c}_{\mathbf{k},\zeta}$ (which have been boosted *down* in energy by $\hbar\omega_{\mathbf{k}}$ to where we expect the atomic ground state should be) but the standard excited-state amplitude $c_{\rm e}$. This explains why we use regular coefficients in Eq. (11.41) but the slow variables in Eq. (11.43). Since by construction $\rho_{\rm ee} + \rho_{\rm gg} = 1$,

$$\partial_t \rho_{\rm gg} = \Gamma \rho_{\rm ee}. \tag{11.44}$$

Finally, the coherences are

$$\rho_{\rm ge} = c_{\rm g} c_{\rm e}^*, \qquad \rho_{\rm eg} = \rho_{\rm ge}^*, \tag{11.45}$$

where we have extended the pure state to include a component in the ground state without spontaneously emitted light:

$$|\psi\rangle = c_{\rm e}|{\rm e}\rangle + c_{\rm g}|{\rm g}\rangle + \sum_{\mathbf{k},\zeta} c_{\mathbf{k},\zeta}|{\rm g}, 1_{\mathbf{k},\zeta}\rangle. \tag{11.46}$$

We introduce this change because otherwise we would have $\rho_{\rm ge} = 0$ for all time. (Note that the above conclusions go through with this change.) The corresponding equation of motion is

$$\partial_t \rho_{\rm ge} = c_{\rm g}\left(i\omega_0 - \frac{\Gamma}{2}\right) c_{\rm e}^* = \left(i\omega_0 - \frac{\Gamma}{2}\right)\rho_{\rm ge}. \tag{11.47}$$

The time derivatives of the $c_{\rm g}$ is zero here, because nothing in the Hamiltonian causes it to evolve.

Notice that what we have derived are exactly the same matrix elements generated by the master equation

$$\partial_t \rho = -\frac{i}{\hbar}[H_{\rm A}, \rho] + \Gamma \mathcal{D}[\sigma]\rho, \tag{11.48}$$

where once again the form of the Lindblad superoperator $\mathcal{D}[\sigma]\rho$ is given by

$$\mathcal{D}[c]\rho := c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right), \tag{11.49}$$

and the atomic Hamiltonian is

$$H_{\rm A} := \hbar\omega_0 |{\rm e}\rangle\langle{\rm e}|. \tag{11.50}$$

That is, the damping term here represents the same damping as in the optical Bloch equations.

11.6 Fermi's golden rule

Now we will rederive the spontaneous-emission rate from a more general approach that applies to *any* quantum decay problem or transition to a continuum of states. We will then recover the result (11.29) for the rate of spontaneous emission in free space. Later, in Section (14.3.10), we will also use this approach to see how the rate of spontaneous emission is modified in the presence of a macroscopic body, such as a mirror or a cavity.

Let's consider a transition from an initial state $|\mathrm{i}\rangle$ to a final state $|\mathrm{f}\rangle$, where $|\mathrm{i}\rangle$ and $|\mathrm{f}\rangle$ are eigenstates of some background Hamiltonian H_0. The transition is due to the *constant* perturbation Hamiltonian H_{int}. We assume the state of the system at $t = 0$ to be $|\mathrm{i}\rangle$, and then consider the effect of H_{int} for $t > 0$. We will solve this problem by noting that we have already solved it in Section 5.2.2. In the setup of that problem, we showed that a two-level atom interacting with a monochromatic field is equivalent to a two-state system interacting with a dc perturbation, after making the rotating-wave approximation and transforming to the rotating frame of the field. Thus, we will make some formal identifications with the two-level atom problem. First, we can identify the free atomic Hamiltonian $\tilde{H}_{\mathrm{A}}$ for the two-level atom in the rotating frame with the background Hamiltonian in the present problem,

$$\tilde{H}_{\mathrm{A}} = -\hbar\Delta|\mathrm{e}\rangle\langle\mathrm{e}| \quad \longleftrightarrow \quad H_0 = \hbar\omega_{\mathrm{fi}}|\mathrm{f}\rangle\langle\mathrm{f}|, \tag{11.51}$$

where $\hbar\omega_{\mathrm{fi}} := E_{\mathrm{f}} - E_{\mathrm{i}}$ is the energy difference of the initial and final states, and we take $E_{\mathrm{i}} = 0$. We can further identify the atom–field interaction Hamiltonian $\tilde{H}_{\mathrm{AF}}$ for the two-level atom with the perturbation Hamiltonian H_{int},

$$\tilde{H}_{\mathrm{AF}} = \frac{\hbar}{2}\left[\Omega^*\sigma + \Omega\sigma^\dagger\right] \quad \longleftrightarrow \quad H_{\mathrm{int}} = \langle\mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle\ |\mathrm{i}\rangle\langle\mathrm{f}| + \mathrm{H.c.}, \tag{11.52}$$

where we are taking the form of $\tilde{H}_{\mathrm{AF}}$ generalized to a complex Rabi frequency Ω, as in Eq. (5.406), and σ is the usual atomic lowering operator. Note that we ignore diagonal matrix elements of H_{int}, as is appropriate for the dipole-interaction Hamiltonian (i.e., we can absorb any diagonal matrix elements of H_{int} into H_0). We showed that the solution to the two-level atom problem is [Eq. (5.60)]

$$P_{\mathrm{e}}(t) = \frac{|\Omega|^2}{|\Omega|^2 + \Delta^2}\sin^2\left(\frac{1}{2}\sqrt{|\Omega|^2 + \Delta^2}\ t\right), \tag{11.53}$$

which gives the excitation probability given that the atom is initially in the ground state. We can map this solution to the current problem by making the identifications $|\mathrm{g}\rangle \longrightarrow |\mathrm{i}\rangle$, $|\mathrm{e}\rangle \longrightarrow |\mathrm{f}\rangle$, $\Delta \longrightarrow \omega_{\mathrm{if}}$, and $\hbar\Omega \longrightarrow 2\langle\mathrm{f}|H_{\mathrm{int}}|\mathrm{i}\rangle$. We will further note that we are treating the interaction as a weak perturbation, so that $|\Omega| \ll |\Delta|$. Thus, the solution to the present perturbation problem is the transition probability to $|\mathrm{f}\rangle$:

$$P_{\mathrm{f}}(t) = \frac{4|\langle\mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2}{\hbar^2\omega_{\mathrm{if}}^2}\sin^2\left(\frac{\omega_{\mathrm{if}}t}{2}\right), \qquad (t \geq 0). \tag{11.54}$$

Now consider the part of the above expression that depends on ω_{if}. It is normalized such that

$$\int_{-\infty}^{\infty} d\omega_{\mathrm{if}}\,\frac{1}{\omega_{\mathrm{if}}^2}\sin^2\left(\frac{\omega_{\mathrm{if}}t}{2}\right) = \frac{\pi t}{2}, \tag{11.55}$$

and the integrand is a localized function in ω_{if} with a width that scales as $1/t$. Thus, for large t, when the integrand becomes a very narrow function of frequency, we may replace the integrand by a delta function:

$$\frac{1}{\omega_{\mathrm{if}}^2}\sin^2\left(\frac{\omega_{\mathrm{if}}t}{2}\right) \longrightarrow \frac{\pi t}{2}\delta(\omega_{\mathrm{if}}). \tag{11.56}$$

Thus, Eq. (11.54) becomes

$$P_{\mathrm{f}}(t) = \frac{2\pi t}{\hbar^2}|\langle\mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(\omega_{\mathrm{if}}) = \frac{2\pi t}{\hbar}|\langle\mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(\hbar\omega_{\mathrm{if}}). \tag{11.57}$$

Thus, in the long-time limit, the excitation probability increases linearly with time. This is clearly valid only for "weak" perturbations and short times such that the excitation probability is small; on the other hand, we had to make the long-time assumption such that the time was long enough to justify the approximation by a delta function. We will return to these constraints below. In any case, the transition rate from $|\mathrm{i}\rangle$ to $|\mathrm{f}\rangle$ is simply the time derivative of the transition probability, and we can thus write the transition rate as

$$\Gamma_{\mathrm{i}\to\mathrm{f}} = \frac{2\pi}{\hbar}|\langle \mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(E_{\mathrm{i}} - E_{\mathrm{f}}). \tag{11.58}$$
(Fermi's golden rule)

The transition rate in this regime of intermediate times is time-independent, and this expression for the transition rate is **Fermi's golden rule**. This statement can also be regarded as a statement of energy conservation: transitions only occur when the energies of the initial and final states match.

The delta function in this expression really only makes sense under an integral over energies, since it represents the transition probability summed over a range of energies. This is a crucial point: it is the existence of a *continuum* of energy levels that causes the time-independent transition rate; otherwise, the transition rate to a discrete state oscillates in time, due to coherent Rabi flopping. Thus, suppose we consider the transition from $|\mathrm{i}\rangle$ to a continuous set $\mathcal{F}$ of states. Then we must sum over the transition rate to all final states $|\mathrm{f}\rangle \in \mathcal{F}$. We will carry out this sum only over a narrow range $(E_{\mathrm{i}} - \varepsilon/2, E_{\mathrm{i}} + \varepsilon/2)$ of final states, where ε defines a range of energies over which H_{int} is constant. Letting $n(E)$ denote the number of states with energy less than E, the sum over transition rates is

$$\begin{aligned}\Gamma_{\mathrm{i}\to\mathcal{F}} &= \frac{2\pi}{\hbar}\int_{n(E_{\mathrm{i}}-\varepsilon/2)}^{n(E_{\mathrm{i}}+\varepsilon/2)} dn(E')\,|\langle \mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(E_{\mathrm{i}} - E') \\ &= \frac{2\pi}{\hbar}\int_{E_{\mathrm{i}}-\varepsilon/2}^{E_{\mathrm{i}}+\varepsilon/2} dE'\,\rho(E')\,|\langle \mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(E_{\mathrm{i}} - E').\end{aligned} \tag{11.59}$$

Here, $\rho(E) := dn/dE$ is the **density of states**, or number of states per unit energy interval. Completing the integral and taking $E_{\mathrm{f}} = E_{\mathrm{i}}$, we arrive at an alternate form of Fermi's golden rule:

$$\Gamma_{\mathrm{i}\to\mathcal{F}} = \frac{2\pi}{\hbar}|\langle \mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\rho(E_{\mathrm{f}}). \tag{11.60}$$
(Fermi's golden rule)

In deriving this, we also had to assume that the density of states was approximately constant over the range of integration, which sets another upper bound on ε. For the delta-function approximation to hold, we needed that the frequency width of the function in the expression (11.56) must be small compared to $\varepsilon/\hbar$, and thus that $t \gg \hbar/\varepsilon$, quantifying the long-time constraint we mentioned above. The short-time constraint is that $\Gamma_{\mathrm{i}\to\mathcal{F}}t \ll 1$, but typically this expression is valid to much longer times by accounting explicitly for depletion of $|\mathrm{i}\rangle$: the decay rate holds so long as the decayed population in the states $\mathcal{F}$ do not influence the decay rate.

11.6.1 Free-Space Decay Rate

Now we show that Fermi's golden rule leads to the correct spontaneous-emission rate in free space. From the Weisskopf–Wigner treatment of spontaneous emission, the dipole interaction Hamiltonian takes the form [see Eq. (11.5)]

$$H_{\mathrm{AF}} = -\sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\left(\sigma + \sigma^\dagger\right)\mathbf{d}_{\mathrm{ge}}\cdot\left(\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})a_{\mathbf{k},\zeta} + \mathbf{f}^*_{\mathbf{k},\zeta}(\mathbf{r})a^\dagger_{\mathbf{k},\zeta}\right). \tag{11.61}$$

The field modes here are labeled as usual by the wave vector $\mathbf{k}$ and the polarization index $\zeta \in \{1, 2\}$, and in free space, from Section 8.4.2, the mode functions are given by

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \frac{1}{\sqrt{V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{11.62}$$

when quantized in the fictitious quantization volume V. Then considering the transition from the initial state $|\mathrm{e}\rangle$ to the set of final states of the form $|\mathrm{g}, 1_{\mathbf{k},\zeta}\rangle$, we can write down the matrix element

$$\langle \mathrm{e}|H_{\mathrm{AF}}|\mathrm{g}, 1_{\mathbf{k},\zeta}\rangle = \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0 V}}\,(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\mathbf{d}_{\mathrm{ge}})\,e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{11.63}$$

We can then compute the density of states as follows, considering only the final states with the atom in the ground state and one photon in some mode. If we assume a cubic quantization volume with $V = L^3$, then the wave vectors are constrained to be

$$k_{x,y,z} = \frac{2\pi n_{x,y,z}}{L} = \frac{2\pi n_{x,y,z}}{\sqrt[3]{V}}, \tag{11.64}$$

where the n_{x^j} are any integers, as a result of the periodic boundary conditions on the quantization box. Thus, in $\mathbf{k}$-space, the states form a cubic lattice with spacing $2\pi/\sqrt[3]{V}$. We can thus associate a cubic volume of $(2\pi)^3/V$ with each state in $\mathbf{k}$-space, where this mini-volume "surrounds" its particular state. Now the set of all states with energy less than E is given by the set of all states in $\mathbf{k}$-space that fall within a radius of $k_E = E/\hbar c$ of $\mathbf{k} = 0$. The volume of the sphere of this radius is $4\pi k_E^3/3$, and thus the number of states is given by dividing this volume by $(2\pi)^3/V$, and then multiplying by 2 to count independent polarizations:

$$n(E) = 2\frac{4\pi k_E^3}{3}\frac{V}{(2\pi)^3} = \frac{E^3 V}{3\pi^2\hbar^3 c^3}. \tag{11.65}$$

Then the density of states is

$$\rho(E) = \frac{dn}{dE} = \frac{E^2 V}{\pi^2\hbar^3 c^3}. \tag{11.66}$$

The relevant initial and final energy in this problem is $E = \hbar\omega_0$, being the energy of the initially excited atom, so that

$$\rho(E_{\mathrm{f}}) = \frac{\omega_0^2 V}{\pi^2\hbar c^3}. \tag{11.67}$$

Putting these pieces together in the golden rule (11.60) we find the rate of spontaneous emission in free space

$$\Gamma = \frac{\omega_0^3 |\langle \mathrm{g}|d|\mathrm{e}\rangle|^2}{3\pi\epsilon_0\hbar c^3}, \tag{11.68}$$

(spontaneous decay rate in free space)

upon taking $\omega_{\mathbf{k}} \approx \omega_0$, and taking $|\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\mathbf{d}_{\mathrm{ge}}|^2 = |\langle \mathrm{g}|d|\mathrm{e}\rangle|^2/3$ for a spherically symmetric atom. This result agrees with Eq. (11.29) from our previous Weisskopf–Wigner calculation.

11.7 Corrections to Exponential Decay

The above results of exponential decay of the atomic excited state is a universal result of unstable quantum systems. However, it is also an approximation, and under physically reasonable assumptions the exponential decay law fails for very short and very long times. This was first discussed by Khalfin on very general grounds.[2] Fonda, Ghirardi, and Rimini have given a comprehensive review, and we follow their treatment for short-time deviations from exponential decay.[3]

11.7.1 Short Times

For short times, we will show that the decay rate in fact vanishes. This result applies broadly, even beyond atomic-level decay. Let us denote the survival probability by

$$P(t) = |c(t)|^2, \tag{11.69}$$

[2] L. A. Khalfin, "Contribution to the Decay Theory of a Quasi-Stationary State," *Soviet Physics JETP* **6**, 1053 (1958).

[3] L. Fonda, G. C. Ghirardi, and A. Rimini, "Decay theory of unstable quantum systems," *Reports on Progress in Physics* **41**, 587 (1978) (doi: 10.1088/0034-4885/41/4/003).

where $c(t)$ is the amplitude of the initial state ($|\mathrm{e}\rangle$ in the spontaneous-emission problem). Then

$$P(0) = c(0) = 1. \tag{11.70}$$

Now let's consider the eigenstates of the Hamiltonian H. For a system with several degrees of freedom, there will in general be multiple eigenstates for a given energy, and we label them by an extra index a. The index a can, for example, represent the set of simultaneous eigenvalues of a complete set of observables commuting with each other and H, if the system has enough symmetry to permit this, but the existence of such observables is not required here. Then we can write the eigenstates as

$$H|E,a\rangle = E|E,a\rangle, \tag{11.71}$$

and the completeness relation becomes

$$\int dE \int da\, |E,a\rangle\langle E,a| = 1. \tag{11.72}$$

Now we can write the coefficient $c(t)$ in terms of the unitary time-evolution operator (assuming a time-independent system, as we have already done by assuming energy eigenstates) as

$$\begin{aligned} c(t) &= \langle\psi(0)|e^{-iHt/\hbar}|\psi(0)\rangle \\ &= \int_{-\infty}^{\infty} dE \int da\, \langle\psi(0)|e^{-iHt/\hbar}|E,a\rangle\langle E,a|\psi(0)\rangle \\ &= \int_{-\infty}^{\infty} dE\, \omega(E) e^{-iEt/\hbar}, \end{aligned} \tag{11.73}$$

where

$$\omega(E) := \int da\, |\langle E,a|\psi(0)\rangle|^2 \tag{11.74}$$

is the Fourier transform of $c(t)$. Now we make the reasonable assumption that the energies E are bounded from below (which happens, for example, if there is a ground state):

$$c(t) = \int_{E_{\min}}^{\infty} dE\, \omega(E) e^{-iEt/\hbar}, \tag{11.75}$$

Note that the integral here is *uniformly* convergent for all t, since $\omega(E) > 0$, and the integral converges by assumption at $t = 0$, so that

$$\int_{E_{\min}}^{\infty} dE\, |\omega(E)| \tag{11.76}$$

is convergent. (The integral for $c(t)$ is hence absolutely convergent.) Thus, we may extend the integral to negative times, and the integral is uniformly convergent for any $t \in \mathbb{R}$. Differentiating (11.75), we find

$$\frac{dc(t)}{dt} = -\frac{i}{\hbar} \int_{E_{\min}}^{\infty} dE\, \omega(E) E e^{-iEt/\hbar}. \tag{11.77}$$

If we assume finite average energy of the initial state,

$$\int_{E_{\min}}^{\infty} dE\, \omega(E) E < \infty, \tag{11.78}$$

then $dc(t)/dt$ exists and is continuous. In particular, $dc(t)/dt$ is continuous at $t = 0$. Now since $\omega(E)$ is real, we have the Fourier-transform property

$$c(-t) = c^*(t), \tag{11.79}$$

and then differentiating

$$P(t) = c(t)c(-t), \tag{11.80}$$

we find

$$\frac{dP(t)}{dt} = \frac{dc(t)}{dt}c(-t) + c(t)\frac{dc(-t)}{dt}. \tag{11.81}$$

Thus, using $\lim_{t\to 0} c(t) = 1$, we find

$$\lim_{t\to 0^+} \frac{dP(t)}{dt} = \lim_{t\to 0^+} \left[\frac{dc(t)}{dt}c(-t) + c(t)\frac{dc(-t)}{dt}\right] = \lim_{t\to 0^+} \left[\frac{dc(t)}{dt} - \frac{dc(t)}{dt}\right] = 0.$$

(short-time nonexponential decay) (11.82)

Thus, we have shown that the *slope* of $P(t)$ vanishes at $t = 0$. $P(t)$ is also symmetric about $t = 0$, and we can see that near $t = 0$, $1 - P(t) = O(t^2)$. By contrast, the exponential decay law requires a *negative* slope at $t = 0$, and so the true decay is always *slower than exponential* for very short times.

11.7.1.1 Quantum Zeno Effect

This slower-than exponential decay can lead to an interesting effect: if the decay is ''interrupted'' by a quantum measurement that distinguishes the initial state from the final states, the evolution is ''reset'' to $t = 0$. With continual, rapid measurements, the system never settles into the usual exponential decay, and the decay rate for the observed system is slower than without observation. This is the essence of the **quantum Zeno effect**.[4] To see this more explicitly, assume that for short times, the survival probability may be written as

$$P(t) \approx 1 - at^2 \tag{11.83}$$

for some constant $a > 0$. Then suppose we make a measurement at time Δt. A single quantum system is projected back into the initial state with probability $P(\Delta t)$, and an ensemble of identically prepared systems will have fraction $P(\Delta t)$ in the initial state after the measurement. In either case, the system is in the same initial state, and then we can repeat the process with more measurements spaced at intervals of duration Δt. After n such measurements, the survival probability is

$$P(n\,\Delta t) \approx \left[1 - a(\Delta t)^2\right]^n \approx 1 - na(\Delta t)^2. \tag{11.84}$$

Noting that the time is $t = n\Delta t$, we can write

$$P(t) \approx 1 - (a\Delta t)t. \tag{11.85}$$

The decay rate in the presence of the measurements has been modified from a to $a\,\Delta t$. As the frequency of the measurements increases, then $\Delta t \longrightarrow 0$ and thus the decay rate is also reduced to zero. Thus, decay is almost completely inhibited for sufficiently frequent measurements.

Over what time scale do we expect the exponential-decay law to be invalid? The relavent time scale is set by the emitted energy during the decay:

$$\tau_{\text{nonexp}} \sim \frac{h}{\Delta E}. \tag{11.86}$$

This sets the time scale over which the excited and ground states cannot be ''resolved.'' For an optical transition, this time scale is $2\pi/\omega_0$, or just the optical period. For visible-wavelength transitions, this time scale is only a couple of fs. This type of nonexponential decay is thus very difficult to observe (and has not been observed thus far), since optical detectors are typically far too slow.

[4] B. Misra and E. C. G. Sudarshan, ''The Zeno's paradox in quantum theory,'' *Journal of Mathematical Physics* **18**, 756 (1977) (doi: 10.1063/1.523304); C. B. Chiu, E. C. G. Sudarshan, and B. Misra, ''Time evolution of unstable quantum states and a resolution of Zeno's paradox,'' *Physical Review D* **16**, 520 (1977) (doi: 10.1103/PhysRevD.16.520).

11.7.2 Long Times

For long times, the decay is also nonexponential. A very general argument[5] appeals to the Paley–Wiener theorem, which states that if the frequency spectrum of $c(t)$ cuts off below some minimum frequency $\omega_{\min}$ (corresponding to a lower energy bound), then $c(t)$ must satisfy

$$\int_{-\infty}^{\infty} dt \, \frac{\left|\log|c(t)|\right|}{1+t^2} < \infty. \tag{11.87}$$

In particular, for the integral to converge, we must have

$$\left|\log|c(t)|\right| \sim t^q \tag{11.88}$$

for $q < 1$ at large times, in which case

$$P(t) = |c(t)|^2 \sim e^{-\alpha t^q}, \tag{11.89}$$

where $\alpha > 0$. Thus, the probability must decay *more slowly* than exponential at late times. In particular, for the two-level atom, an extension of the Weisskopf–Wigner calculation[6] shows that at late times, the decay goes as

$$P(t) \sim \left(\frac{\Gamma}{2\pi\omega_0^3}\right)^2 \frac{1}{t^4}, \tag{11.90}$$

(long-time nonexponential decay)

once the exponential part of the decay has damped away. (We will defer this calculation until Section 15.5.4.) Since $\Gamma/\omega_0 \ll 1$ for optical transitions, the correction is very small. We can estimate the crossover time by setting

$$\left(\frac{\Gamma}{2\pi\omega_0^3}\right)^2 \frac{1}{t^4} \sim e^{-\Gamma t}, \tag{11.91}$$

which gives $\Gamma t \sim 130$ for a typical ratio $\Gamma/\omega_0 \approx 1.2 \times 10^{-8}$. The correction here is *very* small.

[5] L. A. Khalfin, *op. cit.*; L. Fonda, *et al.*, *op. cit.*

[6] P. L. Knight and P. W. Milonni, "Long-Time Deviations from Exponential Decay in Atomic Spontaneous Emission Theory," *Physics Letters* **56A** 275 (1976) (doi: 10.1016/0375-9601(76)90306-6).

11.8 Exercises

Problem 11.1

The **Lamb–Dicke effect**[7] occurs as a narrowing of the radiation spectrum of an atom if it is confined to very small volumes (as can happen for a trapped ion or a neutral atom in an optical lattice). This effect is closely related to the Mössbauer effect[8] for scattering from atoms bound in solids. In this problem you will work out a fairly simple model for this effect.

(a) Assume that the center-of-mass component of the particle is described by the Hamiltonian

$$H_{\text{CM}} = \frac{p^2}{2m} + V(\mathbf{r}), \tag{11.92}$$

where $V(\mathbf{r})$ is the trapping potential. To model the emission process, assume the two-level atom is initially excited, and the field is initially in the vacuum state. Then the emission process for a photon into the $(\mathbf{k}, \zeta)$ mode is the transition

$$|\text{e}, \psi_n\rangle \longrightarrow |\text{g}, \psi_l, 1_{\mathbf{k},\zeta}\rangle, \tag{11.93}$$

where the $|\psi_n\rangle$ are the energy eigenstates of H_{CM} with eigenvalues E_n, and we are assuming that the vibrational state of the atom changes from $n \longrightarrow l$ during the transition. (We will always assume the vibrational energies are small compared to the optical-transition energy.) Assume that the field is quantized in free space, and show that the corresponding transition amplitude is

$$\langle \text{i}|H_{AF}|\text{f}\rangle = -\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \mathbf{d}_{\text{ge}}\langle\psi_n|e^{i\mathbf{k}\cdot\mathbf{r}}|\psi_l\rangle, \tag{11.94}$$

where $|\text{i}\rangle$ and $|\text{f}\rangle$ are the initial and final states, respectively.

(b) What are the different possible frequencies in the radiation spectrum?

(c) From the above arguments, the strength of the $n \longrightarrow l$ transition is proportional to the squared transition amplitude

$$S_{nl} := \left|\langle\psi_n|e^{i\mathbf{k}\cdot\mathbf{r}}|\psi_l\rangle\right|^2. \tag{11.95}$$

Show that the sum of the line strengths for all possible transitions from the initial state is independent of position. In doing so, you have shown that the total decay rate is independent of $V(\mathbf{r})$, and thus equal to the free-space value.

(d) Show that if the atomic wave packet is confined to a region much smaller than λ_0 in each direction, where λ_0 is the resonance wavelength, that the dominant emission line is for the $n \longrightarrow n$ vibrational transition.

Problem 11.2

By modifying the Weisskopf–Wigner derivation of the spontaneous emission rate, derive an expression for the spontaneous decay rate for a spherically symmetric atom a distance z from a perfect, planar mirror. To do this, you will need the half-space mode functions from Section 8.4.3. Also, be careful when taking the continuum limit to obtain the correct prefactor for the integral over $\mathbf{k}$-space.

Problem 11.3

An optical analogue to the quantum Zeno effect occurs in propagation through ideal polarizers. A photon can be regarded as a two-state quantum system, with state $|\psi\rangle = c_{\text{V}}|\text{V}\rangle + c_{\text{H}}|\text{H}\rangle$, where $|\text{V}\rangle$

[7] R. H. Dicke, "The Effect of Collisions upon the Doppler width of Spectral Lines," *Physical Review* **89**, 472 (1953) (doi: 10.1103/PhysRev.89.472).

[8] Rudolf L. Mössbauer, "Kernresonanzfluoreszenz von Gammastrahlung in Ir^{191}, *Zeitschrift für Physik* **151**, 124 (1958) (doi: 10.1007/BF01344210).

indicates vertical (linear) polarization, and $|\mathrm{H}\rangle$ indicates horizontal polarization. With idealized polarizers, we can model the action of a vertical polarizer via the projector $|\mathrm{V}\rangle\langle\mathrm{V}|$, and the action of a horizontal polarizer by $|\mathrm{H}\rangle\langle\mathrm{H}|$.

Note that the probability for a photon to transmit through two crossed polarizers is zero. However, there is a *nonzero* transmission probability if, for example, another polarizer is inserted between the crossed polarizers, but oriented at 45° with respect to them.

Derive an expression for the transmission probability if N such polarizers are inserted between the crossed polarizers, with the nth polarizer making an angle of $n\pi/2(N+1)$ radians with respect to the frontmost polarizer. (That is, the intermediate polarizers uniformly and gradually sweep the polarization through 90°.) Your expression should assume that the photon transmitted through the frontmost polarizer (i.e., the first of the crossed polarizers). Show that your expression converges to unity as $N \longrightarrow \infty$.

This can be interpreted as ''dragging'' the polarization through 90° by making many, slightly different measurements.

Problem 11.4

In this problem, you will work out the theory of spontaneous emission in the presence of a thermal electromagnetic field at temperature T.

(a) Use the formalism for the general Born–Markov master equation in Section 4.5, and apply it to the atom–field interaction *in the rotating-wave approximation* (as in the Weisskopf–Wigner derivation) to derive the master equation

$$\partial_t \rho = -\frac{i}{\hbar}[H_{\mathrm{A}} + H_{\mathrm{eff}}, \rho] + \Gamma[\bar{n}(\omega_0)+1]D[\sigma]\rho + \Gamma\bar{n}(\omega_0)D[\sigma^\dagger]\rho \tag{11.96}$$

for an atom interacting with the electromagnetic field at temperature T, where

$$H_{\mathrm{eff}} = \hbar[\Delta_0 + \Delta(T)]\sigma^\dagger\sigma \tag{11.97}$$

represents the energy shift of the atomic transition (after moving the shifted ground state to zero energy), with the divergent, temperature-independent Lamb shift

$$\Delta_0 = \frac{\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3}{\omega_0-\omega}, \tag{11.98}$$

and the temperature-dependent shift

$$\Delta(T) = \frac{\Gamma}{\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3\bar{n}(\omega)}{\omega_0-\omega}. \tag{11.99}$$

Here, $\bar{n}(\omega)$ is the mean photon number for a mode with frequency ω.

(b) Give an interpretation for all the terms in the master equation (write out equations for the density-matrix elements if you need to).

(c) Write down an expression for $\bar{n}(\omega)$ (nothing fancy, just basic statistical mechanics and Boltzmann statistics).

(d) Argue that $\Delta(T)$ scales as T^4 at low temperatures.

(e) How do the decay terms, Δ_0, and $\Delta(T)$ change if you do *not* make the rotating-wave approximation? Give a qualitative explanation for your results, and indicate whether the low-temperature scaling of $\Delta(T)$ has changed.

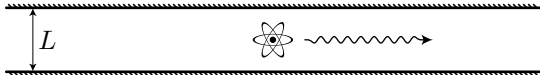
(f) Estimate the temperature-dependent shift for the 780 nm transition in ^{87}Rb (making the crude approximation of treating it as a two-level atom), at room temperature.

Problem 11.5

In the Weisskopf–Wigner derivation of the spontaneous emission rate that we covered in this chapter, we considered the interaction of an atom with the vacuum field, but there was no mention of the Lamb shift, which should also arise from this interaction (as in Problem 11.4). Pinpoint the exact step in the derivation where the Lamb shift disappeared. *Explain.* Also explain how to modify the derivation to obtain an expression for the Lamb shift.

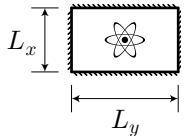
Problem 11.6

(a) A spherically symmetric atom is located halfway between two perfectly reflecting, infinite planar mirrors a distance L apart as shown.



Derive an expression for the decay rate in the limit of small L, such that the cavity acts as a two-dimensional optical waveguide. (Write your answer in terms of the free-space decay rate Γ.) Obviously, any physically correct treatment of this problem must involve three dimensions, so part of the problem is to figure out what ''two-dimensional'' means in this context.

(b) Derive an expression for the decay rate if the atom is at the center of a long, perfectly reflecting, rectangular cavity as shown, in the regime where the cavity acts as a one-dimensional optical waveguide.



Again, part of the problem is to figure out what ''one-dimensional'' means in this context.

Problem 11.7

Derive an expression for the spontaneous-emission rate (Einstein A coefficient) for an atom located a distance a from one of a pair of parallel, infinite, perfectly conducting plates separated by a distance L. Use whatever (quantum-mechanical) formalism you like. Write your result in terms of the free-space rate Γ_0, and plot the resulting rate as a function of L for the case $a = L/2$.

Problem 11.8

(a) Compute the decay rate due to the *magnetic*-dipole transition for the 6.8 GHz ground-state hyperfine ''clock'' transition in ^{87}Rb: $5^2S_1/2, F' = 2 \longrightarrow F = 1$ ($L = L' = 0$, $S = S' = 1/2$, $I = I' = 3/2$). You should proceed by mapping the magnetic-dipole Hamiltonian $H_{\rm AF} = -\boldsymbol{\mu}\cdot\mathbf{B}$ onto the electric-dipole Hamiltonian, then adapt the spontaneous-emission results from this chapter as appropriate to obtain the magnetic-dipole decay rate

$$\Gamma = \frac{\omega_0^3 \mu_{\rm ge}^2}{3\pi\epsilon_0\hbar c^5}. \tag{11.100}$$

For the relevant matrix element of the magnetic-dipole moment, use $\mu_{\rm B}$ as an estimate, or better yet, use the results of Problem 7.6 for a better estimate (noting that $g_S \gg g_I$).

(b) How is the spontaneous-emission rate modified given that the surrounding environment is at room temperature (298 K)? Use the results of Problem 11.4 to help in this calculation.

Problem 11.9

Consider an atom coupled to a single mode of an optical cavity with coupling coefficient g. Assume the intensity-decay of the (empty) cavity is exponential with rate κ. Ignore coupling to other (non-cavity) modes, and do not assume that the atomic and cavity resonances necessarily coincide. Use Fermi's

Golden Rule to derive an expression for the decay rate of the atom in the ''bad-cavity'' limit $\kappa \gg g$. *Hint:* in this limit you can treat the atomic decay directly to continuum, where the cavity has modified the density of vacuum states.

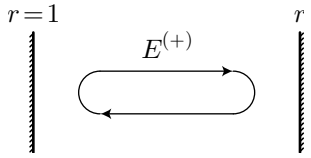
Chapter 12

Coupled-Mode Theory

12.1 Cavity QED

12.1.1 Classical Field in a Single Cavity

To start, let's just consider the simplest case of a classical field in a linear resonator. The resonator consists of two mirrors of field reflection coefficients 1 and r, respectively, surrounding a region of vacuum of length L.



We will assume an approximately monochromatic field of frequency ω. Following the field one round trip around the resonator, we can write

$$E^{(+)}(t+\tau_{\mathrm{rt}})e^{-i\omega(t+\tau_{\mathrm{rt}})} = rE^{(+)}(t)e^{-i\omega t}, \tag{12.1}$$

where $E^{(+)}(t)$ is a slowly varying amplitude for the field, since the fast optical time dependence is written out explicitly, and

$$\tau_{\mathrm{rt}} := \frac{2L}{c} \tag{12.2}$$
(cavity round-trip time)

is the round-trip time of the cavity. Note that we have dropped the polarization of the field, since in our idealized setup it is an invariant. We will assume that the cavity is "good," or not very lossy, so that $|r| \approx 1$. In this case, $E^{(+)}(t+\tau_{\mathrm{rt}})$ is almost the same as $E^{(+)}(t)$, and we can expand $E^{(+)}(t+\tau_{\mathrm{rt}})$ as

$$E^{(+)}(t+\tau_{\mathrm{rt}}) \approx E^{(+)}(t) + \tau_{\mathrm{rt}}\dot{E}^{(+)}(t), \tag{12.3}$$

to first order in τ_{rt}. Putting this into Eq. (12.1), we obtain the rate equation

$$\dot{E}^{(+)}(t) = \frac{re^{i\omega\tau_{\mathrm{rt}}} - 1}{\tau_{\mathrm{rt}}}E^{(+)}(t) \tag{12.4}$$

for the slowly varying field amplitude. For a steady-state solution (i.e., for a stable mode to exist), we must have $\dot{E}^{(+)} = 0$, or

$$re^{i\omega\tau_{\mathrm{rt}}} = 1, \tag{12.5}$$

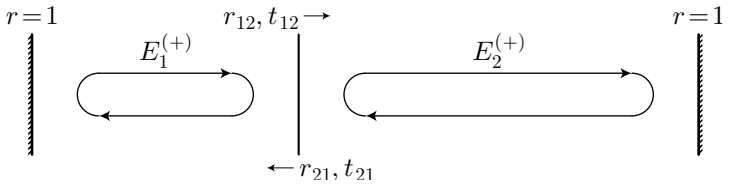
which only happens if $|r| = 1$, and also for

$$\omega\tau_{\mathrm{rt}} + \phi = 2\pi q, \tag{12.6}$$
(resonance condition)

where ϕ is the phase of the reflection, $r = |r|e^{i\phi}$, and q is some integer. It is only for a completely *closed* cavity, $|r| = 1$, where a stable mode exists, and this is precisely the condition that we assumed when quantizing the field. Otherwise, the rate equation (12.4) leads to an exponentially decaying mode (at least within the approximation of a slowly varying amplitude).

12.1.2 Classical Coupled Modes of Two Cavities

Now we can extend this treatment to the case of *two* cavities, which are weakly coupled by a beam splitter where the modulus $|r|$ of the reflection coefficient is close to unity.[1]



We will assume, as for the two-level atom, that only two modes of the individual cavities are approximately "resonant," and so we can treat only these two modes. This, of course, implies that the lengths L_1 and L_2 of the two cavities are appropriately mismatched. The outer mirrors of the double cavity are perfectly reflecting, so that the total system is closed. This will allow us to quantize the coupled fields.

The coupling through the beam splitter is described by field reflection and transmission coefficients r_{12} and t_{12}, as seen by the field $E_1^{(+)}$ in cavity 1, and coefficients r_{21} and t_{21}, as seen by the field $E_2^{(+)}$ in cavity 2. The (lossless) beam splitter induces a unitary transformation on two input fields to generate the output fields. Thus, if the operator for the beam splitter is written

$$U = \begin{bmatrix} t_{21} & r_{12} \\ r_{21} & t_{12} \end{bmatrix}, \tag{12.7}$$

so that the output fields are written in terms of the input fields as

$$\begin{bmatrix} E_{1\leftarrow}^{(+)} \\ E_{2\rightarrow}^{(+)} \end{bmatrix} = U \begin{bmatrix} E_{2\leftarrow}^{(+)} \\ E_{1\rightarrow}^{(+)} \end{bmatrix} \tag{12.8}$$

(with arrows indicating directions of the input and output traveling waves), then it must also have the general unitary form

$$U = \begin{bmatrix} t & r \\ -r^* & t^* \end{bmatrix}, \tag{12.9}$$

with $|r|^2 + |t|^2 = 1$, so that $r = r_{12} = -r_{21}^*$ and $t = t_{12} = t_{21}^*$. (These relations are derivable classically by examining a beam incident on the beam splitter, and the time-reversed process; the resulting relations are then called the **Stokes relations**.) We will assume that $|r| \approx 1$, so that we can treat the coupling as a perturbation on the isolated cavity modes.

We can thus write coupled equations for the fields in basically the same way as for the single cavity:

$$\begin{aligned} E_1^{(+)}(t+\tau_{\mathrm{rt}1})e^{-i\omega_1(t+\tau_{\mathrm{rt}1})} &= r_{12}E_1^{(+)}(t)e^{-i\omega_1 t} + t_{21}E_2^{(+)}(t)e^{-i\omega_2 t} \\ E_2^{(+)}(t+\tau_{\mathrm{rt}2})e^{-i\omega_2(t+\tau_{\mathrm{rt}2})} &= r_{21}E_2^{(+)}(t)e^{-i\omega_2 t} + t_{12}E_1^{(+)}(t)e^{-i\omega_1 t}. \end{aligned} \tag{12.10}$$

Here $\tau_{\mathrm{rt}1}$ and $\tau_{\mathrm{rt}2}$ are the round-trip times $2L_1/c$ and $2L_2/c$ for the two respective cavities, and we have counted the accumulated phase from the fields starting at the beam splitter at $t = 0$. Performing the

[1] For more details on the classical theory of mode coupling in composite resonators, see Robert J. Lang and Amnon Yariv, "Local-field rate equations for coupled optical resonators," *Physical Review A* **34**, 2038 (1986) (doi: 10.1103/PhysRevA.34.2038).

same expansion (assuming weak coupling between the cavities), these coupled equations reduce to the rate equations

$$\frac{d}{dt}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix} = \begin{bmatrix} \dfrac{r_{12}e^{i\omega_1\tau_{\rm rt1}}-1}{\tau_{\rm rt1}} & \dfrac{t_{21}e^{-i\Delta t}}{\tau_{\rm rt1}} \\ \dfrac{t_{12}e^{i\Delta t}}{\tau_{\rm rt2}} & \dfrac{r_{21}e^{i\omega_2\tau_{\rm rt2}}-1}{\tau_{\rm rt2}} \end{bmatrix}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix}, \tag{12.11}$$

where $\Delta := \omega_2 - \omega_1$ is the detuning between the two cavity modes. We can simplify our notation a bit by writing

$$\frac{d}{dt}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix} = -i\begin{bmatrix} \delta\omega_1 & \chi_{12}e^{-i\Delta t} \\ \chi_{21}e^{i\Delta t} & \delta\omega_2 \end{bmatrix}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix}, \tag{12.12}$$

where we have defined the frequency offsets

$$\delta\omega_1 := i\left(\frac{r_{12}e^{i\omega_1\tau_{\rm rt1}}-1}{\tau_{\rm rt1}}\right), \qquad \delta\omega_2 := i\left(\frac{r_{21}e^{i\omega_2\tau_{\rm rt2}}-1}{\tau_{\rm rt2}}\right), \tag{12.13}$$

and the field coupling coefficients

$$\chi_{12} := i\frac{t_{21}}{\tau_{\rm rt1}}, \qquad \chi_{21} := i\frac{t_{12}}{\tau_{\rm rt2}}. \tag{12.14}$$

(classical field-coupling coefficients)

To self-consistently treat the double cavity, we would find the modes by letting $\omega_1, \omega_2 \longrightarrow \omega$, where ω is the eigenfrequency to be found. Then the time derivative vanishes in Eq. (12.12) for eigenmodes of the system, which implies that the determinant of the matrix must also vanish, leading to the condition

$$\left(r_{12}e^{i\omega_1\tau_{\rm rt1}}-1\right)\left(r_{21}e^{i\omega_2\tau_{\rm rt2}}-1\right) = t_{21}t_{12} \tag{12.15}$$

(again, with $\omega_{1,2} \longrightarrow \omega$) that determines the allowed frequencies ω. However, in the perturbative limit that we want to use here, we keep the original frequencies ω_1 and ω_2. We also note that in the perturbative regime, the transmission coefficients t_{12} and t_{21} are $O(\epsilon)$, where ϵ is some small perturbation parameter, and with this definition, the frequencies $\delta\omega_1$ and $\delta\omega_2$ are $O(\epsilon^2)$ if the resonance conditions for the individual cavities are satisfied, since $|r| = \sqrt{1-|t|^2} \approx 1 - |t|^2/2$. Thus, we will simply ignore the diagonal elements of the evolution matrix so that

$$\frac{d}{dt}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix} = \begin{bmatrix} 0 & -i\chi_{12}e^{-i\Delta t} \\ -i\chi_{21}e^{i\Delta t} & 0 \end{bmatrix}\begin{bmatrix} E_1^{(+)} \\ E_2^{(+)} \end{bmatrix}. \tag{12.16}$$

In effect, since $\delta\omega_{1,2}$ have small imaginary parts for any small coupling of the cavities, in neglecting these we are explicitly making the approximation that the isolated-cavity modes are still well defined but coupled together. As a last simplification, we can transform into a rotating frame by defining

$$\tilde{E}_1^{(+)} := E_1^{(+)}e^{i\Delta t}, \qquad \tilde{E}_2^{(+)} := E_2^{(+)}, \tag{12.17}$$

so that

$$\frac{d}{dt}\begin{bmatrix} \tilde{E}_1^{(+)} \\ \tilde{E}_2^{(+)} \end{bmatrix} = \begin{bmatrix} i\Delta & -i\chi_{12} \\ -i\chi_{21} & 0 \end{bmatrix}\begin{bmatrix} \tilde{E}_1^{(+)} \\ \tilde{E}_2^{(+)} \end{bmatrix},$$

(classical coupled-mode equations) (12.18)

and thus we have eliminated the explicit time dependence in the problem. Formally, we see that the dynamics of the two modes are formally equivalent to those of the amplitudes of a two-level atom driven by a classical field (without spontaneous emission), where we identify $2|\chi_{12}|$ with the Rabi frequency Ω, as we see by comparison to Eqs. (5.25). We thus expect Rabi oscillations of the field between the two cavities, characteristic of a pair of coupled harmonic oscillators.

12.1.3 Quantization of the Coupled Modes

We can now write down our quantum description of the coupled cavities by simply identifying the field variables as operators, as we discussed in Chapter 8:

$$E_\alpha^{(+)}(\mathbf{r},t) \longrightarrow -\sqrt{\frac{\hbar\omega_\alpha}{2\epsilon_0}} f_\alpha(\mathbf{r}) a_\alpha(t). \tag{12.19}$$

It is more convenient to write the modes in this case as

$$E_\alpha^{(+)}(\mathbf{r},t) \longrightarrow -\sqrt{\frac{\hbar\omega_\alpha}{2\epsilon_0 V_\alpha}} f'_\alpha(\mathbf{r}) a_\alpha(t), \tag{12.20}$$

where the $f'_\alpha(\mathbf{r})$ are dimensionless mode functions, whose amplitude is of order unity (and of equivalent form for the two cavities), as we discussed in Chapter 10. Then we can interpret the rate equations (12.18) as Heisenberg equations for the field operators $a_1(t)$, $a_2(t)$, where the operators are also understood to be slowly varying (with the optical time dependence factored out). After solving for the operators, we obtain

$$\frac{d}{dt}\begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \end{bmatrix} = \begin{bmatrix} i\Delta & -ig_{12} \\ -ig_{21} & 0 \end{bmatrix} \begin{bmatrix} \tilde{a}_1 \\ \tilde{a}_2 \end{bmatrix},$$

(quantum coupled-mode equations) (12.21)

where the twiddles on the operators remind us that they are in the rotating frame, where the $e^{-i\Delta t}$ time dependence is suppressed, and we have defined the quantum mode-coupling coefficients

$$g_{12} := \chi_{12}\sqrt{\frac{\omega_2 V_1}{\omega_1 V_2}}, \qquad g_{21} := \chi_{21}\sqrt{\frac{\omega_1 V_2}{\omega_2 V_1}}. \tag{12.22}$$

Note that we have dropped the spatial dependence of the modes, as consistent with the classical treatment. We will also assume the two cavities to be mode-matched (as is consistent with a two-mode treatment), which amounts to taking $V_1 = AL_1$ and $V_2 = AL_2$, where the area A is the same for both cavities, and thus $V_1/V_2 = L_1/L_2$. Further, the ratio of frequencies here is an artifact of perturbation theory—had we taken the self-consistent approach described above, we would have had a pair of eigenvalues, giving the frequencies *throughout both cavities* of two different (dressed) modes. Thus, we should set $\omega_1 = \omega_2$ in the coupling coefficients, to obtain

$$g_{12} = \chi_{12}\sqrt{\frac{L_1}{L_2}}, \qquad g_{21} = \chi_{21}\sqrt{\frac{L_2}{L_1}},$$

(quantum mode-coupling coefficients) (12.23)

where, in view of the definitions (12.14), we can write

$$g_{12} = \frac{ict_{21}}{2\sqrt{L_1 L_2}}, \qquad g_{21} = \frac{ict_{12}}{2\sqrt{L_1 L_2}}.$$

(quantum mode-coupling coefficients) (12.24)

The Heisenberg equations here are precisely those that arise from the general Heisenberg equation

$$\dot{\tilde{a}} = -\frac{i}{\hbar}[\tilde{a}, \tilde{H}], \tag{12.25}$$

if we take the Hamiltonian H to be the sum of the free Hamiltonian in the rotating frame,

$$\tilde{H}_0 = -\hbar\Delta\tilde{a}_1^\dagger\tilde{a}_1,$$

(free-field Hamiltonian, rotating frame) (12.26)

and the interaction Hamiltonian,

$$\tilde{H}_{12} = \hbar\left(g_{21}\tilde{a}_1\tilde{a}_2^\dagger + g_{12}\tilde{a}_2\tilde{a}_1^\dagger\right), \tag{12.27}$$

also in the rotating frame. Note from Eqs. (12.24) that with our definition of the coefficients, this Hamiltonian is in fact not Hermitian. This comes from the usual conventions for the reflection and transmission coefficients

(12.7) and (12.9); we really should instead take the convention $t_{21} = -t_{12}^*$ to obtain a Hermitian interaction Hamiltonian. This interaction Hamiltonian is our main result. In particular, note that by redefining the relative phase of the two modes (with the above phase convention in mind), we may always assume a real coupling coefficient, $g_{21} = g_{12} \equiv g \in \mathbb{R}$,

$$\tilde{H}_{12} = \hbar g \left(\tilde{a}_1 \tilde{a}_2^\dagger + \tilde{a}_2 \tilde{a}_1^\dagger \right).$$

(mode-coupling Hamiltonian, rotating frame) (12.28)

Note that had we not taken $\omega_1 = \omega_2$ to obtain Eqs. (12.23), we would not have arrived at a Hermitian interaction Hamiltonian. Notice that the interaction here has exactly the same form as for the atom–field interaction in the Jaynes–Cummings model, as in Eq. (10.9), where we can identify a_1 with the atomic dipole operator σ and a_2 with the quantized field mode a. Note that we have implicitly made a rotating-wave approximation in the classical setup, since we ignored an coupling from $E^{(-)}$ in one cavity to $E^{(+)}$ in the other, since these couplings are far off resonance (as is consistent with the single-mode approximation). Of course, the coupling coefficient g is not related to the one-photon Rabi frequency from the Jaynes–Cummings model, except in the way it appears in the interaction Hamiltonian.

Notice that if we transform out of the rotating frame, the free Hamiltonian is trivial,

$$H_0 = 0,$$

(free-field Hamiltonian, nonrotating frame) (12.29)

while explicit time dependence returns to the interaction Hamiltonian:

$$H_{12} = \hbar g \left(a_1 a_2^\dagger e^{i\Delta t} + a_2 a_1^\dagger e^{-i\Delta t} \right).$$

(mode-coupling Hamiltonian, nonrotating frame) (12.30)

This means that we are now in the interaction picture with respect to the free evolution of the fields (which was equivalent to factoring out the free time dependence of the optical fields). Transforming into the Schrödinger picture yields the alternate free Hamiltonian

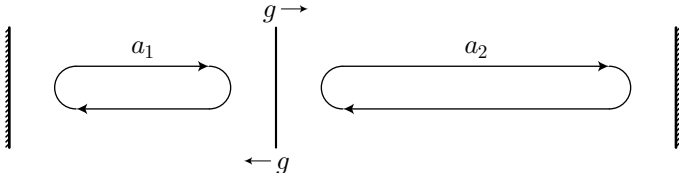
$$H_0 = \hbar\omega_1 \left(a_1^\dagger a_1 + \frac{1}{2} \right) + \hbar\omega_2 \left(a_2^\dagger a_2 + \frac{1}{2} \right), \tag{12.31}$$

as well as an interaction Hamiltonian that is equivalent to the rotating-frame Hamiltonian:

$$H_{12} = \hbar g \left(a_1 a_2^\dagger + a_2 a_1^\dagger \right). \tag{12.32}$$

In fact the rotating frame and the Schrödinger picture are equivalent here except for offsets of the bare energy levels.

We now have essentially the same situation as in the classical case, but now the fields are represented by operators.

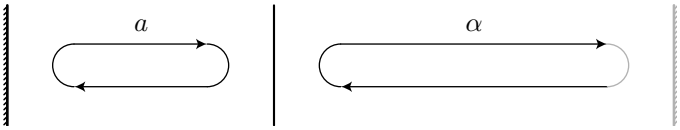


We also have a symmetric coupling between the two fields, representing a Hermitian interaction Hamiltonian.

12.1.4 Cavity Driven by a Classical Field

As an application of the above formalism, we will derive a model for driving a (slightly) lossy cavity with an external, classical field. In this case, cavity 1 will be our cavity, so we will change notations by $a_1 \longrightarrow a$, and we will regard it as being driven by cavity 2. Cavity 2 has a classical field, and with many photons around we may neglect the fluctuations in photon number compared to the mean photon number, making

the replacement $a_2 \longrightarrow \alpha$, assuming cavity 2 is in the coherent state α. While technically the classical field is circulating in the right-hand cavity, this is not necessary—since we will eliminate all the parameters of the classical cavity except the circulating power, we can equally well think of the classical field as an incident traveling wave.



We should then go into the interaction representation with respect to the free dynamics of cavity 2 to obtain the free Hamiltonian for cavity 1,

$$H_0 = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right), \tag{12.33}$$

where we have changed notations $\omega_1 \longrightarrow \omega$, and the interaction Hamiltonian becomes

$$H_{12} = \hbar\mathscr{E}\left(ae^{i\omega_L t} + a^\dagger e^{-i\omega_L t}\right). \tag{12.34}$$

(classical driving Hamiltonian for cavity field)

Here, we have defined the driving amplitude

$$\mathscr{E} := \alpha g, \tag{12.35}$$

(classical-field driving amplitude)

and we have changed notations $\omega_2 \longrightarrow \omega_L$, where ω_L is the "external drive frequency" for cavity 1, and again by appropriate definition of relative phase we may assume α to be a real number (hence $\mathscr{E}$ is also real).

Notice that our setup here is formally equivalent to the two-level atom interacting with a classical field, as we might guess from our remarks above. This correspondence carries through if we identify a with the atomic operator σ, ω with the atomic resonance frequency, ω_L with the frequency of the classical field, and $2\mathscr{E}$ with the Rabi frequency. Also, the interaction Hamiltonian is of the form $x\cos\omega_L t$, as we expect for a forced harmonic oscillator (here, forced by in the incident field).

12.1.4.1 Cavity Decay Rate

Now we can write the driving rate $\mathscr{E}$ in a more useful form. Using $g = |g_{12}|$ and Eqs. (12.24), we find

$$\mathscr{E} = \alpha g = \alpha\frac{c|t_{21}|}{2\sqrt{L_1 L_2}}. \tag{12.36}$$

Now we can define the power P of the "external" field in cavity 2, which is the energy per unit time, or the product of the photon number and photon energy divided by the round-trip time:

$$P = \frac{|\alpha|^2\hbar\omega_2}{\tau_{rt2}}. \tag{12.37}$$

Then we can eliminate α (dropping its phase to focus on the magnitude) to obtain

$$\begin{aligned}\mathscr{E} &= \sqrt{\frac{P\tau_{rt2}}{\hbar\omega_2}}\frac{c|t_{21}|}{2\sqrt{L_1 L_2}} \\ &= \sqrt{\frac{P}{\hbar\omega_2}}\frac{\sqrt{c}|t_{21}|}{\sqrt{2L_1}} \\ &= \sqrt{\frac{P}{\hbar\omega_2}}\frac{|t_{21}|}{\sqrt{\tau_{rt1}}},\end{aligned} \tag{12.38}$$

where we used $\tau_{\rm rt1} = 2L_1/c$ and $\tau_{\rm rt2} = 2L_2/c$. We can then define the **cavity decay rate**[2]

$$\kappa := \frac{|t_{21}|^2}{\tau_{\rm rt1}}, \tag{12.39}$$
(cavity decay rate)

which defines the rate at which energy escapes cavity 1 (in the absence of cavity 2), being the intensity transmission coefficient divided by the round-trip time (to give the rate of energy transport through the output mirror). We also take $\omega_1 = \omega_2 = \omega$ for consistency. Then the driving rate simply becomes

$$\mathscr{E} = \sqrt{\frac{\kappa P}{\hbar\omega}}, \tag{12.40}$$
(classical-field driving amplitude)

where now the frequency ω refers to the driven cavity (cavity 1). Of course, this coupling rate can be complex, with the phase of the input field α. In this form, the coupling rate is independent of the details of the auxiliary cavity (cavity 2), and is thus a general result, assuming the input power is mode-matched with the output port of the cavity. Often, in a real Fabry–Perot cavity, there are multiple loss channels that contribute to κ, in which case an alternate value κ' should be used in the above formula, which is the decay rate that the cavity *would have*, assuming that the input port of the driving field gives the only contribution to κ.

It is common to define the **finesse** of a cavity by

$$\mathscr{F} = \frac{\pi P_{\rm s}^{1/4}}{1 - \sqrt{P_{\rm s}}}, \tag{12.41}$$
(cavity finesse)

where $P_{\rm s}$ is the **survival probability** for a photon after one round trip through the cavity ($P_{\rm s} = |r_{12}|^2$ for the cavity we have set up here). The finesse measures how ''good'' a cavity is, with a large finesse indicating low loss (and well-resolved resonance lines). In the good-cavity limit, we can then write

$$\mathscr{F} \approx \frac{2\pi}{|t_{21}|^2}, \tag{12.42}$$

so that the decay rate becomes

$$\kappa = \frac{2\pi}{\tau_{\rm rt}\mathscr{F}} = 2\pi\frac{\rm FSR}{\mathscr{F}}. \tag{12.43}$$
(decay rate in terms of finesse)

In the last expression, we defined the **free spectral range** $\text{FSR} := 1/\tau_{\rm rt}$, which is the frequency spacing between adjacent modes in the Fabry–Perot resonator. Note that the result here is valid for the *asymmetric* cavity that we started with, where the output coupler is the only source of loss (for example, the factor of 2 disappears for a *symmetric* cavity).

12.1.5 Cavity Decay

The other application of the coupled-mode formalism that we will discuss is the quantum theory of the cavity decay. Since the interaction Hamiltonian (12.32) has the same form as the atom–field interaction in the rotating-wave approximation, we will make heavy use of the Weisskopf–Wigner treatment of atomic spontaneous decay in Chapter 11, and our calculation here will essentially just be mapping the cavity problem onto the atomic problem. First, we will again consider cavity 1 to be the cavity we are modeling, and cavity 2 will contain the ''output field.'' However, we will take the limit as cavity 2 becomes large, to get the free-space limit of the output field. The single-mode approximation for cavity 2 will break down, and we must consider coupling to many modes of cavity 2. Cavity 2 will be initially in the ground state, and always

[2]Here we are defining κ to be the *energy* decay rate of the cavity, in analogy with the population decay of the two-level atom. It is also common to define the decay rate of the *field* using the same symbol, which would differ by a factor of 2 from the definition here.

"approximately" in the ground state so that the radiated energy never re-enters the cavity. The interaction is given by

$$H_{\rm int} = \hbar \sum_q g_q \left(a a_q^\dagger + a^\dagger a_q \right), \tag{12.44}$$

where a is the annihilation operator for cavity 1, and the a_q are the annihilation operators for the output field modes. Thus, as in atomic spontaneous emission, the cavity will by damped by coupling to a "bath" of harmonic oscillators. We can write the coupling coefficient, combining Eqs. (12.39) and (12.24), as

$$g_q = \frac{c|t_{21}|}{2\sqrt{L_1 L_2}} = \frac{|t_{21}|}{\sqrt{\tau_{\rm rt1}\tau_{\rm rt2}}} = \sqrt{\frac{\kappa}{\tau_{\rm rt2}}}, \tag{12.45}$$

where ω is the frequency of the decaying mode of cavity 1, ω_q is the frequency of the qth output mode, and κ is defined as above (note that κ may depend on q via the frequency dependence of the mirror reflectance). It suffices to consider the initial condition of a Fock state $|\psi(0)\rangle = |n\rangle$ for cavity 1, which for short times couples only to states with one less photon in the cavity. Thus it also suffices to consider the state restricted to the manifold given by the superposition

$$|\psi\rangle = c_n |n\rangle + \sum_q c_q |n-1, 1_q\rangle, \tag{12.46}$$

where again the 1_q denotes a single photon present in the qth output mode. This analysis is valid for any other initial state, since any Fock state in the superposition is coupled to a separate set of states from the other Fock states. Writing down the equations of motion for the coefficients, transforming into the rotating frame, and decoupling the equations as in the atomic case leads to

$$\begin{aligned} \dot{\tilde{c}}_n(t) &= -\sum_q n|g_q|^2 \int_0^t dt'\, \tilde{c}_n(t') e^{-i(\omega_q-\omega)(t-t')} \\ &= -\frac{n\kappa}{\tau_{\rm rt2}} \sum_q \int_0^t dt'\, \tilde{c}_n(t') e^{-i(\omega_q-\omega)(t-t')}. \end{aligned} \tag{12.47}$$

The spacing of the output modes is the free spectral range $\Delta\omega_q = 2\pi/\tau_{\rm rt2}$ of the output cavity, and thus in the limit where cavity 2 becomes large, we change the sum to an integral according to

$$\sum_q f(\omega_q)\Delta\omega_q \longrightarrow \int_0^\infty d\omega_{\rm out} f(\omega_{\rm out}), \tag{12.48}$$

where $\omega_{\rm out}$ refers to the output frequency, so that

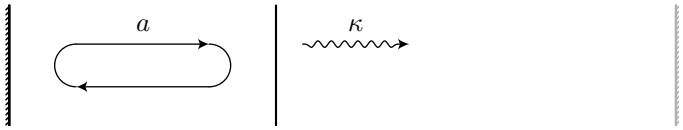
$$\dot{\tilde{c}}_n(t) = -\frac{n\kappa}{2\pi} \int_0^\infty d\omega_{\rm out} \int_0^t dt'\, \tilde{c}_n(t') e^{-i(\omega_{\rm out}-\omega)(t-t')}. \tag{12.49}$$

Carrying out the integrals as in the atomic case amounts to setting $c_n(t') \longrightarrow c_n(t)$, and then introducing an extra factor of π:

$$\dot{\tilde{c}}_n = -\frac{n\kappa}{2}\tilde{c}_n. \tag{12.50}$$

(cavity-decay dynamics)

This is consistent with $|c_n|^2$ decaying at the rate $n\kappa$, as we might expect from the *classical* definition (12.39) of κ.



Thus, we have arrived at the situation where we have eliminated the right-hand cavity by pushing the end mirror away to infinity, and obtained irreversible decay from the remaining cavity into the continuum at rate κ.

12.1.5.1 Master Equation

Still proceeding along the lines of the atomic case, from the above decay equation, the equation of motion for the nth state population is

$$\dot{\rho}_{nn} = -n\kappa\rho_{nn}. \tag{12.51}$$

When tracing over the output field states, the only state coupled to $|n\rangle$ is $|n-1\rangle$, and thus, the $n-1$ level must take up the decayed population

$$\dot{\rho}_{(n-1)(n-1)} = n\kappa\rho_{nn}. \tag{12.52}$$

Finally, transforming out of the rotating frame, using the appropriate combination of coefficients, and tracing out the field gives the equation of motion for the coherence

$$\dot{\rho}_{(n)(n-1)} = \left(i\omega - \frac{n\kappa}{2}\right)\rho_{(n)(n-1)}. \tag{12.53}$$

Our setup was for short times, but in the Markovian approximation (where the emitted field does not act back on the cavity), we simply evolve for a short time and repeat the argument, so that these equations of motion are always valid. Of course, then we must consider all couplings of the above forms between the density matrix elements. Then the equations of motion are precisely those generated by the Lindblad-form master equation

$$\dot{\rho} = -\frac{i}{\hbar}[H_0, \rho] + \kappa\mathcal{D}[a]\rho, \tag{12.54}$$

(cavity-decay master equation)

where again, ignoring the zero-point energy offset,

$$H_0 = \hbar\omega a^\dagger a, \tag{12.55}$$

and the Lindblad superoperator is

$$D[a]\rho := a\rho a^\dagger - \frac{1}{2}\left(a^\dagger a\rho + \rho a^\dagger a\right). \tag{12.56}$$

Recall that the operator form of the cavity-decay master equation has the same form as for atomic spontaneous emission (i.e., the optical Bloch equations), under the identifications $\sigma \longrightarrow a$, $\omega_0 \longrightarrow \omega$, and $\Gamma \longrightarrow \kappa$.

12.2 Input–Output Formalism

The **input–output formalism**[3] is an extension of the above formalism for treating the evolution of systems in the Heisenberg picture, particularly when coupled to a continuum. The difference is that we will now keep explicit track of the inputs and outputs (in the case of a cavity, the input and output *fields*) via Heisenberg-picture operators.

To set this up for a cavity of resonance frequency ω, we begin with the interaction (12.44)

$$H_{\text{int}} = \hbar\sum_q g_q\left(aa_q^\dagger + a^\dagger a_q\right), \tag{12.57}$$

where the coupling to the qth mode outside the cavity is

$$g_q = \sqrt{\frac{\kappa}{\tau_{\text{rt2}}}}, \tag{12.58}$$

[3] M. J. Collett and C. W. Gardiner, ''Squeezing of intracavity and traveling-wave light fields produced in parametric amplification,'' *Physical Review A* **30**, 1386 (1984) (doi: 10.1103/PhysRevA.30.1386); C. W. Gardiner and M. J. Collett, ''Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation,'' *Physical Review A* **31**, 3761 (1985) (doi: 10.1103/PhysRevA.31.3761); C. W. Gardiner and P. Zoller, *Quantum Noise*, second enlarged edition (Springer, 2000); M. Ley and R. Loudon, ''Quantum theory of high-resolution length measurement with a Fabry–Perot interferometer,'' *Journal of Modern Optics* **34**, 227 (1987); R. Graham, ''Quantum Langevin equation and input-output fields for arbitrary linear response,'' *Zeitschrift für Physik B* **76**, 265 (1989) (doi: 10.1007/BF01312694); D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer, 1994), Chapter 7.

where again κ is the cavity decay rate (which could depend on q via the frequency dependence of the output coupler), and $\tau_{\rm rt2}$ is the round-trip time of the ''external cavity,'' in which the exterior field is quantized. Before, we passed over to the continuum limit after setting up a calculation, but now it will be convenient to do so right away. The idea was that we made the replacement

$$\sum_q |g_q|^2 \longrightarrow \frac{\tau_{\rm rt2}}{2\pi}\int_0^\infty d\omega' \frac{\kappa}{\tau_{\rm rt2}} = \int_0^\infty d\omega' \,\frac{\kappa}{2\pi}. \tag{12.59}$$

This is equivalent to making the replacement $g_q \longrightarrow \sqrt{\kappa/2\pi}$ when passing the sum over to an integral over the external-mode frequencies ω', $\sum_q \longrightarrow \int_0^\infty d\omega'$. Making these replacements in the interaction Hamiltonian,

$$H_{\rm int} = \frac{\hbar}{\sqrt{2\pi}}\int_0^\infty d\omega' \sqrt{\kappa(\omega')}\left[ab^\dagger(\omega') + a^\dagger b(\omega')\right], \tag{12.60}$$

(cavity coupling to external modes)

where in the continuous limit, we have changed notation $a_q \longrightarrow b(\omega')$ (b for the external ''bath'' modes), and we are explicitly indicating any frequency dependence of the decay rate, due to frequency dependence of the transmission coefficient (which must ultimately converge to unity as $\omega' \longrightarrow \infty$). Again, this interaction assumes the rotating-wave approximation in omitting terms like $ab(\omega')$ and $a^\dagger b^\dagger(\omega')$. In what follows, it is convenient to extend the lower limit of integration to $-\infty$:

$$H_{\rm int} \approx \frac{\hbar}{\sqrt{2\pi}}\int_{-\infty}^\infty d\omega' \sqrt{\kappa(\omega')}\left[ab^\dagger(\omega') + a^\dagger b(\omega')\right]. \tag{12.61}$$

(cavity coupling to external modes)

This is justified since only bath frequencies ω' near the cavity resonance ω should be important, and ω is much larger than the rates associated with the decay interaction. The bath modes themselves satisfy the commutation relations

$$\left[b(\omega'), b^\dagger(\omega'')\right] = \delta(\omega' - \omega''), \tag{12.62}$$

(bath-mode commutation relation)

which is the continuum version of $[a_q, a_{q'}^\dagger] = \delta_{qq'}$. Also, we have the free-evolution Hamiltonian for the cavity, or the ''system,''

$$H_{\rm sys} = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right) \tag{12.63}$$

(free cavity Hamiltonian)

and the Hamiltonian for the external bath modes

$$H_{\rm ext} = \hbar\int_0^\infty d\omega'\,\omega' b^\dagger(\omega')b(\omega') \approx \hbar\int_{-\infty}^\infty d\omega'\,\omega' b^\dagger(\omega')b(\omega'), \tag{12.64}$$

(free external bath Hamiltonian)

if we drop the zero-point contributions and extend the integral again over negative frequencies.

12.2.1 Quantum Langevin Equation

Now for an arbitrary system operator c, the Heisenberg equation of motion is

$$\partial_t c = -\frac{i}{\hbar}[c, H]. \tag{12.65}$$

In particular, for the cavity annihilation operator a, we have

$$\begin{aligned}\partial_t a &= -\frac{i}{\hbar}[a, H_{\rm sys}] - \frac{i}{\hbar}[a, H_{\rm int}] \\ &= -i\omega a(t) - \frac{i}{\sqrt{2\pi}}\int_{-\infty}^\infty d\omega' \sqrt{\kappa(\omega')}\, b(\omega'),\end{aligned} \tag{12.66}$$

where we have used $[a, a^\dagger] = 1$ and $[a, a^\dagger a] = a$. Similarly for $b(\omega')$,

$$\begin{aligned}\partial_t b(\omega') &= -\frac{i}{\hbar}[b, H_{\rm ext}] - \frac{i}{\hbar}[b, H_{\rm int}] \\ &= -i\omega' b(\omega') - i\sqrt{\frac{\kappa(\omega')}{2\pi}}\, a(t).\end{aligned} \tag{12.67}$$

To solve this latter equation, we can transform to a rotating frame,

$$\begin{aligned}\partial_t \left[b(\omega')e^{i\omega' t}\right] &= \left[\partial_t b(\omega')\right] e^{i\omega' t} + i\omega' b(\omega')\, e^{i\omega' t} \\ &= -i\sqrt{\frac{\kappa(\omega')}{2\pi}}\, a(t)\, e^{i\omega' t}.\end{aligned} \tag{12.68}$$

Integrating from some *past* time t_0 to t, we obtain

$$b(\omega')e^{i\omega' t} - b_0(\omega')e^{i\omega' t_0} = -i\sqrt{\frac{\kappa(\omega')}{2\pi}} \int_{t_0}^{t} dt'\, a(t')\, e^{i\omega' t'}, \tag{12.69}$$

where $b_0(\omega') := b(\omega')|_{t=t_0}$. We can then rewrite this equation as

$$b(\omega') = b_0(\omega')\, e^{-i\omega'(t-t_0)} - i\sqrt{\frac{\kappa(\omega')}{2\pi}} \int_{t_0}^{t} dt'\, a(t')\, e^{-i\omega'(t-t')}. \tag{12.70}$$

Putting this into Eq. (12.66), we find

$$\partial_t a = -i\omega a(t) - \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega' \sqrt{\kappa(\omega')} b_0(\omega')\, e^{-i\omega'(t-t_0)} - \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega'\, \kappa(\omega') \int_{t_0}^{t} dt'\, a(t')\, e^{-i\omega'(t-t')}. \tag{12.71}$$

To proceed, we now make the Markov approximation by ignoring the frequency dependence of the decay rate,

$$\kappa(\omega') \approx \kappa = (\text{constant}). \tag{12.72}$$

Strictly speaking, this cannot be true, but can be a good approximation over the frequency range of interest—the resonance linewidth in the case of the optical cavity. Then we can name the integral in the second term of Eq. (12.71) such that it becomes $\sqrt{\kappa}\, a_{\rm in}(t)$, where

$$a_{\rm in}(t) := \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega'\, b_0(\omega')\, e^{-i\omega'(t-t_0)} \tag{12.73}$$

(input field operator)

is the **input field operator**, which we will interpret in just a bit. The last term of Eq. (12.71) then becomes

$$-\frac{\kappa}{2\pi} \int_{t_0}^{t} dt'\, a(t') \int_{-\infty}^{\infty} d\omega'\, e^{-i\omega'(t-t')} = -\frac{\kappa}{2} a(t), \tag{12.74}$$

where we have used

$$\int_{-\infty}^{\infty} d\omega'\, e^{-i\omega'(t-t')} = 2\pi\delta(t-t') \tag{12.75}$$

and

$$\int_{t_0}^{t} dt'\, a(t')\, \delta(t-t') = \frac{a(t)}{2}, \tag{12.76}$$

since the delta function is "split," with half the contribution of the exponential factor in Eq. (12.75) being picked up here. Putting these pieces together, we have the Heisenberg equation

$$\partial_t a = -i\omega a(t) - \frac{\kappa}{2} a(t) - \sqrt{\kappa}\, a_{\rm in}(t), \tag{12.77}$$

(quantum Langevin equation)

called the **quantum Langevin equation**, since as we will see, the second term represents damping, and the last term represents quantum noise. This last term also represents an *input* to the system, since it represents the influence of the external modes $b(\omega')$ at time t_0 in the past on the present system operator $a(t)$.

12.2.1.1 Evolution of the Mean

To get a bit more insight into the Langevin equation (12.77), recall the Schrödinger-picture description of cavity damping represented by the master equation (12.54). The master equation implies the equation of motion

$$\partial_t\langle A\rangle = \mathrm{Tr}[A\partial_t\rho] = -\frac{i}{\hbar}\langle[A, H_{\mathrm{sys}}]\rangle + \kappa\left\langle a^\dagger A a - \frac{1}{2}\left(a^\dagger a A + A a^\dagger a\right)\right\rangle \tag{12.78}$$

for an arbitrary system operator A. Setting $A = a$,

$$\partial_t\langle a\rangle = -i\omega\langle a\rangle + \frac{\kappa}{2}\left\langle a^\dagger a a - a a^\dagger a\right\rangle = -i\omega\langle a\rangle - \frac{\kappa}{2}\langle a\rangle\,. \tag{12.79}$$

This is equivalent to the expectation value of the Langevin equation (12.77), but evidently $\langle a_{\mathrm{in}}(t)\rangle = 0$ for this case. This is consistent with our expectation, since the reservoir that leads to this master equation is in the vacuum state, and the reservoir expectation value for each mode vanishes: $\langle 0|b_0(\omega')|0\rangle = 0$. This means that the expectation value damps exponentially away:

$$\langle a(t)\rangle = \langle a(0)\rangle\, e^{-i\omega t}e^{-\kappa t/2}. \tag{12.80}$$

Obviously, though, the operator a must have more to it than just the expectation value—in particular, it has *fluctuations* about the mean. Otherwise, for example, the commutator $[a(t), a^\dagger(t)]$ would decay to zero, but it must be unity for all times. The input a_{in} thus acts as a quantum noise that represents the fluctuations of a—fluctuations that are required in the presence of damping to ensure that commutators are preserved. This is one manifestation of the **fluctuation–dissipation relation** (Section 14.3.8.1).

This interpretation of the input operators as noise terms is reinforced by computing the commutator of $a_{\mathrm{in}}(t)$,

$$\begin{aligned}\left[a_{\mathrm{in}}(t), a_{\mathrm{in}}^\dagger(t')\right] &= \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega'\int_{-\infty}^{\infty}d\omega''\,\left[b_0(\omega'), b_0^\dagger(\omega'')\right]\,e^{-i\omega'(t-t_0)}e^{i\omega''(t'-t_0)}\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega'\int_{-\infty}^{\infty}d\omega''\,\delta(\omega'-\omega'')\,e^{-i\omega'(t-t_0)}e^{i\omega''(t'-t_0)}\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega'\,e^{-i\omega'(t-t')},\end{aligned} \tag{12.81}$$

so that

$$\left[a_{\mathrm{in}}(t), a_{\mathrm{in}}^\dagger(t')\right] = \delta(t-t'). \tag{12.82}$$
(input-operator commutator)

The input operator thus appears to have the character of white noise, since its correlation function is a delta function (i.e., the power spectrum is flat). We will formalize this notion better after defining some fundamental concepts in stochastic calculus.

12.2.2 Output Field

We can proceed again as before, but instead of integrating Eq. (12.68) from a past time t_0 to t, we can integrate from t to a *future* time t_1, to obtain

$$b(\omega') = b_1(\omega')\,e^{-i\omega'(t-t_1)} + i\sqrt{\frac{\kappa(\omega')}{2\pi}}\int_t^{t_1}dt'\,a(t')\,e^{-i\omega'(t-t')}, \tag{12.83}$$

where $b_1(\omega') := b(\omega')|_{t=t_1}$. Again putting this into Eq. (12.66), we find

$$\partial_t a = -i\omega a(t) - \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}d\omega'\sqrt{\kappa(\omega')}b_1(\omega')\,e^{-i\omega'(t-t_1)} + \frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega'\,\kappa(\omega')\int_t^{t_1}dt'\,a(t')\,e^{-i\omega'(t-t')}. \tag{12.84}$$

Making the Markov approximation, defining the *output* field

$$a_{\mathrm{out}}(t) := \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}d\omega'\,b_1(\omega')\,e^{-i\omega'(t-t_1)}, \tag{12.85}$$
(output field operator)

and carrying out the integrals as before, we find the alternate, *time-reversed* Langevin equation

$$\partial_t a = -i\omega a(t) + \frac{\kappa}{2} a(t) - \sqrt{\kappa}\, a_{\text{out}}(t). \tag{12.86}$$
(time-reversed Langevin equation)

Here, the output operator $a_{\text{out}}(t)$ represents the coupling of the system to *future* bath modes, and thus we interpret this to be the system output. However, since the influence of $a_{\text{out}}(t)$ is in the future, this equation represents the *backwards* evolution of the system—hence the negative damping term.

Using essentially the same calculation leading up to Eq. (12.82), we find that the output-field commutator

$$\left[a_{\text{out}}(t), a_{\text{out}}^\dagger(t')\right] = \delta(t-t'). \tag{12.87}$$
(output-operator commutator)

is also a temporal delta function. Not surprisingly, the output operator $a_{\text{out}}(t)$ has the same spectral properties as the input $a_{\text{in}}(t)$.

12.2.3 Input–Output Relation

To relate the input and output fields, we start by integrating Eq. (12.70) over all frequencies (in the Markov approximation),

$$\begin{aligned}\int_{-\infty}^{\infty} d\omega'\, b(\omega') &= \int_{-\infty}^{\infty} d\omega'\, b_0(\omega')\, e^{-i\omega'(t-t_0)} - i\sqrt{\frac{\kappa}{2\pi}} \int_{t_0}^{t} dt'\, a(t') \int_{-\infty}^{\infty} d\omega'\, e^{-i\omega'(t-t')} \\ &= -\sqrt{2\pi} i a_{\text{in}}(t) - i\sqrt{\frac{\pi\kappa}{2}}\, a(t),\end{aligned} \tag{12.88}$$

which we can rewrite as

$$\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega'\, b(\omega') = a_{\text{in}}(t) + \frac{\sqrt{\kappa}}{2}\, a(t). \tag{12.89}$$

In particular, this means that the combination of operators on the right-hand side commutes with *any* system operator $c(t)$, since it is independent of the state of the bath at the same time. Similarly, we can integrate Eq. (12.83) over all frequencies,

$$\begin{aligned}\int_{-\infty}^{\infty} d\omega'\, b(\omega') &= \int_{-\infty}^{\infty} d\omega'\, b_1(\omega')\, e^{-i\omega'(t-t_1)} + i\sqrt{\frac{\kappa}{2\pi}} \int_{t}^{t_1} dt'\, a(t') \int_{-\infty}^{\infty} d\omega'\, e^{-i\omega'(t-t')} \\ &= -\sqrt{2\pi} i a_{\text{out}}(t) + i\sqrt{\frac{\pi\kappa}{2}}\, a(t),\end{aligned} \tag{12.90}$$

which we can rewrite as

$$\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega'\, b(\omega') = a_{\text{out}}(t) - \frac{\sqrt{\kappa}}{2}\, a(t). \tag{12.91}$$

Again, the combination of operators on the right-hand side commutes with any system operator $c(t)$. Comparing Eq. (12.91) with Eq. (12.89), we find the important relation

$$a_{\text{out}}(t) - a_{\text{in}}(t) = \sqrt{\kappa}\, a(t) \tag{12.92}$$
(input–output relation)

for the input, output, and system fields.

12.2.4 General Heisenberg Equations

Above, we derived the Heisenberg–Langevin equations for the cavity annihilation operator $a(t)$. However, it is also useful to derive Langevin equations for an *arbitrary* system operator $c(t)$. We leave the derivation as an exercise; the results are

$$\partial_t c = -\frac{i}{\hbar}\left[c, H_{\text{sys}}\right] - \left\{\left[c, a^\dagger\right]\left(\frac{\kappa}{2} a + \sqrt{\kappa}\, a_{\text{in}}(t)\right) - \left(\frac{\kappa}{2} a^\dagger + \sqrt{\kappa}\, a_{\text{in}}^\dagger(t)\right)[c, a]\right\}$$
(quantum Langevin equation) (12.93)

and

$$\partial_t c = -\frac{i}{\hbar}[c, H_{\rm sys}] - \left\{ [c, a^\dagger] \left(-\frac{\kappa}{2} a + \sqrt{\kappa}\, a_{\rm out}(t) \right) - \left(-\frac{\kappa}{2} a^\dagger + \sqrt{\kappa}\, a_{\rm out}^\dagger(t) \right) [c, a] \right\}. \quad \text{(time-reversed Langevin equation)} \tag{12.94}$$

These obviously reduce to the previous Langevin equations if $c \longrightarrow a$.

12.2.5 Causality

Suppose again that $c(t)$ is some Heisenberg-picture system operator. Since we integrate Eq. (12.93) forward in time to find the evolution of $c(t)$ in response to $a_{\rm in}(t)$, we can see that $c(t)$ only depends on $a_{\rm in}(t')$ in the past ($t' < t$). Expressed as a commutator, this statement is

$$[c(t), a_{\rm in}(t')] = 0 \quad (t' > t). \tag{12.95}$$

Similarly, integrating Eq. (12.94) gives the influence of $a_{\rm out}(t)$ on $c(t)$ in the *past*, so

$$[c(t), a_{\rm out}(t')] = 0 \quad (t' < t). \tag{12.96}$$

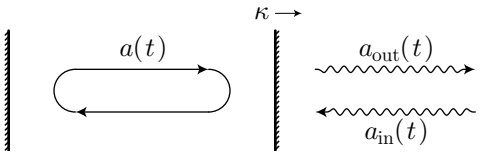
But returning to Eq. (12.92), we see that both $a_{\rm in}(t)$ and $a_{\rm out}(t)$ give rise to the same commutator (up to a minus sign), but on different sections of the time axis. We can thus combine Eqs. (12.92), (12.95), and (12.96) to obtain the general commutators

$$\begin{aligned} [c(t), a_{\rm in}(t')] &= -\theta(t - t')\sqrt{\kappa}\,[c(t), a(t')] \\ [c(t), a_{\rm out}(t')] &= \quad \theta(t' - t)\sqrt{\kappa}\,[c(t), a(t')]\,. \end{aligned} \quad \text{(input–output commutators)} \tag{12.97}$$

These commutators are a general statement of causality with respect to the system and input/output operators. The general strategy for the input–output formalism, then, is to specify the input field $a_{\rm in}(t)$ for a system, use the Langevin equation (12.93) to determine the influence of the input on a system operator $c(t)$ [in particular, $a(t)$], and then use the input–output relation (12.92) to determine the system output $a_{\rm out}(t)$.

12.2.6 Example: Reflections from a Cavity

Let's consider the same one-sided cavity as usual, now in the input-output formalism.[4]



Further, let's resolve the cavity-field operator $a(t)$ into frequency components $a(\omega')$ via the usual Fourier transform:

$$a(\omega') = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt\, a(t)\, e^{i\omega'(t - t_0)}. \tag{12.98}$$

We can resolve the input and output fields in the same way; for example, examination of Eqs. (12.73) and (12.85) shows that $a_{\rm in}(\omega') = ib_0(\omega')$ and $a_{\rm out}(\omega') = ib_1(\omega')$. Then the Langevin equation (12.77)

$$\partial_t a = -i\omega a(t) - \frac{\kappa}{2} a(t) - \sqrt{\kappa}\, a_{\rm in}(t) \tag{12.99}$$

becomes

$$-i\omega' a(\omega') = -i\omega a(\omega') - \frac{\kappa}{2} a(\omega') - \sqrt{\kappa}\, a_{\rm in}(\omega'), \tag{12.100}$$

[4] Here, we are following M. J. Collett and C. W. Gardiner, *op. cit.*

or

$$\sqrt{\kappa}\, a_{\rm in}(\omega') = \left[i(\omega' - \omega) - \frac{\kappa}{2} \right] a(\omega'). \tag{12.101}$$

Then we can use the input–output relation (12.92) to eliminate the cavity field $a(\omega') = [a_{\rm out}(\omega') - a_{\rm in}(\omega')]/\sqrt{\kappa}$, with the result

$$\kappa a_{\rm in}(\omega') = \left[i(\omega' - \omega) - \frac{\kappa}{2} \right] [a_{\rm out}(\omega') - a_{\rm in}(\omega')], \tag{12.102}$$

or

$$\left[i(\omega' - \omega) + \frac{\kappa}{2} \right] a_{\rm in}(\omega') = \left[i(\omega' - \omega) - \frac{\kappa}{2} \right] a_{\rm out}(\omega'), \tag{12.103}$$

so that

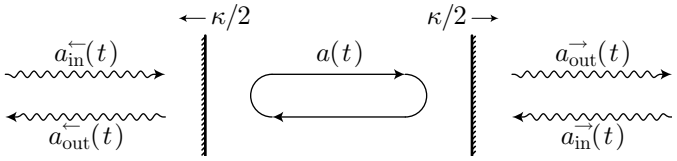
$$a_{\rm out}(\omega') = \left(\frac{i(\omega' - \omega) + \kappa/2}{i(\omega' - \omega) - \kappa/2} \right) a_{\rm in}(\omega').$$

(input–output relation for one-sided cavity) (12.104)

This equation demonstrates that in steady state, the intensity reflected at each frequency is the same as the incident intensity (there can be no net flux). However, there is a phase shift that vanishes at large detunings away from the cavity resonance (so that the reflection is mostly directly from the output mirror), and that becomes a π phase shift exactly at resonance, when the output field is entirely radiated by the cavity. Note that we also see here the explicit phase convention for reflecting from the output mirror, for which the field acquires a factor of -1 ($+1$ for a reflection from the same mirror, but from inside the cavity).

12.2.7 Example: Cavity Transmission

Now if we consider a *two-sided* cavity,[5] the generalization is straightforward: we have input and output operators for each dissipation channel. We will use the notation $a_{\rm in}^{\leftarrow}(t)$ and $a_{\rm out}^{\leftarrow}(t)$ for the input and output to the left-hand-side mirror, $a_{\rm in}^{\rightarrow}(t)$ and $a_{\rm out}^{\rightarrow}(t)$ for the input and output to the right-hand-side mirror.



We can generalize the Langevin equation (12.77) by simply adding both inputs,

$$\partial_t a = -i\omega a(t) - \frac{\kappa}{2} a(t) - \sqrt{\frac{\kappa}{2}} \left[a_{\rm in}^{\leftarrow}(t) + a_{\rm in}^{\rightarrow}(t) \right], \tag{12.105}$$

begin careful to add damping terms for each input. Here we assume a symmetric cavity, with decay rate $\kappa/2$ for each mirror, and thus a *total* decay rate of κ. Switching to frequency space as before,

$$\left[-i(\omega' - \omega) + \frac{\kappa}{2} \right] a(\omega') = -\sqrt{\frac{\kappa}{2}} \left[a_{\rm in}^{\leftarrow}(\omega') + a_{\rm in}^{\rightarrow}(\omega') \right], \tag{12.106}$$

and stipulating that there is no input from the right-hand side ($a_{\rm in}^{\rightarrow} = 0$),

$$\left[i(\omega' - \omega) - \frac{\kappa}{2} \right] a(\omega') = \sqrt{\frac{\kappa}{2}}\, a_{\rm in}^{\leftarrow}(\omega'). \tag{12.107}$$

Now noting that the input–output relation (12.92) holds for each pair of input and output operators separately (since they represent separate baths),

$$a_{\rm out}^{\leftarrow}(\omega') - a_{\rm in}^{\leftarrow}(\omega') = \sqrt{\frac{\kappa}{2}}\, a(\omega') = a_{\rm out}^{\rightarrow}(\omega') - a_{\rm in}^{\rightarrow}(\omega'), \tag{12.108}$$

[5] Again, we are following M. J. Collett and C. W. Gardiner, *op. cit.*

again with $a_{\text{in}}^{\rightarrow} = 0$, we can eliminate $a(\omega')$ in favor of $a_{\text{out}}^{\rightarrow}(\omega')$ to obtain

$$\left[i(\omega' - \omega) - \frac{\kappa}{2}\right] a_{\text{out}}^{\rightarrow}(\omega') = \frac{\kappa}{2}\, a_{\text{in}}^{\leftarrow}(\omega'), \tag{12.109}$$

so that the transmitted field becomes

$$a_{\text{out}}^{\rightarrow}(\omega') = \left[\frac{\kappa/2}{i(\omega' - \omega) - \kappa/2}\right] a_{\text{in}}^{\leftarrow}(\omega'). \tag{12.110}$$
(transmitted field)

This is a Lorentzian response with a full-width at half-maximum of κ, and perfect transmission on resonance, $a_{\text{out}}^{\rightarrow}(\omega) = a_{\text{in}}^{\leftarrow}(\omega)$. Then from Eq. (12.108),

$$a_{\text{out}}^{\leftarrow}(\omega') = a_{\text{in}}^{\leftarrow}(\omega') + a_{\text{out}}^{\rightarrow}(\omega') = a_{\text{in}}^{\leftarrow}(\omega') + \left[\frac{\kappa/2}{i(\omega' - \omega) - \kappa/2}\right] a_{\text{in}}^{\leftarrow}(\omega'), \tag{12.111}$$

and thus the output field becomes

$$a_{\text{out}}^{\leftarrow}(\omega') = \left[\frac{i(\omega' - \omega)}{i(\omega' - \omega) - \kappa/2}\right] a_{\text{in}}^{\leftarrow}(\omega'), \tag{12.112}$$
(reflected field)

which is, of course, whatever didn't make it through the cavity. Returning to Eq. (12.107), we find that the internal field is given by

$$a(\omega') = \left[\frac{\sqrt{\kappa/2}}{i(\omega' - \omega) - \kappa/2}\right] a_{\text{in}}^{\leftarrow}(\omega'). \tag{12.113}$$
(transmitted field)

This is again a Lorentzian response, with a maximum amplitude at resonance of $-\sqrt{\kappa/2}\, a_{\text{in}}^{\leftarrow}(\omega' = \omega)$. In terms of intensities, this is

$$|a(\omega' = \omega)|^2 = \frac{2}{\kappa}|a_{\text{in}}^{\leftarrow}(\omega' = \omega)|^2. \tag{12.114}$$

From Eq. (12.43) the decay rate is $\kappa = 2\pi/\tau_{\text{rt}}\mathscr{F} = 2\pi(\text{FSR})/\mathscr{F}$ in terms of the finesse and round-trip time (free spectral range), so

$$|a(\omega' = \omega)|^2 = \frac{\mathscr{F}}{2\pi}\left(\frac{2|a_{\text{in}}^{\leftarrow}(\omega' = \omega)|^2}{\text{FSR}}\right). \tag{12.115}$$
(cavity buildup)

The usual classical result is that the cavity intensity builds up to a factor of $\mathscr{F}/2\pi$ over the input intensity (for large $\mathscr{F}$); the factor of FSR/2 here emphasizes the different normalization of the input/output operators compared to the cavity operator, since the cavity operator is defined in a bounded region, while the input-output operators are defined on unbounded ones.

12.2.8 Example: Driven Cavity

Suppose we have a one-sided cavity driven by a classical field. Then the cavity Hamiltonian is

$$H_{\text{cav}} = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right), \tag{12.116}$$

and the Hamiltonian for the external driving laser at frequency ω_{L} is, from Eq. (12.34),

$$H_{\text{L}} = \hbar\mathscr{E}\left(ae^{i\omega_{\text{L}}t} + a^\dagger e^{-i\omega_{\text{L}}t}\right). \tag{12.117}$$

The quantum Langevin equation (12.77) is thus modified to include the free evolution of both Hamiltonians, with the result

$$\partial_t a = -i\omega a(t) - i\mathscr{E}e^{-i\omega_{\text{L}}t} - \frac{\kappa}{2}a(t) - \sqrt{\kappa}\, a_{\text{in}}(t). \tag{12.118}$$

Defining the rotating-frame operator

$$\tilde{a}(t) := a(t)e^{i\omega_{\rm L}t} \tag{12.119}$$

(with a similar definition for $\tilde{a}_{\rm in}$), the Langevin equation becomes

$$\partial_t\tilde{a} = (\partial_t a)e^{i\omega_{\rm L}t} + i\omega_{\rm L}\tilde{a} = i(\omega_{\rm L}-\omega)\tilde{a}(t) - i\mathscr{E} - \frac{\kappa}{2}\tilde{a}(t) - \sqrt{\kappa}\,\tilde{a}_{\rm in}(t). \tag{12.120}$$

With the external bath in the vacuum state, the expectation value of this equation is

$$\partial_t\langle\tilde{a}\rangle = i(\omega_{\rm L}-\omega)\langle\tilde{a}(t)\rangle - i\mathscr{E} - \frac{\kappa}{2}\langle\tilde{a}(t)\rangle\,. \tag{12.121}$$

In steady state, $\partial_t\langle\tilde{a}\rangle = 0$, so

$$\alpha := \langle\tilde{a}(t\to\infty)\rangle = \frac{i\mathscr{E}}{i(\omega_{\rm L}-\omega)-\kappa/2}. \tag{12.122}$$
(cavity steady state)

But due to the presence of the vacuum input field in Eq. (12.120), the field $a(t)$ fluctuates about this value. In fact, we have already seen in Section 5.6.1.3 that the steady state of the damped harmonic oscillator is a coherent state. That is consistent with what we see here, since if the steady state is $|\alpha\rangle$, then by definition the coherent state satisfies $a|\alpha\rangle = \alpha|\alpha\rangle$, so that $\langle a\rangle = \langle\alpha|a|\alpha\rangle = \alpha$. The effect of the input is then to superpose vacuum fluctuations on the classical motion $\alpha e^{-i\omega_{\rm L}t}$ of the cavity steady state.

On resonance, then, the steady-state amplitude is given by

$$|\alpha|^2 = \frac{4\mathscr{E}^2}{\kappa^2} = \frac{4P}{\kappa\hbar\omega} = \frac{2\mathscr{F}P}{\pi(\mathrm{FSR})\hbar\omega} \tag{12.123}$$
(resonant cavity steady state)

in terms of the input power P, where we used $\mathscr{E} = \sqrt{\kappa P/\hbar\omega}$ from Eq. (12.40). Recall that for a coherent state $|\alpha|^2$ is the mean photon number, since $\langle\alpha|a^\dagger a|\alpha\rangle = |\alpha|^2$.

Finally, note that rather than include a driving Hamiltonian, we could have specified the classical driving field as part of the input field. That is, if the quantum Langevin equation *without* the contribution from $H_{\rm L}$ is

$$\partial_t a = -i\omega a(t) - \frac{\kappa}{2}a(t) - \sqrt{\kappa}\,a_{\rm in}(t), \tag{12.124}$$

Then making the replacement

$$a_{\rm in}(t) \longrightarrow \frac{i\mathscr{E}}{\sqrt{\kappa}}e^{-i\omega_{\rm L}t} + a_{\rm in}(t) \tag{12.125}$$

leads to the same driven Langevin equation (12.118). That is, the input field is in a coherent state, with amplitude $\alpha/2$ [the same factor of 2 that we saw in Eq. (12.115) crops up here as well], and again $a_{\rm in}(t)$ represents the input-field vacuum fluctuations.

12.2.9 Example: Atomic Motion in an Optical Cavity

Now consider a single atom in a resonantly driven cavity. The cavity Hamiltonian is again

$$H_{\rm cav} = \hbar\omega\left(a^\dagger a + \frac{1}{2}\right), \tag{12.126}$$

the Hamiltonian for the external driving laser exactly on resonance, $\omega_{\rm L} = \omega$, is again

$$H_{\rm L} = \hbar\mathscr{E}\left(ae^{i\omega t} + a^\dagger e^{-i\omega t}\right), \tag{12.127}$$

the free evolution of the atom is

$$H_{\rm A} = \frac{p^2}{2m} + \hbar\omega_0\sigma^\dagger\sigma, \tag{12.128}$$

and the Jaynes–Cummings Hamiltonian (10.9) for the atom–field interaction is

$$H_{\rm AF} = \hbar g(x)\left(\sigma a^\dagger + \sigma^\dagger a\right), \tag{12.129}$$

where

$$g(x) = g\cos(kx) \tag{12.130}$$

gives the spatial dependence of the cavity QED coupling rate due to the standing-wave field in the cavity along the longitudinal direction.

With this collection of Hamiltonians, the quantum Langevin equation (12.77) becomes

$$\partial_t a = -i\omega a(t) - i\mathscr{E}e^{-i\omega t} - ig\cos(kx)\sigma(t) - \frac{\kappa}{2}a(t) - \sqrt{\kappa}\,a_{\rm in}(t), \tag{12.131}$$

which becomes

$$\partial_t \tilde{a} = -i\mathscr{E} - ig\cos(kx)\tilde{\sigma}(t) - \frac{\kappa}{2}\tilde{a}(t) - \sqrt{\kappa}\,\tilde{a}_{\rm in}(t), \tag{12.132}$$

upon entering the rotating frame of the laser field, with $\tilde{a}(t) = a(t)e^{i\omega t}$ and $\tilde{\sigma}(t) = \sigma(t)e^{i\omega t}$. Similarly, the general quantum Langevin equation (12.93) gives

$$\partial_t \sigma = -i\omega_0\sigma(t) - ig\cos(kx)[\sigma(t), \sigma^\dagger(t)]a - \frac{\Gamma}{2}\sigma(t) - \sqrt{\Gamma}\sigma_{\rm in}(t), \tag{12.133}$$

where we used $\sigma^2 = 0$ and $[\sigma, \sigma^\dagger\sigma] = \sigma$. Also, we are including dissipation of the atom via spontaneous emission, where Γ is the decay rate into modes other than the cavity mode (which is approximately the free-space rate, so long as the cavity does not subtend a large solid angle from the atom). This dissipation has the same form as for the cavity, but with the replacements $\kappa \longrightarrow \Gamma$ and $a \longrightarrow \sigma$. In the rotating frame, this becomes

$$\partial_t \tilde{\sigma} = i\Delta\tilde{\sigma}(t) - ig\cos(kx)[\tilde{\sigma}(t), \tilde{\sigma}^\dagger(t)]\tilde{a}(t) - \frac{\Gamma}{2}\tilde{\sigma} - \sqrt{\Gamma}\tilde{\sigma}_{\rm in}(t), \tag{12.134}$$

where we now have the usual detuning

$$\Delta := \omega - \omega_0 \tag{12.135}$$

of the laser field from the atomic resonance. Working in the limit of large detuning $|\Delta| \gg \Gamma, \kappa, g$, the atom is only weakly excited, and thus we can take $\sigma\sigma^\dagger \approx 1$ and $\sigma^\dagger\sigma \approx 0$, since the expectation values of these operators represent the populations of the ground and excited states, respectively. Thus, we have

$$\partial_t \tilde{\sigma} = \left(i\Delta - \frac{\Gamma}{2}\right)\tilde{\sigma}(t) - ig\cos(kx)\tilde{a}(t) - \sqrt{\Gamma}\tilde{\sigma}_{\rm in}(t) \tag{12.136}$$

for the atomic evolution.

12.2.9.1 Adiabatic Approximation

Now to make the **adiabatic approximation**. The idea is that the time scale for atomic motion is much slower than any of the time scales $|\Delta|, \Gamma, \kappa, g$ representing the cavity or internal atomic dynamics. Since these other processes are explicitly damped at the (fast) rates $\kappa/2$ and $\Gamma/2$, respectively, we can assume that they are always in quasi-equilibrium with respect to the atomic motion, so that $\partial_t\tilde{\sigma} = 0$ and $\partial_t\tilde{a} = 0$. This is sensible since to a good approximation, the atom will not respond to the fast fluctuations in $\tilde{\sigma}(t)$ and $\tilde{a}(t)$. Further to this end, then, we can set the input noise terms in Eqs. (12.132) and (12.136) to zero. This is equivalent to replacing $\tilde{\sigma}(t)$ and $\tilde{a}(t)$ by their expectation values $\langle\tilde{\sigma}(t)\rangle$ and $\langle\tilde{a}(t)\rangle$, again since, to a first approximation, the atomic motion should not respond to the fast fluctuations, so long as we are careful about the procedure. Thus, Eq. (12.136) becomes

$$\tilde{\sigma}(t) \approx \frac{ig\cos(kx)}{i\Delta - \Gamma/2}\tilde{a}(t) \approx \frac{g}{\Delta}\cos(kx)\tilde{a}(t), \tag{12.137}$$

and Eq. (12.132) becomes

$$0 \approx -i\mathscr{E} - ig\cos(kx)\tilde{\sigma}(t) - \frac{\kappa}{2}\tilde{a}(t) \approx -i\mathscr{E} - i\frac{g^2}{\Delta}\cos^2(kx)\tilde{a}(t) - \frac{\kappa}{2}\tilde{a}(t) \tag{12.138}$$

$$\tilde{a}(t) \approx \frac{-i\mathscr{E}}{i\dfrac{g^2}{\Delta}\cos^2(kx) + \dfrac{\kappa}{2}} \tag{12.139}$$

As in Section 12.2.8, we can define

$$\alpha := \frac{2\mathscr{E}}{\kappa} \tag{12.140}$$

as the free-cavity coherent-state amplitude. Then expanding Eq. (12.139) to lowest order in Δ^{-1},

$$\tilde{a}(t) \approx \frac{-i\alpha}{1 + i\dfrac{2g^2}{\kappa\Delta}\cos^2(kx)} \approx -i\alpha\left[1 - i\frac{g^2}{\kappa\Delta}\cos^2(kx)\right], \tag{12.141}$$

so that we now have adiabatic expressions for both the atom and cavity lowering operators.

The atomic Hamiltonian (12.128) involves the combination $\sigma^\dagger\sigma$, so it is also useful to derive an adiabatic relation for this operator. Note that for weak excitation, as we implicitly saw in coherent vs. incoherent scattering from the two-level atom in Section 5.7.2,

$$\langle\sigma^\dagger\sigma\rangle \approx \langle\sigma^\dagger\rangle\langle\sigma\rangle\,. \tag{12.142}$$

This is because in the (far-detuned) weak-excitation limit, the spontaneous-emission rate from Eq. (5.272) for classical-field excitation with Rabi frequency Ω is $\Gamma\Omega^2/4\Delta^2$. The correction to factoring the expectation value in Eq. (12.142) is $O(\Delta^{-2})$, and thus ignorable in this regime. From Eq. (12.137)

$$\tilde{\sigma}^\dagger\tilde{\sigma} \approx \frac{g^2}{\Delta^2}\cos^2(kx)a^\dagger a, \tag{12.143}$$

since nothing changes upon transforming out of the rotating frame. The atomic Hamiltonian (12.144), once transformed into the rotating frame, becomes

$$H_{\text{eff}} = \frac{p^2}{2m} - \hbar\Delta\tilde{\sigma}^\dagger\tilde{\sigma}, \tag{12.144}$$

since this Hamiltonian correctly generates the first term in the Heisenberg equation (12.134). But with Eq. (12.143), we have the effective Hamiltonian

$$\boxed{H_{\text{eff}} = \frac{p^2}{2m} - \hbar\frac{g^2}{\Delta}\cos^2(kx)a^\dagger a}$$

(effective Hamiltonian for atomic motion) (12.145)

for the atomic motion in response to the cavity field $a^\dagger a$ in the adiabatic approximation.

Now considering the motion of $a^\dagger a$ under H_{eff}, which now replaces H_{A} and H_{AF}, we can use Eq. (12.93) to obtain

$$\begin{aligned}\partial_t a^\dagger a &= i\mathscr{E}\left(\tilde{a} - \tilde{a}^\dagger\right) - \left\{a^\dagger\left(\frac{\kappa}{2}a + \sqrt{\kappa}\,a_{\text{in}}(t)\right) + \left(\frac{\kappa}{2}a^\dagger + \sqrt{\kappa}\,a_{\text{in}}^\dagger(t)\right)a\right\} \\ &= i\mathscr{E}\left(\tilde{a} - \tilde{a}^\dagger\right) - \kappa a^\dagger a - \sqrt{\kappa}\left[\tilde{a}^\dagger\tilde{a}_{\text{in}}(t) - \tilde{a}_{\text{in}}^\dagger(t)\tilde{a}\right]\end{aligned} \tag{12.146}$$

where we used $[a^\dagger a, a] = [a^\dagger, a]a = -a$ and $[a^\dagger a, a^\dagger] = a^\dagger[a, a^\dagger] = a^\dagger$, with the same relations holding in the rotating frame, where $a^\dagger a = \tilde{a}^\dagger\tilde{a}$. In the adiabatic approximation, we again set the noise terms to zero on average (noting that, e.g., $a(t)$ and $a_{\text{in}}(t)$ are statistically independent), and set the time derivative to zero, with the result

$$a^\dagger a \approx \frac{i\mathscr{E}}{\kappa}\left(\tilde{a} - \tilde{a}^\dagger\right) = -\alpha\frac{\left(\tilde{a} - \tilde{a}^\dagger\right)}{2i}. \tag{12.147}$$

Now using Eq. (12.141),

$$a^\dagger a \approx \alpha^2. \tag{12.148}$$

Then the effective Hamiltonian (12.145) becomes

$$H_{\text{eff}} \approx \frac{p^2}{2m} - \hbar\frac{\alpha^2 g^2}{\Delta}\cos^2(kx), \tag{12.149}$$

(effective Hamiltonian for atomic motion)

which represents the *average* effective potential seen by the atoms. Of course, the operator nature of Eq. (12.145) shows that there will be *fluctuations* in the effective potential due to fluctuations in the cavity field, which are in turn due to cavity decay. Physically, think of it this way: each time a photon escapes the cavity at a random time, α^2 jumps discontinuously downward by 1, and this amplitude smoothly recovers due to the driving field until the next random jump.

12.3 Exercises

Problem 12.1

Derive the decay rate for an optical cavity in the high-finesse limit using Fermi's Golden Rule. Assume that the decay occurs on the continuous family $|n\rangle \longrightarrow |n-1, 1_q\rangle$ of transitions.

Problem 12.2

Derive the input–output Heisenberg equations of motion (12.93) and (12.94)

$$\begin{aligned}\partial_t c &= -\frac{i}{\hbar}\,[c, H_{\rm sys}] - \left\{[c, a^\dagger]\left(\frac{\kappa}{2}a + \sqrt{\kappa}\,a_{\rm in}(t)\right) - \left(\frac{\kappa}{2}a^\dagger + \sqrt{\kappa}\,a_{\rm in}^\dagger(t)\right)[c, a]\right\} \\ \partial_t c &= -\frac{i}{\hbar}\,[c, H_{\rm sys}] - \left\{[c, a^\dagger]\left(-\frac{\kappa}{2}a + \sqrt{\kappa}\,a_{\rm out}(t)\right) - \left(-\frac{\kappa}{2}a^\dagger + \sqrt{\kappa}\,a_{\rm out}^\dagger(t)\right)[c, a]\right\}\end{aligned} \tag{12.150}$$

for an arbitrary system operator c, for the system and bath interaction as we set up in class.

Problem 12.3

Use the general Born–Markov master-equation formalism (Section 4.5) to derive the master equation for cavity decay (i.e., for a damped harmonic oscillator, coupled to a reservoir in the vacuum state). You should end up with the usual Lindblad master equation of the form

$$\partial_t \rho = -\frac{i}{\hbar}[H, \rho] + \kappa D[a]\rho. \tag{12.151}$$

To make things simpler, ignore the Lamb shift, and assume an equal coupling of the cavity to all reservoir modes (i.e., a frequency-independent decay rate).

Problem 12.4

(a) Show for a damped cavity, where the cavity annihilation operator a satisfies the quantum Langevin equation

$$\partial_t a = -i\omega a(t) - \frac{\kappa}{2}a(t) - \sqrt{\kappa}\,a_{\rm in}(t), \tag{12.152}$$

that for a bath in the vacuum state, that

$$\partial_t \langle a\rangle = -i\omega\langle a(t)\rangle - \frac{\kappa}{2}\langle a(t)\rangle\,. \tag{12.153}$$

(b) The above result might seem alarming, as it appears that operators are damping away to nothing. However, the noisy character of the input operator in the vacuum state compensates for the damping. To illustrate this, show explicitly that the commutator $[a, a^\dagger]$ is time-invariant in spite of the damping.

Problem 12.5

Consider a single-mode optical cavity with mode annihilation operator $a(t)$, with corresponding input and output operators $a_{\rm in}(t)$ and $a_{\rm out}(t)$, respectively, and decay rate κ.

(a) Show that $[a_{\rm out}^\dagger(t), a_{\rm out}(t')] = [a_{\rm in}^\dagger(t), a_{\rm in}(t')]$. What is the physical interpretation of this result?

(b) Show that $\left\langle a_{\rm out}^\dagger(t)\, a_{\rm out}(t')\right\rangle = \kappa\langle a^\dagger(t)\, a(t')\rangle$, assuming a vacuum input. What is the physical interpretation of this result?

(c) Show that $\langle a_{\rm out}(t)\, a_{\rm out}(t')\rangle = \kappa\langle a[\max(t,t')]\, a[\min(t,t')]\rangle$, assuming a vacuum input. What is the physical interpretation of this result?

Chapter 13

Mechanical Effects of the Quantum Vacuum

Here we will examine mechanical forces on atoms due to the quantum vacuum. Of course, there is no absorption from the vacuum, but scattering of virtual photons can still produce forces. Complementary to these forces are shifts of atomic energy levels, which we will also calculate. The main effect we will examine is the **Casimir–Polder effect**,[1] where a ground-state atom near a conducting surface, with the field in the vacuum state, is attracted to the surface. We will also investigate the **Lamb shift** of atomic transitions, which we can interpret as an ac Stark shift due to the vacuum field. Since the introduction of a conducting boundary modifies the field modes, we can thus interpret the Casimir–Polder effect as a space-dependent Lamb shift due to the local modification of the vacuum field modes. (However, unlike the Lamb shift, which must be observed by probing *transitions* among atomic levels, the Casimir–Polder shift is observable for an atom in a *single* level due to the mechanical action of the shift.)

13.1 Setup

We want to compute the effect of the vacuum field due to a plane conductor on an atom. We can write the Hamiltonian for the free atom and field as

$$H_0 = \sum_j \hbar\omega_{j0} |\mathrm{e}_j\rangle\langle \mathrm{e}_j| + \sum_{\mathbf{k},\zeta} \hbar\omega_{\mathbf{k}} \left(a^\dagger_{\mathbf{k},\zeta} a_{\mathbf{k},\zeta} + \frac{1}{2} \right), \tag{13.1}$$

where as usual, the ground state has zero energy, $\omega_{j0} = (E_j - E_0)/\hbar$ is the transition frequency of the $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}_j\rangle$ atomic transition, the wave vector $\mathbf{k}$ labels the field modes of different frequency and orientation, the index ζ labels the two independent polarizations, and $a_{\mathbf{k},\zeta}$ is the annihilation operator for the $(\mathbf{k},\zeta)$ mode. We will write the eigenstates that we need of the free Hamiltonian in the form $|\alpha, n_{\mathbf{k},\zeta}\rangle$, which means that the atom is in state $|\alpha\rangle$, while the field mode $(\mathbf{k},\zeta)$ has n photons (other modes not explicitly labeled are in the vacuum state).

[1] The original, and very readable, reference is of course H. B. G. Casimir and D. Polder, ''The Influence of Retardation on the London-van der Waals Forces,'' *Physical Review* **73**, 360 (1948) (doi: 10.1103/PhysRev.73.360). However, they only obtain asymptotic results. Our results here, which include the intermediate regime, are consistent with the results of the seminal calculation of G. Barton, ''Frequency shifts near an interface: Inadequacy of two-level atomic models,'' *Journal of Physics B* **7**, 2134 (1974) (doi: 10.1088/0022-3700/7/16/012); as well as the more recent calculations by A. O. Barut and J. P. Dowling, ''Quantum electrodynamics based on self-energy, without second quantization: The Lamb shift and long-range Casimir-Polder van der Waals forces near boundaries,'' *Phys. Rev. A*, **36**, 2550 (1987) (doi: 10.1103/PhysRevA.36.2550); and D. Meschede, W. Jhe, and E. A. Hinds, ''Radiative properties of atoms near a conducting plane: An old problem in a new light,'' *Phys. Rev. A*, **41**, 1587 (1990) (doi: 10.1103/PhysRevA.41.1587). See also S. Haroche, ''Cavity Quantum Electrodynamics,'' in *Fundamental Systems in Quantum Optics: Proceedings of the Les Houches Summer School, Session LIII, 1990*, J. Dalibard, J. M. Raimond, and J. Zinn-Justin, Eds. (Elsevier, 1992), Course 13, p. 767.

The quantum electric field from Eq. (8.56) is

$$\mathbf{E}(\mathbf{r},t) = -\sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}} \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) a_{\mathbf{k},\zeta}(t) + \text{H.c.}, \tag{13.2}$$

where the $a_{\mathbf{k},\zeta}$ are the field annihilation operators, and the unit-normalized mode functions of frequency $\omega_{\mathbf{k}}$ in half-space from Section 8.4.3 are

$$\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) = \sqrt{\frac{2}{V}} \left(\hat{\varepsilon}_{\mathbf{k},\zeta,\parallel} \sin k_z z - i \hat{\varepsilon}_{\mathbf{k},\zeta,z} \cos k_z z \right) e^{i\mathbf{k}_\parallel \cdot \mathbf{r}}, \tag{13.3}$$

where $V = L^3$ is the quantization volume. Recall that we are applying perfectly conducting boundary conditions at $z = 0$ and $z = L$, and periodic boundary conditions at $x = 0$ and $x = L$, as well as $y = 0$ and $y = L$. In the above mode functions, $\hat{\varepsilon}_{\mathbf{k},\zeta}$ is the unit polarization vector of the mode, and the $\hat{x}_\alpha$ are the Cartesian unit vectors along the x_α-direction. Recall also that the wave vectors are given by

$$k_x = \frac{2\pi n_x}{L}, \qquad k_y = \frac{2\pi n_y}{L}, \qquad k_z = \frac{\pi n_z}{L}, \tag{13.4}$$

where n_x and n_y are any integers, and n_z is nonnegative. In the above quantization volume, we regard the atom as being located at $(L/2, L/2, z)$, although as we will see the transverse location is unimportant.

13.2 Atom–Vacuum Interaction

We now account for the coupling of the atom with the field. We use the usual dipole form of the atom–field interaction Hamiltonian

$$H_{\text{AF}} = -\mathbf{d} \cdot \mathbf{E}, \tag{13.5}$$

where we neglect the contribution of the polarization term in Eq. (9.38) in this gauge. With the above expression (13.2) of the electric field, along with the usual form of the dipole operator, we can write the interaction explicitly *without* the rotating-wave approximation as

$$H_{\text{AF}} = -\sum_j \sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}} (\sigma_j + \sigma_j^\dagger) \langle \text{g}|\mathbf{d}|\text{e}_j\rangle \cdot \left[\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\, a_{\mathbf{k},\zeta} + \mathbf{f}^*_{\mathbf{k},\zeta}(\mathbf{r})\, a^\dagger_{\mathbf{k},\zeta} \right], \tag{13.6}$$

where $\sigma_j := |\text{g}\rangle\langle j|$, and we have assumed the dipole matrix element $\langle \text{g}|\mathbf{d}|\text{e}_j\rangle$ to be real.

The idea here is to compute the energy shift of the ground state $|\text{g}\rangle$ in perturbation theory, which will give the standard Casimir–Polder potential. Up to second order,

$$V_{\text{CP}} = \langle \text{g}|H_0|\text{g}\rangle + \langle \text{g}|H_{\text{AF}}|\text{g}\rangle + \sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle \text{g}|H_{\text{AF}}|\text{e}_j, 1_{\mathbf{k},\zeta}\rangle|^2}{E_{\text{g},0} - E_{\text{e}_j,1_{\mathbf{k},\zeta}}}. \tag{13.7}$$

The first two terms vanish, and the only nonvanishing contribution in the second-order term from H_{AF} will be from the $\sigma_j a_{\mathbf{k},\zeta}$ and $\sigma_j^\dagger a^\dagger_{\mathbf{k},\zeta}$ terms, which are not energy-conserving and thus are usually dropped in the rotating-wave approximation. Then we insert the atom and field energies:

$$V_{\text{CP}} = -\sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle \text{g}|H_{\text{AF}}|\text{e}_j, 1_{\mathbf{k},\zeta}\rangle|^2}{\hbar(\omega_{j0} + \omega_{\mathbf{k}})}. \tag{13.8}$$

Using Eq. (13.6), we can write

$$V_{\text{CP}}(\mathbf{r}) = -\sum_j \sum_{\mathbf{k},\zeta} \frac{\omega_{\mathbf{k}}}{2\epsilon_0} \frac{|\langle \text{g}|\mathbf{d}|\text{e}_j\rangle \cdot \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2}{(\omega_{j0} + \omega_{\mathbf{k}})} \tag{13.9}$$

if the atom is at position $\mathbf{r}$ and using Eq. (13.3) for the mode functions,

$$V_{\text{CP}}(z) = -\sum_j \sum_{\mathbf{k},\zeta} \frac{\omega_{\mathbf{k}}}{\epsilon_0 V} \frac{\left|\langle \text{g}|\hat{\varepsilon}_{\mathbf{k},\zeta,\parallel} \cdot \mathbf{d}|\text{e}_j\rangle\right|^2 \sin^2 k_z z + \left|\langle \text{g}|\hat{\varepsilon}_{\mathbf{k},\zeta,z} \cdot \mathbf{d}|\text{e}_j\rangle\right|^2 \cos^2 k_z z}{(\omega_{j0} + \omega_{\mathbf{k}})}, \tag{13.10}$$

where we have dropped cross-terms involving the parallel and perpendicular components of the dipole operator [these vanish in the upcoming angular integrals—convince yourself of this after working up through Eqs. (13.25)]. For compactness, we will henceforth write

$$d_{j,\parallel}^2 \equiv d_{j,x}^2 + d_{j,y}^2 = \left|\langle \text{g}|\hat{x} \cdot \mathbf{d}|\text{e}_j\rangle\right|^2 + \left|\langle \text{g}|\hat{y} \cdot \mathbf{d}|\text{e}_j\rangle\right|^2, \qquad d_{j,z}^2 \equiv \left|\langle \text{g}|\hat{z} \cdot \mathbf{d}|\text{e}_j\rangle\right|^2, \tag{13.11}$$

so that Eq. (13.10) becomes

$$V_{\text{CP}} = -\frac{1}{\epsilon_0 V} \sum_j \sum_{\mathbf{k},\zeta} \frac{\omega_{\mathbf{k}}}{(\omega_{j0} + \omega_{\mathbf{k}})} \left\{ [(\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{x})^2 + (\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{y})^2] \frac{1}{2} d_{j,\parallel}^2 \sin^2 k_z z + (\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{z})^2 d_{j,z}^2 \cos^2 k_z z \right\}, \tag{13.12}$$

where $d_{j,x}^2 = d_{j,y}^2$ for a spherically symmetric atom (technically, we are not assuming a spherically symmetric atom, but because of the symmetry of the setup, the x and y dependence of the solution must be equivalent to that of a spherically symmetric atom), and we have again discarded vanishing cross-terms. All that remains is to evaluate the sums over atomic states and field modes.

The polarization sum is easy to evaluate, using the result from Eq. (8.190)

$$\sum_\zeta (\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{r}_\alpha)(\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{r}_\beta) = \delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}, \tag{13.13}$$

which becomes

$$\sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{r}_\alpha|^2 = 1 - \frac{k_\alpha^2}{k^2}, \tag{13.14}$$

for $\alpha = \beta$, and we have introduced the absolute value to handle complex polarization vectors (in half space we may assume they are real, and the polarization sum is independent of complex basis transformations). In particular, we may write the two required sums

$$\begin{aligned} \sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta,z}|^2 &= \sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{z}|^2 = 1 - \frac{k_z^2}{k^2} \\ \sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta,\parallel}|^2 &= \sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{x}|^2 + \sum_\zeta |\hat{\varepsilon}_{\mathbf{k},\zeta} \cdot \hat{y}|^2 = 2 - \frac{k_x^2}{k^2} - \frac{k_y^2}{k^2} = 1 + \frac{k_z^2}{k^2}. \end{aligned} \tag{13.15}$$

Then we can perform the polarization sums in the level shift, with the result

$$V_{\text{CP}} = -\frac{1}{\epsilon_0 V} \sum_j \sum_{\mathbf{k}} \frac{\omega_{\mathbf{k}}}{(\omega_{j0} + \omega_{\mathbf{k}})} \left\{ \frac{1}{2} \left(1 + \frac{k_z^2}{k^2}\right) d_{j,\parallel}^2 \sin^2 k_z z + \left(1 - \frac{k_z^2}{k^2}\right) d_{j,z}^2 \cos^2 k_z z \right\}. \tag{13.16}$$

Expanding out the $\sin^2$ and $\cos^2$ functions,

$$\begin{aligned} V_{\text{CP}} = -\frac{1}{2\epsilon_0 V} \sum_j \sum_{\mathbf{k}} \frac{\omega_{\mathbf{k}}}{(\omega_{j0} + \omega_{\mathbf{k}})} \Bigg\{ &\left[\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right) + \frac{k_z^2}{k^2} \left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) \right] \\ &- \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) + \frac{k_z^2}{k^2} \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right) \right] \cos(2k_z z) \Bigg\}. \end{aligned} \tag{13.17}$$

Before evaluating the sum, we must be careful to remove divergences, which will simplify this expression.

13.3 Renormalization

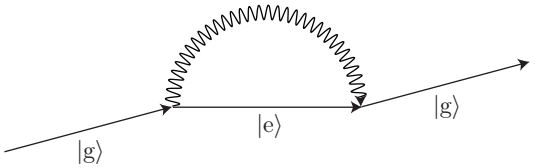
Note that to compute the force on the atom due to the wall, we are really computing the *difference* in energy between this situation and the limit of a large box, where the particle is arbitrarily far away from the surface (we will also take the limit as the box size goes to infinity). That is, the ''renormalized'' potential is

$$\lim_{L,z_0\longrightarrow\infty} V_{\rm CP}\left(\frac{L}{2},\frac{L}{2},z\right) - V_{\rm CP}\left(\frac{L}{2},\frac{L}{2},z_0\right). \tag{13.18}$$

Hence the z-independent terms in Eq. (13.17), which are divergent under the mode summation, do not contribute to the answer, so we can explicitly drop it. Thus, we have

$$V_{\rm CP} = \frac{1}{2\epsilon_0 V}\sum_j \sum_{\mathbf{k}} \frac{\omega_{\mathbf{k}}}{(\omega_{j0}+\omega_{\mathbf{k}})}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) + \frac{k_z^2}{k^2}\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\right]\cos(2k_z z), \tag{13.19}$$

and what remains is to evaluate the wave-vector summation. The terms that we have dropped, which are present in free space, are responsible for the **Lamb shift** that we will return to later. In fact, from the perturbation result (13.9), which contains contributions from the $\sigma^\dagger a^\dagger$ and σa terms of the interaction Hamiltonian, we see that the effect we are considering corresponds to the following ''one-loop'' graph.



This corresponds to a ground-state atom emitting a photon and becoming excited, and then reabsorbing the photon and returning to the ground state. Roughly speaking, if the virtual photon ''bounces off the mirror,'' then we count it as contributing to the Casimir–Polder potential. Otherwise we count it as part of the free-space Lamb shift.

13.4 Large-Box Limit

When the box becomes large ($L \longrightarrow \infty$), the spacing between the modes becomes small. In this limit, an integral of a function is equivalent to a sum weighted by the mode spacings. As usual, we can write

$$\sum_{\mathbf{k}} f(\mathbf{k})\,\Delta k_x\,\Delta k_y\,\Delta k_z \longrightarrow \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z\, f(\mathbf{k}) \tag{13.20}$$

for an arbitrary function $f(\mathbf{k})$. Since

$$\Delta k_x = \frac{2\pi}{L_x}, \qquad \Delta k_y = \frac{2\pi}{L_y}, \qquad \Delta k_z = \frac{\pi}{L_z}, \tag{13.21}$$

we can thus make the formal replacement

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{\pi(2\pi)^2}\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{0}^{\infty} dk_z, \tag{13.22}$$

where $V = L^3$. Thus, we are left with the expression

$$\begin{aligned} V_{\rm CP} &= \frac{1}{8\pi^3\epsilon_0}\sum_j \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{0}^{\infty} dk_z \frac{k}{(k_{j0}+k)}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) + \frac{k_z^2}{k^2}\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\right]\cos(2k_z z) \\ &= \frac{1}{16\pi^3\epsilon_0}\sum_j \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{k}{(k_{j0}+k)}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) + \frac{k_z^2}{k^2}\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\right]\cos(2k_z z), \end{aligned} \tag{13.23}$$

where $\omega_{\mathbf{k}} = ck$, and $\omega_{j0} = ck_{j0}$. We now just need to evaluate the integrals here.

13.5 Spherical Coordinates

In Eq. (13.23), we basically just need to evaluate the integrals

$$\begin{aligned} I_1 &= \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{k}{(k+k_0)} \cos(2k_z z) \\ I_2 &= \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{k_z^2}{k(k+k_0)} \cos(2k_z z). \end{aligned} \tag{13.24}$$

Writing I_1 in spherical coordinates, and carrying out the ϕ (azimuthal) integral,

$$\begin{aligned} I_1 &= 2\pi \int_0^{\infty} dk \int_0^{\pi} d\theta \, \frac{k^3 \sin\theta \cos(2kz\cos\theta)}{k+k_0} \\ &= 2\pi \int_0^{\infty} dk \, \frac{k^3}{(k+k_0)} \frac{\sin(2kz)}{kz} \\ &= -\frac{\pi}{2z} \partial_z^2 \int_0^{\infty} dk \, \frac{\sin(2kz)}{k+k_0}. \end{aligned} \tag{13.25}$$

It appears that we are cheating by pulling the ∂_z^2 out of the integral, since the integral is, strictly speaking, divergent since the integrand asymptotically scales as k for large k. However, in writing down the interaction Hamltonian (13.6), what we neglected is the fact that the atom is not perfectly localized, and the Hamiltonian should in fact be averaged over a localized distribution that represents the atom's extent. If this distribution is $h(\mathbf{r})$, then each factor of $\sin k_z z$ and $\cos k_z z$ in Eq. (13.10) and the subsequent expressions is multiplied by the Fourier transform $\tilde{h}(\mathbf{k})$. Then factors of $\cos 2k_z z$ in the subsequent expressions are multiplied by $\tilde{h}^2(\mathbf{k})$, which cuts off the high-frequency ends of the above integral. For example, if $h(\mathbf{r})$ is a Gaussian function, then $\tilde{h}^2(\mathbf{k})$ is also Gaussian, and we have no problems with the convergence of the integrals. We will not explicitly include these cutoff functions, except to note that they ensure that all the mode integrals are convergent, and omitting them is equivalent to performing the integrals properly and then taking the limit as the size of the atom vanishes.

To evaluate the above integral, we need to introduce the **auxiliary functions to the sine and cosine integrals**,[2] which have the definitions

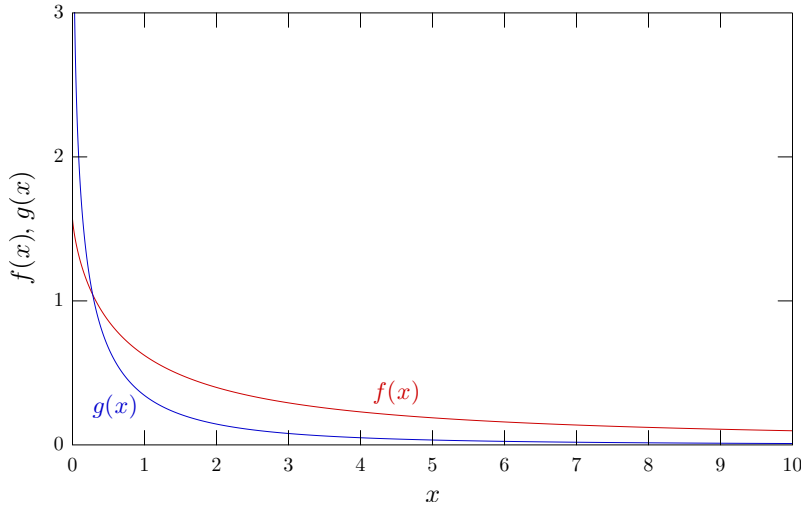
$$\begin{aligned} f(z) &= \sin z \, \mathrm{Ci}(z) + \cos z \left[\frac{\pi}{2} - \mathrm{Si}(z) \right] \\ g(z) &= -\cos z \, \mathrm{Ci}(z) + \sin z \left[\frac{\pi}{2} - \mathrm{Si}(z) \right]. \end{aligned} \tag{13.26}$$

Here, the sine integral $\mathrm{Si}(x)$ and cosine integral $\mathrm{Ci}(x)$ are defined by

$$\begin{aligned} \mathrm{Si}(z) &= \int_0^z \frac{\sin t}{t} \, dt \\ \mathrm{Ci}(z) &= -\int_z^{\infty} \frac{\cos t}{t} \, dt = \gamma + \log z + \int_0^z \frac{\cos t - 1}{t} \, dt, \end{aligned} \tag{13.27}$$

where γ is Euler's constant. The auxiliary functions are plotted here.

[2] See Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), pp. 231-3.



We note that even though they are composed of oscillatory functions, they are not themselves oscillatory. Now using the integral representation of the auxiliary function $f(z)$ from Problem 13.1,[3]

$$\int_0^\infty \frac{\sin(ax)}{x+\beta}\,dx = f(a\beta) \qquad (|\arg\beta| < \pi,\ a > 0), \tag{13.28}$$

we can evaluate the above integral to obtain

$$I_1 = -\frac{\pi}{2z}\partial_z^2 f(2k_0 z). \tag{13.29}$$

We can also write I_2 in spherical coordinates, and after carrying out the ϕ (azimuthal) integral,

$$\begin{aligned} I_2 &= 2\pi \int_0^\infty dk \int_0^\pi d\theta\, \frac{k^3 \sin\theta\cos^2\theta\cos(2kz\cos\theta)}{k+k_0} \\ &= -\frac{\pi}{2}\partial_z^2 \int_0^\infty dk\, \frac{k}{k+k_0}\int_0^\pi d\theta\, \sin\theta\cos(2kz\cos\theta) \\ &= -\frac{\pi}{2}\left(\partial_z^2\frac{1}{z}\right)\int_0^\infty dk\,\frac{\sin(2kz)}{k+k_0} \\ &= -\frac{\pi}{2}\left(\partial_z^2\frac{1}{z}\right) f(2k_0 z). \end{aligned} \tag{13.30}$$

In evaluating the derivatives, it is useful to note the equivalence of the operators

$$\left(\frac{1}{2z}\partial_z^2 - \frac{1}{z^2}\partial_z + \frac{1}{z^3}\right) \equiv \frac{1}{2}\left(\partial_z^2\frac{1}{z}\right), \tag{13.31}$$

where both are assumed to operate on some function of z. Notice that I_1 and I_2 differ only by the ordering of ∂_z^2 and $(1/z)$.

We can then write the Casimir–Polder potential as

$$V_{\text{CP}} = \frac{1}{16\pi^3\epsilon_0}\sum_j \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) I_{1j} + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right) I_{2j}\right], \tag{13.32}$$

where I_{1j} is the same as I_1 and I_{2j} is the same as I_2 but with $k_0 \longrightarrow k_{j0}$. Putting in the above values of I_1 and I_2, we find

$$V_{\text{CP}} = -\frac{1}{(4\pi\epsilon_0)8\pi}\sum_j \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\left(\frac{1}{z}\partial_z^2\right) + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_z^2\frac{1}{z}\right)\right] f(2k_{j0}z). \tag{13.33}$$

[3]See Abramowitz and Stegun, *op. cit.*, or I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 7th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2007), Formula 3.722.1.

This is not quite our final form. To simplify the antisymmetric part (involving the difference of $d^2_{j,\parallel}$ and $d^2_{j,z}$), we must reintroduce the Thomas–Reiche–Kuhn sum rule.

13.5.1 Thomas–Reiche–Kuhn Sum Rule

In the classical analysis of the Lorentz model in Section 1.2.1, we wrote down the Thomas–Reiche–Kuhn (TRK) sum rule

$$\sum_j f_{0j} = 1. \tag{13.34}$$

For a fine-structure transition $J \longrightarrow J'$, we have the expression

$$\Gamma_{J'J} = \frac{e^2\omega^2_{J'J}}{2\pi\epsilon_0 m_e c^3}\frac{2J+1}{2J'+1} f_{JJ'} \tag{13.35}$$

that relates the oscillator strength to the decay rate. We further have the relation

$$\Gamma_{J'J} = \frac{\omega^3_{J'J}}{3\pi\epsilon_0\hbar c^3}\frac{2J+1}{2J'+1}|\langle J\|e\mathbf{r}\|J'\rangle|^2 \tag{13.36}$$

relating the decay rate to the lifetime, in which case the TRK sum rule becomes

$$\sum_{J'} \omega_{JJ'}|d_{JJ'}|^2 = \frac{3\hbar e^2}{2m}. \tag{13.37}$$

This is closer to the usual quantum-mechanical form for the TRK sum rule, which we now derive directly.

Recall that we derived the relation

$$\begin{aligned}\langle 0|p_\alpha|j\rangle &= -i\frac{m}{\hbar}\langle 0|[r_\alpha, H]|j\rangle \\ &= im\omega_{0j}\langle 0|r_\alpha|j\rangle,\end{aligned} \tag{13.38}$$

between the position and momentum matrix elements in Section 9.3.2, Eq. (9.52), where

$$\omega_{0j} := \frac{E_0 - E_j}{\hbar} \tag{13.39}$$

is again the *signed* frequency of the transition between the energy eigenstates. The TRK sum rule follows from the following:

$$\begin{aligned}\sum_j \omega_{0j}|\langle 0|r_\alpha|j\rangle|^2 &= \frac{1}{2}\sum_j (\omega_{0j} - \omega_{j0})\,|\langle 0|r_\alpha|j\rangle|^2 \\ &= -\frac{i}{2m}\sum_j \left(\langle 0|p_\alpha|j\rangle\langle j|r_\alpha|0\rangle - \langle 0|r_\alpha|j\rangle\langle j|p_\alpha|0\rangle\right) \\ &= -\frac{i}{2m}\langle 0|[p_\alpha, r_\alpha]|0\rangle \\ &= -\frac{\hbar}{2m}.\end{aligned} \tag{13.40}$$

Thus, we have established the TRK sum rule (note the subscript ordering on ω_{j0}):

$$\sum_j \omega_{j0}|\langle 0|r_\alpha|j\rangle|^2 = \frac{\hbar}{2m}. \tag{13.41}$$

(Thomas–Reiche–Kuhn sum rule)

Using Eq. (13.38), we can also establish the alternate form of the TRK sum rule:

$$\sum_j \frac{|\langle 0|p_\alpha|j\rangle|^2}{\omega_{j0}} = \frac{m\hbar}{2}. \tag{13.42}$$

In the Casimir–Polder potential, two combinations of dipole sums are particularly useful, as a corollary to the TRK sum rule:

$$\begin{aligned} \sum_j \omega_{j0}\left(d_{j,\parallel}^2 + 2d_{j,z}^2\right) &= \frac{2\hbar e^2}{m} \\ \sum_j \omega_{j0}\left(d_{j,\parallel}^2 - 2d_{j,z}^2\right) &= 0. \end{aligned} \tag{13.43}$$

In writing these relations down, we have used $\hat{r}_\alpha \cdot \mathbf{d}_{0j} = e\langle 0|r_\alpha|j\rangle$. Of course, the label 0 can refer to *any* energy eigenstate, not just the ground state.

As we indicated before in Section 1.2.1, the TRK sum rule can converge rather slowly, involving many bound and unbound states for a convergence of matrix elements in an atom.[4] We can see in Eq. (13.41) a bit of why this is: In the "E gauge" (Section 9.3.2), calculations typically involve such sums over squared position matrix elements (where the corresponding sums in the "A gauge" involve sums over the squared momentum matrix elements as in Eq. (13.42). The E-gauge summations typically converge more quickly, since the dipole matrix elements to high-energy states typically drop off quickly (high-energy states tend to be far away from the nucleus, unlike low-energy states). However, the sum (13.41) is also weighted by the transition frequency, which again gives weight to high-energy states, and slowing the convergence of the sum.

13.5.2 Simplification

In Eq. (13.33), we have the derivative of $f(z)$. The derivatives of the auxiliary functions are given by

$$\begin{aligned} \partial_z f(z) &= -g(z) \\ \partial_z^2 f(z) &= -\partial_z g(z) = \frac{1}{z} - f(z), \end{aligned} \tag{13.44}$$

so that the second derivative generates a term of the form

$$\sum_j \left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{2k_{j0}}{z^2}, \tag{13.45}$$

which vanishes due to the TRK sum rule. Thus, we can write Eq. (13.33) as

$$V_{\text{CP}} = \frac{1}{(4\pi\epsilon_0)8\pi}\sum_j \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{4k_{j0}^2}{z} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_z^2\frac{1}{z}\right)\right] f(2k_{j0}z).$$

(Casimir–Polder potential for ground-state atom near perfect mirror) (13.46)

Now defining the scaled coordinates

$$z_j' := 2k_{j0}z, \tag{13.47}$$

we can write

$$V_{\text{CP}} = \frac{1}{(4\pi\epsilon_0)\pi c^3}\sum_j \omega_{j0}^3\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{1}{z_j'} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_{z_j'}^2\frac{1}{z_j'}\right)\right] f(z_j')$$

(Casimir–Polder potential, scaled coordinates) (13.48)

as the potential shift of the ground state.

[4] See Peter W. Milonni and Joseph H. Eberly, *Lasers* (Wiley, 1988), p. 239, where the terms due to ionized states in hydrogen make up a significant part of the sum.

13.5.3 Spherical Symmetry

For an atom with a spherically symmetric ground state (S-orbital),

$$d_{j,\parallel}^2/2 - d_{j,z}^2 = \frac{1}{2}\left(|\langle \mathrm{g}|ex|\mathrm{e}_j\rangle|^2 + |\langle \mathrm{g}|ey|\mathrm{e}_j\rangle|^2\right) - |\langle \mathrm{g}|ez|\mathrm{e}_j\rangle|^2 = 0, \tag{13.49}$$

while

$$d_{j,\parallel}^2/2 + d_{j,z}^2 = \frac{1}{2}\left(|\langle \mathrm{g}|ex|\mathrm{e}_j\rangle|^2 + |\langle \mathrm{g}|ey|\mathrm{e}_j\rangle|^2\right) + |\langle \mathrm{g}|ez|\mathrm{e}_j\rangle|^2 = 2d_{j,z}^2. \tag{13.50}$$

Thus, the potential simplifies, and becomes

$$V_{\mathrm{CP}} = -\frac{1}{(4\pi\epsilon_0)4\pi}\sum_j d_{j,z}^2 \partial_z^2 \frac{1}{z} f(2k_{j0}z).$$

(Casimir–Polder potential, spherically symmetric atom) (13.51)

Still, the $f(z)$ function is unusual, and we will get some better feeling for it by examining its asymptotics and by carrying out the derivatives.

13.6 Asymptotic Behavior

At small distances, we can use $f(0) = \pi/2$ to obtain from Eq. (13.51) the form

$$V_{\mathrm{CP}} = -\frac{1}{(4\pi\epsilon_0)}\frac{1}{4z^3}\sum_j d_{j,z}^2.$$

(near-field van der Waals potential, spherically symmetric atom) (13.52)

We can see that this expression agrees with the classical result for the interaction of an induced, *static* dipole with its image. The interaction energy of two dipoles $\mathbf{d}_1$ and $\mathbf{d}_2$ is

$$V_{\mathrm{dip}} = \frac{\mathbf{d}_1\cdot\mathbf{d}_2 - 3(\hat{r}_{12}\cdot\mathbf{d}_1)(\hat{r}_{12}\cdot\mathbf{d}_2)}{(4\pi\epsilon_0)r_{12}^3}, \tag{13.53}$$

where $\mathbf{r}_{12}$ is the vector for the displacement between the two dipoles. For a dipole a distance z from the conducting surface interacting with its image, we have $r_{12} = 2z$, so that the dipole interaction energy reduces to

$$V_{\mathrm{dip}} = -\frac{1}{(4\pi\epsilon_0)}\frac{1}{16z^3}\left(d_\parallel^2 + 2d_\perp^2\right), \tag{13.54}$$

where $d_\parallel$ and $d_\perp$ are the parallel and perpendicular components of the dipole, respectively. We have also added an extra factor of 1/2, since don't want the direct interaction energy, but rather the energy required to bring the dipole from distance $+\infty$ to z from the surface (equivalently, the energy we derived amounts to a field integral over all space, but we only want the integral over the dipole fields over *half* space). This agrees with Eq. (13.52) if we interpret the dipole moments as ground-state expectation values

$$d_{\parallel,\perp}^2 \equiv \langle \mathrm{g}|d_{\parallel,\perp}^2|\mathrm{g}\rangle, \tag{13.55}$$

we find the quantum instantaneous dipole energy

$$V_{\mathrm{dip}} = -\frac{1}{(4\pi\epsilon_0)}\frac{1}{16z^3}\langle \mathrm{g}|\left(d_\parallel^2 + 2d_\perp^2\right)|\mathrm{g}\rangle. \tag{13.56}$$

If we further make the identifications

$$d_\parallel^2 = e^2(x^2+y^2), \qquad d_\perp^2 = e^2z^2, \tag{13.57}$$

so that assuming an isotropic atom,

$$\langle \mathrm{g}|x^2|\mathrm{g}\rangle = \langle \mathrm{g}|y^2|\mathrm{g}\rangle = \langle \mathrm{g}|z^2|\mathrm{g}\rangle, \tag{13.58}$$

we can write the static-dipole interaction as

$$V_{\rm dip} = -\frac{1}{(4\pi\epsilon_0)}\frac{1}{4z^3}\langle {\rm g}|e^2z^2|{\rm g}\rangle.$$

(near-field van der Waals potential, spherically symmetric atom) (13.59)

This is equivalent to Eq. (13.52) after inserting the identity summation over atomic states in the matrix element. This is the usual **van der Waals** atom–surface interaction. We can see that in the short range, this effect is largely a classical dipole–image interaction, except for the interpretation of the dipole as an expectation value. The force here is thus due to zero-point fluctuations of the atomic dipole, which induce a mean-square dipole that interacts with its image to produce the force.

Asymptotically, $f(z) \sim 1/z$ and $g(z) \sim 1/z^2$ for $|\arg z| < \pi$, so that for the spherically symmetric atom at large z,

$$V_{\rm CP} \sim -\frac{3c}{(4\pi\epsilon_0)4\pi z^4}\sum_j \frac{d_{j,z}^2}{\omega_{j0}} = -\frac{3\hbar c\alpha_0}{(4\pi\epsilon_0)8\pi z^4},$$

(long-range Casimir–Polder potential, spherically symmetric atom) (13.60)

where

$$\alpha_0 = \alpha(0) = \sum_j \frac{2d_{j,z}^2}{\hbar\omega_{j0}} \tag{13.61}$$

is the (classical or small-signal) static polarizability from Eq. (1.32), with $\omega, \gamma_j \longrightarrow 0$, and Eq. (5.213) to convert the oscillator strength to the dipole matrix element. Thus, we recover the standard Casimir–Polder result in the large-z limit.[5] Again, the effect is due to fluctuations of the atomic dipole, but now the dipole interacts with its *retarded* image, due to the long distance to the mirror and back. The retardation means that the dipole is no longer completely correlated with its image, and this is why the potential falls off more quickly (like z^{-4} instead of z^{-3}) in the far-field regime. Note, however, that in this regime, the Casimir–Polder effect is regarded in this regime as a true quantum effect of the *field*, since a semiclassical argument with retardation does not reproduce the correct potential without an extra choice of field-operator ordering.[6]

13.7 Excited-Level Shift

Now let us consider the corresponding shift of some excited atomic level $|\alpha\rangle$. Then we must modify the expression (13.9) to read

$$V_\alpha = -\sum_{j\neq\alpha}\sum_{\mathbf{k},\zeta}\frac{\omega_{\mathbf{k}}}{2\epsilon_0}\frac{|\langle\alpha|\mathbf{d}|j\rangle\cdot\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2}{(\omega_{j\alpha}+\omega_{\mathbf{k}})}, \tag{13.62}$$

When summing over a state $|j\rangle$ of *higher* energy than $|\alpha\rangle$, then the term has the same form as before, since $|\alpha\rangle$ acts effectively as a ground state. However, if $|j\rangle$ is of *lower* energy than $|\alpha\rangle$, then $|\alpha\rangle$ takes on the role of an excited state for the transition, and $\omega_{j\alpha} < 0$. In the interaction Hamiltonian, these modified terms are due to the usual energy-conserving terms $\sigma_j a^\dagger_{\mathbf{k},\zeta}$ and $\sigma_j^\dagger a_{\mathbf{k},\zeta}$, as opposed to the energy-nonconserving terms $\sigma_j a_{\mathbf{k},\zeta}$ and $\sigma_j^\dagger a^\dagger_{\mathbf{k},\zeta}$ terms that are responsible for the ground-state shift.

To avoid problems with the pole in the k integration, we note that in perturbation theory, the integral is always taken to be the Cauchy principle value (Section 14.1.4.2), so that the singularity causes no difficulty in principle. The derivation then carries through as for the ground state, with possibly negative wave numbers in the solution (13.33):

$$V_\alpha = -\frac{1}{(4\pi\epsilon_0)8\pi}\sum_j\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\left(\frac{1}{z}\partial_z^2\right) + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_z^2\frac{1}{z}\right)\right] f(2k_{j\alpha}z), \tag{13.63}$$

[5] Peter W. Milonni, *The Quantum Vacuum* (Academic Press, San Diego, 1994), p. 107.

[6] Stephen M. Barnett, Alain Aspect, and Peter W. Milonni, "On the quantum nature of the Casimir–Polder interaction," *Journal of Physics B: Atomic, Molecular, and Optical Physics* **33**, L143 (2000) (doi: 10.1088/0953-4075/33/4/106).

with

$$f(2k_{j\alpha}z) = f[\mathrm{sgn}(k_{j\alpha})2|k_{j\alpha}|z]. \tag{13.64}$$

Note that since we now have in principle a complex-valued function, we are explicitly taking the real part of the right-hand side. We can simplify this a bit by noting that $\mathrm{Si}(-z) = -\mathrm{Si}(z)$. However, $\mathrm{Ci}(z)$ is a bit more complicated: $\mathrm{Ci}(z)$ has a branch cut along the $(-z)$-axis, and it picks up an additional term when the sign of the argument changes:[7]

$$\begin{aligned} \mathrm{Ci}(-z) &= \mathrm{Ci}(z) - i\pi \qquad (0 < \arg z < \pi) \\ \mathrm{Ci}(-z) &= \mathrm{Ci}(z) + i\pi \qquad (-\pi < \arg z < 0). \end{aligned} \tag{13.65}$$

Note then that strictly speaking, we should not have a negative real argument of $f(z)$. However, we are implementing the Cauchy principle value, which consists of deforming the k integral by adding $\pm i0$ to k, and then averaging the results, so that the two possible extra terms cancel. Thus, for our purposes, we may write $\mathrm{Ci}(-z) = \mathrm{Ci}(z)$ for real z, and effectively,

$$f(-z) = -f(z) + \pi\cos z, \tag{13.66}$$

also for z real. Thus,

$$f(2k_{j\alpha}z) = \mathrm{sgn}(\omega_{j\alpha})f(2|k_{j\alpha}|z) + \Theta(\omega_{\alpha j})\pi\cos(2|k_{j\alpha}|z), \tag{13.67}$$

where $\Theta(z)$ is the Heaviside step function. That is, if the frequency is negative (for a term in the sum corresponding to a lower-energy level), the sign of the $f(z)$ function changes, and an extra term appears. Thus,

$$V_\alpha = -\sum_j \frac{\mathrm{sgn}(\omega_{j\alpha})}{(4\pi\epsilon_0)8\pi}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\left(\frac{1}{z}\partial_z^2\right) + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_z^2\frac{1}{z}\right)\right]\Big[f(2|k_{j\alpha}|z) - \Theta(\omega_{\alpha j})\pi\cos(2|k_{j\alpha}|z)\Big]. \tag{13.68}$$

Thus, extra, *oscillatory* terms are present when the level $|\alpha\rangle$ is an excited state—recall that $f(z)$ is *not* oscillatory. Redefining the scaled coordinates

$$z_j' := 2|k_{j\alpha}|z, \tag{13.69}$$

we can write the potential shift as

$$V_\alpha = -\sum_j \frac{\mathrm{sgn}(\omega_{j\alpha})|\omega_{j\alpha}|^3}{(4\pi\epsilon_0)\pi c^3}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\left(\frac{1}{z_j'}\partial_{z_j'}^2\right) + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_{z_j'}^2\frac{1}{z_j'}\right)\right]\Big[f(z_j') - \Theta(\omega_{\alpha j})\pi\cos z_j'\Big]. \tag{13.70}$$

The TRK sum rule again applies as in Eq. (13.48), and we can thus evaluate the derivatives in the above expression to obtain

$$V_\alpha = \sum_j \frac{\mathrm{sgn}(\omega_{j\alpha})|\omega_{j\alpha}|^3}{(4\pi\epsilon_0)\pi c^3}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{1}{z_j'} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left(\partial_{z_j'}^2\frac{1}{z_j'}\right)\right]\Big[f(z_j') - \Theta(\omega_{\alpha j})\pi\cos z_j'\Big]. \tag{13.71}$$

This is our final result for the level shift of *any* level due to the presence of the conducting plane. However, let's condense the notation just a bit more and write

$$V_\alpha = \sum_j \frac{\mathrm{sgn}_j|\omega_{j\alpha}|^3}{(4\pi\epsilon_0)\pi c^3}\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{1}{z_j'} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\partial_{z_j'}^2\frac{1}{z_j'}\right]\Big[f(z_j') - \Theta_j\pi\cos z_j'\Big], \tag{13.72}$$

(Casimir–Polder potential, excited level)

where $\mathrm{sgn}_j = \mathrm{sgn}(\omega_{j\alpha})$ is negative only when level α has higher energy than level j, and $\Theta_j = \Theta(\omega_{\alpha j})$ is unity in the same case and vanishing otherwise.

[7] See Abramowitz and Stegun, *op. cit.*, p. 232, Formula 5.2.20, or Gradshteyn and Ryzhik, *op. cit.*, Formula 8.233.2.

13.7.1 Classical Antenna Behavior

Suppose now that we focus only on the oscillatory terms:

$$V_\alpha^{(\mathrm{osc})} = \sum_{j<\alpha} \frac{|\omega_{j\alpha}|^3}{4\pi\epsilon_0 c^3} \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2 \right) \frac{1}{z'_j} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2 \right) \left(\partial_{z'_j}^2 \frac{1}{z'_j} \right) \right] \cos z'_j. \tag{13.73}$$

We can regard the sum here as extending over all states (even degenerate ones), in which case the formula

$$\Gamma_{jj'} = \frac{|\omega_{j'j}|^3 |\langle j|d|j'\rangle|^2}{3\pi\epsilon_0 \hbar c^3} \tag{13.74}$$

applies for the $|j'\rangle \longrightarrow |j\rangle$ decay path. Thus, we find

$$V_\alpha^{(\mathrm{osc})} = \sum_{j<\alpha} \frac{3}{4} \hbar \Gamma_{\alpha j} \left[\left(\hat{\varepsilon}_{j,\parallel}^2/2 - \hat{\varepsilon}_{j,\perp}^2 \right) - \left(\hat{\varepsilon}_{j,\parallel}^2/2 + \hat{\varepsilon}_{j,\perp}^2 \right) \partial_{z'_j}^2 \right] \frac{\cos z'_j}{z'_j}, \tag{13.75}$$

where now $\hat{\varepsilon}_\parallel$ and $\hat{\varepsilon}_\perp$ are the projections of the dipole unit vector onto the components parallel and perpendicular to the surface, respectively. By comparison to the classical expression of a dipole near a mirror, from Eq. (1.130),

$$\delta\omega_0 = \frac{3}{4}\gamma \left[\left(\hat{\varepsilon}_\parallel^2/2 - \hat{\varepsilon}_\perp^2 \right) - \left(\hat{\varepsilon}_\parallel^2/2 + \hat{\varepsilon}_\perp^2 \right) \partial_{z'}^2 \right] \frac{\cos z'}{z'}, \tag{13.76}$$

where $z' := 2k_0 z$, we see that the oscillatory part of the shift is explained by the classical model (at least for a single transition, for here we must sum over all lower energy levels). Thus, we can interpret the part of the potential unique to excited states as a classical dipole potential due to the spontaneously radiated field. However, what the classical model misses is the $f(z)$ part, which is a manifestation of the quantum vacuum. As we noted above, the ground-state Casimir–Polder potential is an effect of reflecting virtual photons, and we now see that the excited-state potential also includes effects due to *real* photons bouncing from the mirror.

To examine this more quantitatively, consider a two-level atom,[8] in which case the *excited-state shift* corresponding to Eq. (13.72) is

$$\frac{\delta E_\mathrm{e}}{\hbar} = \frac{3\Gamma}{4\pi} \left[\left(\hat{\varepsilon}_\parallel^2/2 - \hat{\varepsilon}_\perp^2 \right) - \left(\hat{\varepsilon}_\parallel^2/2 + \hat{\varepsilon}_\perp^2 \right) \partial_{z'}^2 \right] \frac{\pi \cos z' - f(z')}{z'}. \tag{13.77}$$

For large z', $f(z') \ll \pi \cos z'$, and we recover the classical result (13.76). For small z', $\pi \cos z' - f(z') = \pi/2 + O(z')$, and thus the quantum near-field (van der Waals) shift is *half* the classical value. If we choose to identify the classical and quantum mean-square dipoles differently to make them agree in the near-field, then the quantum shift will be double the classical shift in the far field.[9] Thus, because of the vacuum contribution, there is no simple way to exactly identify the shift of the atomic excited state of the two-level atom with that of the classical dipole radiator.

[8] Note that the two-level atom has some artifacts due to different cancellations than in the full summation for a real atom. For example, if we calculate the *transition frequency shift* to compare to the classical case, we get the combination $\pi \cos z' - 2f(z')$, which vanishes to zeroth order. Thus, the $1/z^3$ leading-order contribution vanishes, which is not the case for real atoms, because the coupling to other levels still generate this term. See G. Barton, *op. cit.*

[9] E. A. Hinds and V. Sandoghdar, "Cavity QED level shifts of simple atoms," *Physical Review A* **43**, 398 (1991) (doi: 10.1103/PhysRevA.43.398). See also S. Haroche, *op. cit.*

13.8 Power–Zienau Transformation in Half-Space

One little detail that we have ignored is that the *complete* dipole interaction Hamiltonian that we derived before in Eq. (9.38) has the form

$$\begin{aligned} H_{\rm AF} &= \frac{1}{2}\int d^3r\,\rho(\mathbf{r})\phi(\mathbf{r}) + e\mathbf{r}_{\rm e}\cdot\mathbf{E}^{\perp}(0) + \frac{1}{2\epsilon_0}\int d^3r\,P^{\perp 2}(\mathbf{r}) \\ &= \frac{1}{2}\int d^3r\,\rho(\mathbf{r})\phi(\mathbf{r}) - \mathbf{d}\cdot\mathbf{E}^{\perp}(0) + \frac{1}{2\epsilon_0}\int d^3r\,P^{\perp 2}(\mathbf{r}) \\ &= \tilde{H}_{\rm Coulomb} + \tilde{H}_{\mathbf{d}\cdot\mathbf{E}} + \tilde{H}_{\rm self}, \end{aligned} \tag{13.78}$$

where the last term is a polarization energy due to the dipole, with

$$P^{\perp}_{\alpha}(\mathbf{r}) := -er_{{\rm e},\beta}\delta^{\perp}_{\alpha\beta}(\mathbf{r}). \tag{13.79}$$

Note that we are also now including the interaction of the electron charge density $\rho(\mathbf{r})$ with the *scalar* potential, because now we have a coupling to a *longitudinal* electric field: an instantaneous dipole moment of the atom shows up as an instantaneous image dipole, to enforce the boundary conditions at the conducting plane. The $\mathbf{r}\cdot\mathbf{E}$ term does not account for this, as it was derived from the coupling to the vector potential and thus includes only the coupling to the *transverse* field. We have neglected the first and last terms thus far, although we have gotten the correct interaction, which yields the instantaneous dipole-dipole coupling at short range and the retarded scaling at long range. We have already shown in Eq. (13.56) that the instantaneous dipole-image interaction due to $H_{\rm Coulomb}$ is given by

$$\Delta E_{\rm Coulomb} = -\frac{1}{(4\pi\epsilon_0)}\frac{1}{16z^3}\langle\alpha|\left(d_{\parallel}^2 + 2d_{\perp}^2\right)|\alpha\rangle. \tag{13.80}$$

We have neglected the last term in Eq. (13.78), the dipole self-energy. This easy to justify, since the self-energy is integrated over all space and thus independent of $\mathbf{r}$. The contribution is thus z-independent and disappears in the renormalization of the Casimir–Polder potential. Note that this term *is* important in computing the Lamb shift, as we will show later.

Of course, the above interaction Hamiltonian was derived using a field commutator, specifically a *free-space* field commutator. However, for this calculation we have been living in *half* space, and we should verify that our interaction Hamiltonian is still appropriate. From our previous treatment of the Power–Zienau transformation, specifically Eq. (9.33), the transformed electric field operator is given by

$$\begin{aligned} \tilde{E}^{\perp}_{\beta}(\mathbf{r}) &= U(\mathbf{r}_{\rm e})E^{\perp}_{\beta}(\mathbf{r})U^{\dagger}(\mathbf{r}_{\rm e}) \\ &= E^{\perp}_{\beta}(\mathbf{r}) + \frac{ie}{\hbar}r_{{\rm e},\alpha}\Big[A_{\alpha}(z\hat{z}), E^{\perp}_{\beta}(\mathbf{r})\Big], \end{aligned} \tag{13.81}$$

if we regard the atomic position to be $z\hat{z}$, with the conductor located at $z = 0$ (the transverse location is arbitrary). We now use the commutator [from Eq. (8.215)]

$$[A_{\alpha}(\mathbf{r},t), E_{\beta}(\mathbf{r}',t)] = -\frac{i\hbar}{\epsilon_0}\left[\delta^{\perp}_{\alpha\beta}(\mathbf{r}-\mathbf{r}') - \delta^{\top}_{\alpha\beta}(\mathbf{r}^{-}-\mathbf{r}')\right], \tag{13.82}$$

where the transverse and "reflected transverse" delta functions are [from Eqs. (8.178) and (8.214), respectively]

$$\begin{aligned} \delta^{\perp}_{\alpha\beta}(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_{\alpha}k_{\beta}}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}} \\ \delta^{\top}_{\alpha\beta}(\mathbf{r}) &= \frac{1}{(2\pi)^3}\int d^3k\,\left(\delta^{-}_{\alpha\beta} - \frac{k^{-}_{\alpha}k_{\beta}}{k^2}\right)e^{i\mathbf{k}\cdot\mathbf{r}}. \end{aligned} \tag{13.83}$$

(Recall that $k_\alpha^- = k_\alpha$ except $k_z^- = -k_z$, and $\delta_{\alpha\beta}^- = \delta_{\alpha\beta}$ except that $\delta_{zz}^- = -1$.) We can thus write the transformed field in half-space as

$$\tilde{E}_\beta^\perp(\mathbf{r}) = E_\beta^\perp(\mathbf{r}) + \frac{e}{\epsilon_0} r_{\mathrm{e},\alpha} \left[\delta_{\alpha\beta}^\perp(\mathbf{r} - z\hat{z}) - \delta_{\alpha\beta}^\top(\mathbf{r} + z\hat{z})\right]. \tag{13.84}$$

We now see that the transformed field has an extra contribution at the location of the image dipole. The field energy thus transforms as

$$\begin{aligned} U(\mathbf{r}_\mathrm{e}) \left[\int_{z>0} d^3r \left[E^\perp(\mathbf{r})\right]^2\right] U^\dagger(\mathbf{r}_\mathrm{e}) &= \int_{z>0} d^3r \left[U(\mathbf{r}_\mathrm{e})\mathbf{E}^\perp(\mathbf{r})U^\dagger(\mathbf{r}_\mathrm{e})\right]^2 \\ &= \int_{z>0} d^3r \left[\mathbf{E}^\perp(\mathbf{r})\right]^2 + \frac{2e}{\epsilon_0}\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp(z\hat{z}) + \frac{2e}{\epsilon_0}\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\top(-z\hat{z}) \\ &\quad + \frac{1}{\epsilon_0^2}\int_{z>0} d^3r \left[\mathbf{P}^\perp(\mathbf{r} - z\hat{z})\right]^2 + \frac{1}{\epsilon_0^2}\int_{z>0} d^3r \left[\mathbf{P}^\top(\mathbf{r} + z\hat{z})\right]^2 \\ &\quad + \frac{2}{\epsilon_0^2}\int d^3r\, \mathbf{P}^\perp(\mathbf{r} - z\hat{z})\cdot\mathbf{P}^\top(\mathbf{r} + z\hat{z}). \end{aligned} \tag{13.85}$$

We have defined here the reflected field

$$E_\alpha^\top(\mathbf{r}) := \frac{1}{2}\int_{z>0} d^3r' \left[\delta_{\alpha\beta}^\top(\mathbf{r}' - \mathbf{r}) + \delta_{\beta\alpha}^\top(\mathbf{r}' - \mathbf{r})\right] E_\beta^\perp(\mathbf{r}'), \tag{13.86}$$

which for our purposes here is the same as the usual transverse field, but with the opposite sign for the z-component, and the reflected polarization

$$P_\alpha^\top(\mathbf{r}) := -e r_{\mathrm{e},\beta}\delta_{\alpha\beta}^\top(\mathbf{r}). \tag{13.87}$$

The interaction Hamiltonian then becomes

$$\begin{aligned} \tilde{H}_{\mathrm{AF}} &= \frac{1}{2}\int d^3r\, \rho(\mathbf{r})\phi(\mathbf{r}) + e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp(z\hat{z}) + e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\top(-z\hat{z}) \\ &\quad + \frac{1}{2\epsilon_0}\int_{z>0} d^3r \left[\mathbf{P}^\perp(\mathbf{r} + z\hat{z})\right]^2 + \frac{1}{2\epsilon_0}\int_{z>0} d^3r \left[\mathbf{P}^\perp(\mathbf{r} - z\hat{z})\right]^2 + \frac{1}{\epsilon_0}\int_{z>0} d^3r\, \mathbf{P}^\perp(\mathbf{r} - z\hat{z})\cdot\mathbf{P}^\top(\mathbf{r} + z\hat{z}). \end{aligned} \tag{13.88}$$

Thus we see that in half-space, there is an additional dipole-field interaction term and two additional self-energy terms. However, the field operator $\mathbf{E}^\top(-z\hat{z})$ vanishes, since it refers to the field amplitude behind the conductor, where it vanishes. The other polarization terms reduce very simply, using the properties

$$[\mathbf{P}^\top(\mathbf{r})]^2 = [\mathbf{P}^\perp(\mathbf{r}^-)]^2 \tag{13.89}$$

and

$$\mathbf{P}^\perp(\mathbf{r})\cdot\mathbf{P}^\top(\mathbf{r}') = \mathbf{P}^\top(\mathbf{r})\cdot\mathbf{P}^\perp(\mathbf{r}'). \tag{13.90}$$

In this case, the interaction Hamiltonian becomes

$$\begin{aligned} \tilde{H}_{\mathrm{AF}} &= \frac{1}{2}\int d^3r\, \rho(\mathbf{r})\phi(\mathbf{r}) + e\mathbf{r}_\mathrm{e}\cdot\mathbf{E}^\perp(z) + \frac{1}{2\epsilon_0}\int_{\text{all } z} d^3r \left[\mathbf{P}^\perp(\mathbf{r})\right]^2 + \frac{1}{2\epsilon_0}\int_{\text{all } z} d^3r\, \mathbf{P}^\perp(\mathbf{r} - z\hat{z})\cdot\mathbf{P}^\top(\mathbf{r} + z\hat{z}) \\ &= \tilde{H}_{\mathrm{Coulomb}} + \tilde{H}_{\mathbf{d}\cdot\mathbf{E}} + \tilde{H}_{\mathrm{self}} + \tilde{H}_{\text{dipole-image}}. \end{aligned} \tag{13.91}$$

Thus, we recover the usual dipole self-energy term, plus a second that is evidently due to the interaction of the dipole with its image in the mirror. The usual term is z-independent and does not contribute to the Casimir–Polder potential after renormalization. We will examine this term more closely in Section 13.12.2.1. However, the new term *does*, and we will now evaluate it.

To evaluate the dipole-image part of the Hamiltonian,

$$\tilde{H}_{\text{dipole-image}} = \frac{1}{2\epsilon_0}\int_{\text{all } z} d^3r\, \mathbf{P}^{\perp}(\mathbf{r}-z\hat{z})\cdot\mathbf{P}^{\perp}(\mathbf{r}+z\hat{z}), \tag{13.92}$$

we first evaluate the integral

$$\begin{aligned}
(I_3)_{\alpha\gamma} &= \int d^3r\, \delta^{\perp}_{\alpha\beta}(\mathbf{r}-z\hat{z})\,\delta^{\top}_{\gamma\beta}(\mathbf{r}+z\hat{z}) \\
&= \frac{1}{(2\pi)^6}\int d^3r \int d^3k \int d^3k' \left(\delta_{\alpha\beta}-\frac{k'_\alpha k'_\beta}{k'^2}\right)\left(\delta^{-}_{\gamma\beta}-\frac{k^{-}_\gamma k_\beta}{k^2}\right) e^{i\mathbf{k}\cdot(\mathbf{r}-z\hat{z})}e^{i\mathbf{k}'\cdot(\mathbf{r}+z\hat{z})} \\
&= \frac{1}{(2\pi)^3}\int d^3k \int d^3k' \left(\delta_{\alpha\beta}-\frac{k'_\alpha k'_\beta}{k'^2}\right)\left(\delta^{-}_{\gamma\beta}-\frac{k^{-}_\gamma k_\beta}{k^2}\right)\delta(\mathbf{k}+\mathbf{k}')\, e^{-i(k_z-k'_z)z} \\
&= \frac{1}{(2\pi)^3}\int d^3k \left(\delta_{\alpha\beta}-\frac{k_\alpha k_\beta}{k^2}\right)\left(\delta^{-}_{\gamma\beta}-\frac{k^{-}_\gamma k_\beta}{k^2}\right) e^{-i2k_z z} \\
&= \frac{1}{(2\pi)^3}\int d^3k \left(\delta^{-}_{\alpha\gamma}-\frac{k_\alpha k^{-}_\gamma}{k^2}\right) e^{-i2k_z z}.
\end{aligned} \tag{13.93}$$

Note that the z coordinate here is the location of the atom, *not* a component of the integration variable $\mathbf{r}$. Now we can evaluate the components on a case-by-case basis. Note that the integrand is axially symmetric, except for the tensor part, and thus the integral vanishes if $\alpha \neq \gamma$, as in this case the dependence on the axial angle ϕ will be $\sin\theta$, $\cos\theta$, or $\sin\theta\cos\theta$. We can also see that the $\delta^{-}_{\alpha\gamma}$ *never* contributes, in view of the integral

$$\begin{aligned}
\int d^3k\, e^{-i2k_z z} &= \int d^3k\, \cos(2kz\cos\theta) \\
&= 2\pi\int_0^\infty dk\, k^2 \int_0^\pi d\theta\, \sin\theta\cos(2kz\cos\theta) \\
&= \frac{2\pi}{z}\int_0^\infty dk\, k\sin(2kz) \\
&= \lim_{\sigma\to 0}\frac{2\pi}{z}\int_0^\infty dk\, k\sin(2kz)e^{-k\sigma} \\
&= \lim_{\sigma\to 0}\frac{8\pi\sigma}{(\sigma^2+4z^2)^2} \\
&= 0.
\end{aligned} \tag{13.94}$$

If $\alpha=\gamma=z$, then we have

$$\begin{aligned}
(I_3)_{zz} &= \frac{1}{(2\pi)^3}\int d^3k\, \frac{k_z^2}{k^2}e^{-i2k_z z} \\
&= \frac{1}{(2\pi)^3}\left(-\frac{1}{4}\partial_z^2\right)\int d^3k \frac{1}{k^2}\cos(2k_z z) \\
&= \frac{1}{(2\pi)^2}\left(-\frac{1}{4}\partial_z^2\right)\int_0^\infty dk \int_0^\pi d\theta\, \sin\theta\cos(2kz\cos\theta) \\
&= \frac{1}{(2\pi)^2}\left(-\frac{1}{4}\partial_z^2\right)\int_0^\infty dk\, \frac{\sin(2kz)}{kz} \\
&= \frac{1}{(2\pi)^2}\left(-\frac{1}{4}\partial_z^2\right)\frac{\pi}{2z} \\
&= -\frac{1}{16\pi z^3}.
\end{aligned} \tag{13.95}$$

On the other hand, if $\alpha=\gamma=x$ or y, then

$$(I_3)_{xx} = (I_3)_{yy} = -\frac{1}{(2\pi)^3}\int d^3k\, \frac{k_\parallel^2}{2k^2}e^{-i2k_z z} = -\frac{1}{32\pi z^3}, \tag{13.96}$$

where we have used the fact that the x- and y-directions are equivalent in this axisymmetric problem, and $k_\parallel^2 = k_x^2 + k_y^2 = k^2 - k_z^2$. Thus,

$$(I_3)_{\alpha\beta} = -\frac{1}{32\pi z^3}\left(\delta_{\alpha x}\delta_{\beta x} + \delta_{\alpha y}\delta_{\beta y} + 2\delta_{\alpha z}\delta_{\beta z}\right). \tag{13.97}$$

To second order in the atomic dipole moment (i.e., to order e^2), it is sufficient to compute the shift due to the dipole-image Hamiltonian to first order in perturbation theory. Thus, for the shift of level α,

$$\begin{aligned}\Delta E_{\text{dipole-image},\alpha} &= \langle\alpha|\tilde{H}_{\text{dipole-image}}|\alpha\rangle \\ &= \frac{e^2}{2\epsilon_0}\langle\alpha|r_{\text{e},\mu}r_{\text{e},\nu}|\alpha\rangle(I_3)_{\mu\nu} \\ &= \frac{1}{(4\pi\epsilon_0)16z^3}\langle\alpha|\left[d_{j,\parallel}^2 + 2d_{j,z}^2\right]|\alpha\rangle.\end{aligned} \tag{13.98}$$

This is the opposite of the static-dipole energy shift due to the Coulomb Hamiltonian.

Note that this result—where contributions boundary terms from the Power–Zienau transformation cancel the static image energies—*always* holds, independent of the shape of the boundary. Thus, when using the dipole Hamiltonian, it is sufficient to use the *free-space* version without worrying about Coulomb interactions with images.[10]

13.9 Calculation in the Coulomb Gauge

Now we will show that the same Casimir–Polder potential obtains if we use the A-gauge form

$$H_{\text{AF}}^{(A)} = \frac{e}{m_{\text{e}}}\mathbf{p}_{\text{e}}\cdot\mathbf{A} + \frac{e^2}{2m_{\text{e}}}A^2 \tag{13.99}$$

of the interaction Hamiltonian from Eq. (9.49), including the sometimes-neglected A^2 term. The ground-state shift of the first term follows from adapting Eq. (13.32) to read

$$\Delta E_{\mathbf{p}\cdot\mathbf{A}} = \frac{1}{16\pi^3\epsilon_0}\sum_j\left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)I'_{1j} + \left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)I'_{2j}\right], \tag{13.100}$$

where the above form is the same as before, but written in terms of new integrals

$$\begin{aligned}I'_1 &= k_0^2\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\int_{-\infty}^{\infty}dk_z\,\frac{1}{k(k+k_0)}\cos(2k_z z) \\ I'_2 &= k_0^2\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\int_{-\infty}^{\infty}dk_z\,\frac{k_z^2}{k^3(k+k_0)}\cos(2k_z z).\end{aligned} \tag{13.101}$$

This is because the matrix elements of the $\mathbf{r}\cdot\mathbf{E}$ and $\mathbf{p}\cdot\mathbf{A}$ interaction Hamiltonians differ in magnitude by a factor of ω/ω_{j0}, as we showed in Section 9.3.2. Thus, the integrands are multiplied by factors of $(k_{j0}/k)^2$ compared to the previous calculation in the E gauge. Evaluating the I'_1 integral,

$$\begin{aligned}I'_1 &= 2\pi k_0^2\int_0^{\infty}dk\int_0^{\pi}d\theta\,\frac{k\sin\theta\cos(2kz\cos\theta)}{(k+k_0)} \\ &= 2\pi k_0^2\int_0^{\infty}dk\,\frac{k}{(k+k_0)}\frac{\sin(2kz)}{kz} \\ &= \frac{2\pi k_0^2}{z}\int_0^{\infty}dk\,\frac{\sin(2kz)}{(k+k_0)} \\ &= \frac{2\pi k_0^2}{z}f(2k_0 z).\end{aligned} \tag{13.102}$$

[10] E. A. Power and T. Thirunamachandran, "Quantum electrodynamics in a cavity," *Physical Review A* **25**, 2473 (1982) (doi: 10.1103/PhysRevA.25.2473).

Recalling that

$$I_1 = -\frac{\pi}{2}\left(\frac{1}{z}\partial_z^2\right) f(2k_0 z) = \frac{2\pi k_0^2}{z} f(2k_0 z) - \frac{\pi k_0}{z^2}, \tag{13.103}$$

we showed above that the second term vanished in the sum over levels by the TRK sum rule (13.43). Thus, the two integrals give equivalent results for the asymmetric-dipole part of the Casimir–Polder potential.

To evaluate the second integral I_2', the procedure is similar to the case for I_2:

$$\begin{aligned}
I_2' &= 2\pi k_0^2 \int_0^\infty dk \int_0^\pi d\theta \, \frac{k \sin\theta \cos^2\theta \cos(2kz\cos\theta)}{(k+k_0)} \\
&= 4\pi \int_0^\infty dk \, \frac{k}{(k+k_0)} \left(\frac{\sin(2kz)}{2kz} + 2\frac{\cos(2kz)}{(2kz)^2} - 2\frac{\sin(2kz)}{(2kz)^3} \right) \\
&= -\pi k_0^2 \left(\frac{1}{2z}\partial_z^2 - \frac{1}{z^2}\partial_z + \frac{1}{z^3} \right) \int_0^\infty dk \, \frac{\sin(2kz)}{k^2(k+k_0)} \\
&= -\frac{\pi}{2} k_0^2 \left(\partial_z^2 \frac{1}{z} \right) \int_0^\infty dk \, \frac{\sin(2kz)}{k^2(k+k_0)} \\
&= -\frac{\pi}{2} k_0^2 \left(\partial_z^2 \frac{1}{z} \right) \left(-4\partial_z^{-2} \right) \int_0^\infty dk \, \frac{\sin(2kz)}{(k+k_0)} \\
&= -\frac{\pi}{2} k_0^2 \left(\partial_z^2 \frac{1}{z} \right) \left(-4\partial_z^{-2} \right) f(2k_0 z) \\
&= -\frac{\pi}{2} \left(\partial_z^2 \frac{1}{z} \right) f(2k_0 z) + \frac{\pi^2}{4} \left(\partial_z^2 \frac{1}{z} \right) + \pi k_0 \left(\partial_z^2 \frac{1}{z} \right) [z \log(2k_0 z) - z] \\
&= -\frac{\pi}{2} \left(\partial_z^2 \frac{1}{z} \right) f(2k_0 z) + \frac{\pi^2}{2z^3} - \frac{\pi k_0}{z^2},
\end{aligned} \tag{13.104}$$

where we used the antiderivative formula

$$\partial_z^{-2} f(z) = -f(z) + \pi/2 + z\log z - z. \tag{13.105}$$

The $\pi/2$ is one of the constants of integration from evaluating the antiderivatives, while the other gives a term of the form cz, which vanishes under the derivative. These constants are set by noting that the integral in the fourth line above vanishes for $z = 0$, and that $f(0) = \pi/2$. Now recalling from Eq. (13.30) that

$$I_2 = -\frac{\pi}{2} \left(\partial_z^2 \frac{1}{z} \right) f(2k_0 z), \tag{13.106}$$

we see that there are *two* terms in I_2' that do not appear in I_2 that we must explain. The second term in I_2' leads to an energy shift of the form

$$\frac{1}{32\pi\epsilon_0 z^3} \sum_j \left(d_{j,\parallel}^2/2 + d_{j,z}^2 \right) = \frac{1}{4\pi\epsilon_0} \frac{1}{16 z^3} \langle \mathrm{g} | \left(d_\parallel^2 + d_z^2 \right) | \mathrm{g} \rangle. \tag{13.107}$$

This has the same form, except for the opposite sign, as the static dipole energy (13.54), if we interpret the classical squared dipole moments as quantum expectation values. Thus, this term *cancels* the static Coulomb energy of the instantaneous dipole moment interacting with the boundary, which we have so far neglected to include, given by the Hamiltonian

$$H_{\mathrm{Coulomb}} = \frac{1}{2} \int d^3 r \rho(\mathbf{r}) \phi(\mathbf{r}), \tag{13.108}$$

where $\rho(\mathbf{r})$ is the charge density corresponding to the atomic dipole, and $\phi(\mathbf{r})$ is the scalar potential, which gives the energy contribution of the *longitudinal* field. This energy corresponds to an *unretarded* energy, and clearly must be canceled to produce the correct quantum (retarded) shift.

The third term in I_2', which also does not appear in I_1, also scales as k_0 and therefore, by the TRK sum rule (13.43), becomes independent of the level:

$$\frac{k_{j0}}{16\pi^2\epsilon_0 z^2}\sum_j\left(d_{j,\parallel}^2/2+d_{j,z}^2\right)=-\frac{e^2\hbar}{16\pi^2\epsilon_0 m_{\rm e}cz^2}. \tag{13.109}$$

We will show below that this term exactly cancels the contribution from the A^2 term of the A-gauge interaction Hamiltonian.

Thus, we see that the A-gauge interaction Hamiltonian gives the same result as the total E-gauge dipole interaction Hamiltonian. The same is true of the excited-level shifts, because the extra $\cos(2k_0z)$ terms that appear there do not generate extra terms when differentiated or integrated twice, and the extra terms we generated turned out to be level-independent. Thus, the same branch-cut argument above produces the same terms in the A gauge. However we see explicitly here the importance of summing over all the excited levels; had we made a two-level atom approximation, we would have gotten different results in the two gauges.[11]

Then what about the rest of the A-gauge Hamiltonian? We still have left the field self-energy part

$$H_{\rm AF,2}=\frac{e^2}{2m_{\rm e}}A^2 \tag{13.110}$$

of the interaction. Since our calculation is valid to lowest (second) nonvanishing order in the dipole matrix element, we can compute the shift due to this Hamiltonian to order e^2, and thus it suffices to compute the shift to *first* order in perturbation theory:

$$\begin{aligned}\Delta E_{A^2}&=\langle {\rm g}|H_{\rm AF,2}|{\rm g}\rangle\\&=\frac{e^2}{2m_{\rm e}}\langle {\rm g}|A^2|{\rm g}\rangle.\end{aligned} \tag{13.111}$$

Using the expression

$$\mathbf{A}(\mathbf{r},t)=\sum_{\mathbf{k},\zeta}i\sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}\epsilon_0}}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})a_{\mathbf{k},\zeta}(t)+{\rm H.c.} \tag{13.112}$$

for the quantum vector potential in terms of the mode functions from Eq. (8.61), we find that only the terms of the form $a_{{\rm k},\zeta}a_{{\rm k},\zeta}^\dagger$, where both operators correspond to the *same* mode, contribute to the vacuum expectation value above. Thus,

$$\Delta E_{A^2}=\frac{e^2\hbar}{4m_{\rm e}\epsilon_0}\sum_{\mathbf{k},\zeta}\frac{1}{\omega_{\mathbf{k}}}|\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2. \tag{13.113}$$

Using the half-space mode functions in Eq. (13.3), the shift then becomes

$$\begin{aligned}\Delta E_{A^2}&=\frac{e^2\hbar}{2m_{\rm e}\epsilon_0 V}\sum_{\mathbf{k},\zeta}\frac{1}{\omega_{\mathbf{k}}}\left[|\hat{\varepsilon}_{\mathbf{k},\zeta,\parallel}|^2\sin^2 k_zz+|\hat{\varepsilon}_{\mathbf{k},\zeta,z}|^2\cos^2 k_zz\right]\\&=\frac{e^2\hbar}{2m_{\rm e}\epsilon_0 V}\sum_{\mathbf{k}}\frac{1}{\omega_{\mathbf{k}}}\left[\left(1+\frac{k_z^2}{k^2}\right)\sin^2 k_zz+\left(1-\frac{k_z^2}{k^2}\right)\cos^2 k_zz\right]\\&=\frac{e^2\hbar}{2m_{\rm e}\epsilon_0 V}\sum_{\mathbf{k}}\frac{1}{\omega_{\mathbf{k}}}\left[1+\frac{k_z^2}{k^2}\left(\sin^2 k_zz-\cos^2 k_zz\right)\right]\\&=\frac{e^2\hbar}{2m_{\rm e}\epsilon_0 V}\sum_{\mathbf{k}}\frac{1}{\omega_{\mathbf{k}}}\left[1-\frac{k_z^2}{k^2}\cos(2k_zz)\right].\end{aligned} \tag{13.114}$$

As usual, we can change this to an integral over all reciprocal space in the large-volume limit:

$$\Delta E_{A^2}=\frac{e^2\hbar}{8\pi^2 m_{\rm e}\epsilon_0 c}\int_0^\pi d\theta\,\sin\theta\int_0^\infty dk\,k\left[1-\frac{k_z^2}{k^2}\cos(2k_zz)\right]. \tag{13.115}$$

[11] This point was made eloquently by G. Barton, *op. cit.*

The first term in the integral is z-independent, and gives a correction to the Lamb shift:

$$\Delta E_{A^2,\text{Lamb}} = \frac{e^2\hbar}{4\pi^2 m_\text{e}\epsilon_0 c}\int_0^\infty dk\, k. \tag{13.116}$$

Renormalization thus removes this part from the Casimir–Polder shift. The remainder of the energy shift is, using the same integration procedures as before,

$$\begin{aligned}
\Delta E_{A^2} &= -\frac{e^2\hbar}{8\pi^2 m_\text{e}\epsilon_0 c}\int_0^\infty dk \int_0^\pi d\theta\, \sin\theta \frac{k_z^2}{k}\cos(2k_z z) \\
&= -\frac{e^2\hbar}{8\pi^2 m_\text{e}\epsilon_0 c}\int_0^\infty dk\, k \int_0^\pi d\theta\, \sin\theta\cos^2\theta\cos(2kz\cos\theta) \\
&= -\frac{e^2\hbar}{8\pi^2 m_\text{e}\epsilon_0 c}\left(-\frac{1}{4}\partial_z^2\frac{1}{z}\right)\int_0^\infty dk\,\frac{\sin(2kz)}{k^2} \\
&= -\frac{e^2\hbar}{8\pi^2 m_\text{e}\epsilon_0 c}\left(\partial_z^2\frac{1}{z}\partial_z^{-2}\right)\int_0^\infty dk\,\sin(2kz) \\
&= -\frac{e^2\hbar}{16\pi^2 m_\text{e}\epsilon_0 c}\left(\partial_z^2\frac{1}{z}\partial_z^{-2}\right)\frac{1}{z} \\
&= -\frac{e^2\hbar}{16\pi^2 m_\text{e}\epsilon_0 c}\left(\partial_z^2\frac{1}{z}\right)(z\log z - z) \\
&= -\frac{e^2\hbar}{16\pi^2 m_\text{e}\epsilon_0 c}\partial_z^2\log z \\
&= -\frac{e^2\hbar}{16\pi^2 m_\text{e}\epsilon_0 c z^2} = -\frac{\hbar^2\alpha_0}{4\pi m_\text{e} z^2},
\end{aligned} \tag{13.117}$$

where $\alpha_0 := e^2/4\pi\epsilon_0\hbar c$ is the fine-structure constant. There is here no constant of integration from evaluating the antiderivative, since the integral in the third step above vanishes for $z = 0$, and $\lim_{z\to\infty}(z\log z - z) = 0$; the other constant of integration contributes a linear term that vanishes subsequently under the derivative. Note that this is exactly the *opposite* of the extra term that we found in Eq. (13.109), and thus this contribution cancels that one. Again, this term is level-independent, and thus this cancellation occurs for *any* atomic level, not just the ground state.

13.10 Evaluation

If we are to evaluate the Casimir–Polder potential, it is useful to expand out the rather compact form of Eq. (13.72). We can do this by noting

$$\partial_{z_j'}^2\frac{1}{z_j'}\Big[f(z_j') - \Theta_j\pi\cos z_j'\Big] = \frac{1}{z_j'^2} + \left(\frac{2}{z_j'^3} - \frac{1}{z_j'}\right)\Big[f(z_j') - \Theta_j\pi\cos z_j'\Big] + \frac{2}{z_j'^2}\Big[g(z_j') - \Theta_j\pi\sin z_j'\Big], \tag{13.118}$$

and thus

$$\begin{aligned}
V_\alpha = &\sum_j \frac{\text{sgn}_j|\omega_{j\alpha}|^3}{(4\pi\epsilon_0)\pi c^3}\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right)\frac{1}{z_j'}\Big[f(z_j') - \Theta_j\pi\cos z_j'\Big] \\
&-\sum_j \frac{\text{sgn}_j|\omega_{j\alpha}|^3}{(4\pi\epsilon_0)\pi c^3}\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\left[\frac{1}{z_j'^2} + \left(\frac{2}{z_j'^3} - \frac{1}{z_j'}\right)\Big[f(z_j') - \Theta_j\pi\cos z_j'\Big] + \frac{2}{z_j'^2}\Big[g(z_j') - \Theta_j\pi\sin z_j'\Big]\right],
\end{aligned}$$

(general Casimir–Polder potential, expanded derivatives) (13.119)

we can then evaluate this expression computationally by summing over all the states, evaluating the auxiliary functions in terms of the sine and cosine integrals.

13.11 Numerical Evaluation: ^{87}Rb

The dipole moments defined above must be written in terms of the appropriate dipole matrix element connecting two hyperfine levels $|F, m_F\rangle$ and $|F', m'_F\rangle$ (primes referring to the excited level). The hyperfine dipole moment can be factored in terms of a Wigner 3-j symbol,

$$\langle F\ m_F|er_q|F'\ m'_F\rangle = \langle F\|e\mathbf{r}\|F'\rangle(-1)^{F'-1+m_F}\sqrt{2F+1}\begin{pmatrix} F' & 1 & F \\ m'_F & q & -m_F \end{pmatrix}, \tag{13.120}$$

where I is the nuclear spin and J is the composite electron spin. This can be further factored in terms of a 6-j symbol and a reduced matrix element for the fine-structure transition:

$$\begin{aligned}\langle F\|e\mathbf{r}\|F'\rangle &\equiv \langle J\ I\ F\|e\mathbf{r}\|J'\ I'\ F'\rangle \\ &= \langle J\|e\mathbf{r}\|J'\rangle(-1)^{F'+J+1+I}\sqrt{(2F'+1)(2J+1)}\begin{Bmatrix} J & J' & 1 \\ F' & F & I \end{Bmatrix}.\end{aligned} \tag{13.121}$$

This form is particularly convenient, as this dipole matrix element may be written in terms of the partial lifetime for the decay path $J' \longrightarrow J$:

$$\frac{1}{\tau_{J'J}} = \frac{\omega_0^3}{3\pi\epsilon_0\hbar c^3}\frac{2J+1}{2J'+1}|\langle J\|e\mathbf{r}\|J'\rangle|^2. \tag{13.122}$$

Experimental measurements usually consider only the total lifetime of a level, given by summing over all possible decay paths

$$\frac{1}{\tau_{J'}} = \sum_J \frac{1}{\tau_{J'J}}. \tag{13.123}$$

However, using the partial lifetimes avoids confusion with myriad normalization conventions for dipole matrix elements, oscillator strengths, etc.

We are only considering broadband "light," so to good approximation we do not need to explicitly consider hyperfine splittings. Thus, it is convenient to simply sum over the dipole matrix elements originating from a particular hyperfine state $|F, m_F\rangle$. For the ground and D_1 ($5P_{1/2}$) states, we can evaluate this sum with the above formulae, and it turns out to be particularly simple:

$$\sum_{F'}|\langle F, m_F|er_0|F', m_F\rangle|^2 = \frac{1}{3}|\langle J\|e\mathbf{r}\|J'\rangle|^2. \tag{13.124}$$

For the D_2 excited states, the sum is somewhat more complicated but can be written in terms of a number of compact forms. Taking $I = 3/2$, $J = 3/2$ for the excited state in the D_2 line ($5P_{3/2}$), there are two possibilities depending on the type of transition. If we consider coupling to the ground ($5S_{1/2}$) state, then $J = 1/2$, $J' = 3/2$, and for the perpendicular dipole moment,

$$\sum_{F'}|\langle F, m_F|er_0|F', m_F\rangle|^2 = |\langle J\|e\mathbf{r}\|J'\rangle|^2 \times \begin{cases} (3-m_F)(3+m_F)/30, & F'=3 \\ 1/6, & F'=2 \\ (1+6m_F^2)/30, & F'=1 \\ 1/6, & F'=0 \end{cases}. \tag{13.125}$$

For the parallel dipole moment,

$$\begin{aligned}\sum_{F'}\left(|\langle F, m_F|er_1|F', (m_F-1)\rangle|^2 + |\langle F, m_F|er_{-1}|F', (m_F+1)\rangle|^2\right) = \\ |\langle J\|e\mathbf{r}\|J'\rangle|^2 \times \begin{cases} (12+m_F^2)/30, & F'=3 \\ 1/2, & F'=2 \\ (13-6m_F^2)/30, & F'=1 \\ 1/6m_F^2, & F'=0 \end{cases}.\end{aligned} \tag{13.126}$$

If we consider coupling to higher-lying excited states, then $J = 3/2$, $J' = 1/2, 3/2, 5/2$, and for the perpendicular dipole moment,

$$\sum_{F'} |\langle F, m_F|er_0|F', m_F\rangle|^2 = |\langle J\|e\mathbf{r}\|J'\rangle|^2 \times \begin{cases} (3-m_F)(3+m_F)/15, & F=3, J'=1/2 \\ 1/3, & F=2, J'=1/2 \\ (1+6m_F^2)/15, & F=1, J'=1/2 \\ 1/3, & F=0, J'=1/2 \\ (9+4m_F^2)/75, & F=3, J'=3/2 \\ 1/3, & F=2, J'=3/2 \\ (17+24m_F^2)/75, & F=1, J'=3/2 \\ 1/3, & F=0, J'=3/2 \\ (29-m_F^2)/75, & F=3, J'=5/2 \\ 1/3, & F=2, J'=5/2 \\ (7+2m_F^2)/25, & F=1, J'=5/2 \\ 1/3, & F=0, J'=5/2 \end{cases}. \tag{13.127}$$

For the parallel dipole moment,

$$\sum_{F'} \left(|\langle F, m_F|er_1|F', (m_F-1)\rangle|^2 + |\langle F, m_F|er_{-1}|F', (m_F+1)\rangle|^2 \right) = |\langle J\|e\mathbf{r}\|J'\rangle|^2 \times \begin{cases} (6+m_F^2)/15, & F=3, J'=1/2 \\ 2/3, & F=2, J'=1/2 \\ (14-6m_F^2)/15, & F=1, J'=1/2 \\ 2/3, & F=0, J'=1/2 \\ (66-45m_F^2)/75, & F=3, J'=3/2 \\ 2/3, & F=2, J'=3/2 \\ (41-24m_F^2)/75, & F=1, J'=3/2 \\ 2/3, & F=0, J'=3/2 \\ (46+m_F^2)/75, & F=3, J'=5/2 \\ 2/3, & F=2, J'=5/2 \\ (18-2m_F^2)/25, & F=1, J'=5/2 \\ 2/3, & F=0, J'=5/2 \end{cases}. \tag{13.128}$$

These formulae are sufficient to determine the dipole moments in the Casimir–Polder energy shift for the excited state.

13.11.1 Tabulated Data

Now we tabulate the lines involved in the trap-depth calculations. For the ground-state shift, we need the series of $nP_{3/2}$ and $nP_{1/2}$ transitions, as given here. Sources here are listed as Steck[12], NIST[13], Morton[14], Gomez[15], and Safronova[16].

[12]Daniel A. Steck, "Rubidium 87 D Line Data," available online at `http://steck.us/alkalidata`.

[13]NIST Atomic Spectra Database (version 3.0), Available online at `http://physics.nist.gov/PhysRefData /ASD/index.html`.

[14]Donald C. Morton, "Atomic Data for Resonance Absorption Lines. II. Wavelengths Longward of the Lyman Limit for Heavy Elements," *Astrophys. J. Supp. Ser.* **130**, 403 (2000) (doi: 10.1086/317349).

[15]E. Gomez, F. Baumer, A. D. Lange, G. D. Sprouse, and L. A. Orozco, "Lifetime measurement of the 6*s* level of rubidium," *Phys. Rev. A* **72**, 012502 (2005) (doi: 10.1103/PhysRevA.72.012502).

[16]M. S. Safronova, Carl J. Williams, and Charles W. Clark, "Relativistic many-body calculations of electric-dipole matrix elements, lifetimes, and polarizabilities in rubidium," *Phys. Rev. A* **69**, 022509 (2004) (doi: 10.1103/PhysRevA.69.022509).

transition	λ (nm)	source	$\tau_{J'J}$ (ns)	type	source
$5S_{1/2} \longrightarrow 5P_{3/2}$	780.241209686(13)	Steck	26.236(11)	expt	Steck
$5S_{1/2} \longrightarrow 5P_{1/2}$	794.9788509(8)	Steck	27.679(27)	expt	Steck
$5S_{1/2} \longrightarrow 6P_{3/2}$	420.2989(10)	NIST	565(28)	expt	Morton
$5S_{1/2} \longrightarrow 6P_{1/2}$	421.6726(10)	NIST	667(33)	expt	Morton
$5S_{1/2} \longrightarrow 7P_{3/2}$	358.8073(10)	NIST	2530(130)	expt	Morton
$5S_{1/2} \longrightarrow 7P_{1/2}$	359.2597(10)	NIST	3460(170)	expt	Morton
$5S_{1/2} \longrightarrow 8P_{3/2}$	334.9658(10)	NIST	7300(360)	expt	Morton
$5S_{1/2} \longrightarrow 8P_{1/2}$	335.1775(10)	NIST	$1.122(56) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 9P_{3/2}$	322.8911(10)	NIST	$1.563(78) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 9P_{1/2}$	323.0088(10)	NIST	$2.60(13) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 10P_{3/2}$	315.8444(10)	NIST	$2.96(15) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 10P_{1/2}$	315.9173(10)	NIST	$4.98(25) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 11P_{3/2}$	311.3468(10)	NIST	$3.98(20) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 11P_{1/2}$	311.3950(10)	NIST	$7.87(39) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 12P_{3/2}$	308.2893(10)	NIST	$6.71(34) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 12P_{1/2}$	308.3229(10)	NIST	$1.43(72) \times 10^5$	expt	Morton
$5S_{1/2} \longrightarrow 13P_{3/2}$	306.1131(10)	NIST	$9.52(48) \times 10^4$	expt	Morton
$5S_{1/2} \longrightarrow 13P_{1/2}$	306.1375(10)	NIST	$2.22(11) \times 10^5$	expt	Morton

The additional data for the treatment of the D_2 excited ($5P_{3/2}$) state are as follows:

transition	λ (nm)	source	$\tau_{J'J}$ (ns)	type	source
$5P_{3/2} \longrightarrow 6S_{1/2}$	1366.875(10)	NIST	68.35(26)	expt	Gomez
$5P_{3/2} \longrightarrow 7S_{1/2}$	741.02136(10)	NIST	219.7(88)	theory	Safronova
$5P_{3/2} \longrightarrow 8S_{1/2}$	616.13310(10)	NIST	458(18)	theory	Safronova
$5P_{3/2} \longrightarrow 4D_{3/2}$	1529.261(10)	NIST	563(23)	theory	Safronova
$5P_{3/2} \longrightarrow 4D_{5/2}$	1529.366(10)	NIST	93.7(37)	theory	Safronova
$5P_{3/2} \longrightarrow 5D_{3/2}$	776.15716(10)	NIST	1490(340)	theory	Safronova
$5P_{3/2} \longrightarrow 5D_{5/2}$	775.97855(10)	NIST	254(56)	theory	Safronova
$5P_{3/2} \longrightarrow 6D_{3/2}$	630.09666(10)	NIST	1586(20)	theory	Safronova
$5P_{3/2} \longrightarrow 6D_{5/2}$	630.00670(10)	NIST	269(20)	theory	Safronova

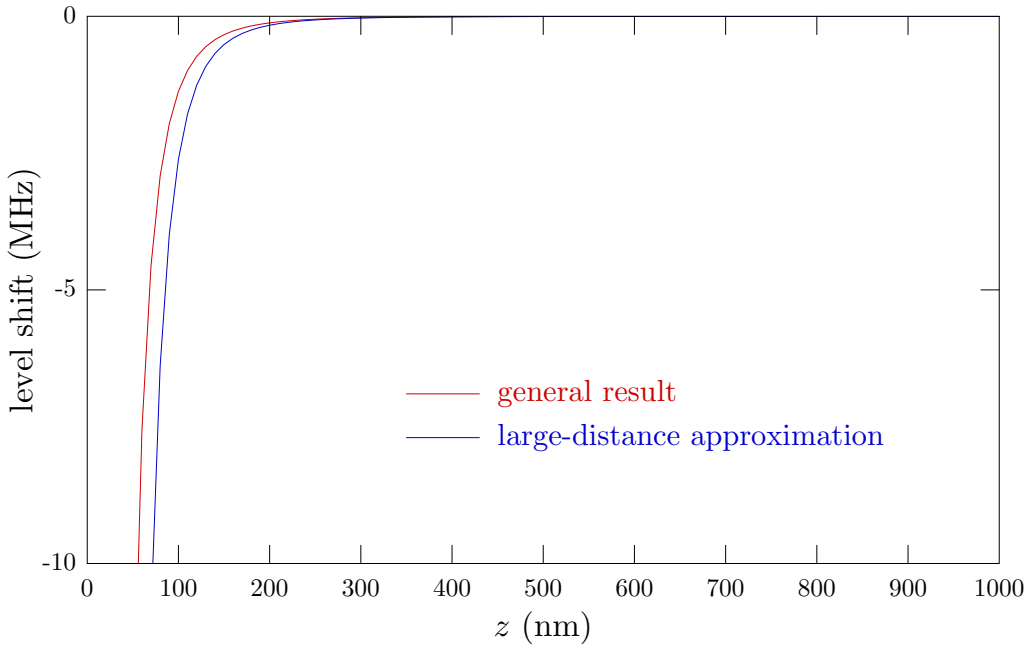
The additional data for the treatment of the D_1 excited ($5P_{1/2}$) state are as follows:

transition	λ (nm)	source	$\tau_{J'J}$ (ns)	type	source
$5P_{1/2} \longrightarrow 6S_{1/2}$	1323.879(10)	NIST	136.71(51)	expt	Gomez
$5P_{1/2} \longrightarrow 7S_{1/2}$	728.20028(10)	NIST	419(17)	theory	Safronova
$5P_{1/2} \longrightarrow 8S_{1/2}$	607.24355(10)	NIST	870(35)	theory	Safronova
$5P_{1/2} \longrightarrow 4D_{3/2}$	1475.644(10)	NIST	103(41)	theory	Safronova
$5P_{1/2} \longrightarrow 5D_{3/2}$	762.10304(10)	NIST	335(82)	theory	Safronova
$5P_{3/2} \longrightarrow 6D_{3/2}$	620.80263(10)	NIST	339(29)	theory	Safronova

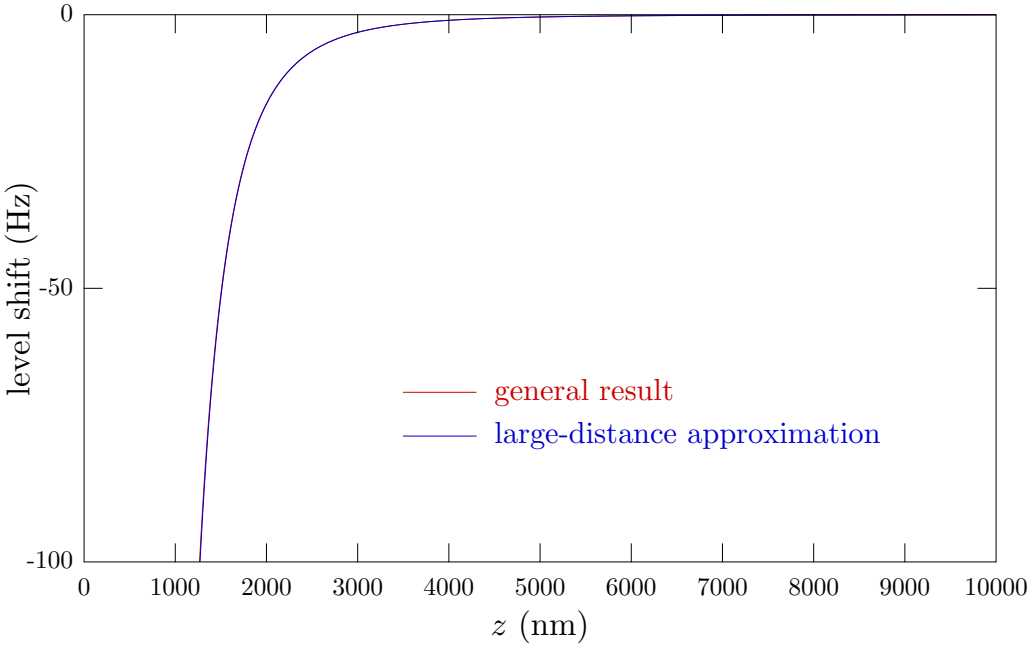
13.11.2 Results

The ground-level shift is shown here with the long-distance Casimir–Polder result ($\sim 1/z^4$), where the deviation is clear at distances much smaller than the main transition wavelengths around 800 nm. The shifts in the long-distance approximation computed by summing the dipole moments agree with those computed from the static polarizability value[17] of $\alpha_0 = h \cdot 0.122\ 306(16)$ Hz/(V/cm)^2 to within a few percent.

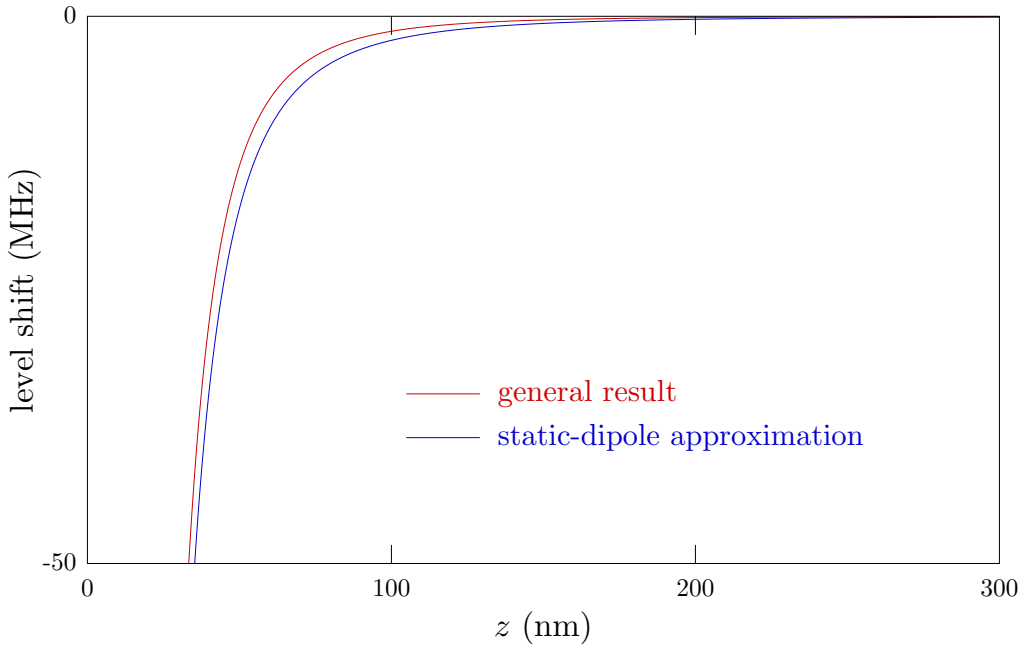
[17] Daniel A. Steck, "Rubidium 87 D Line Data" unpublished, available online at `http://steck.us/alkalidata`.



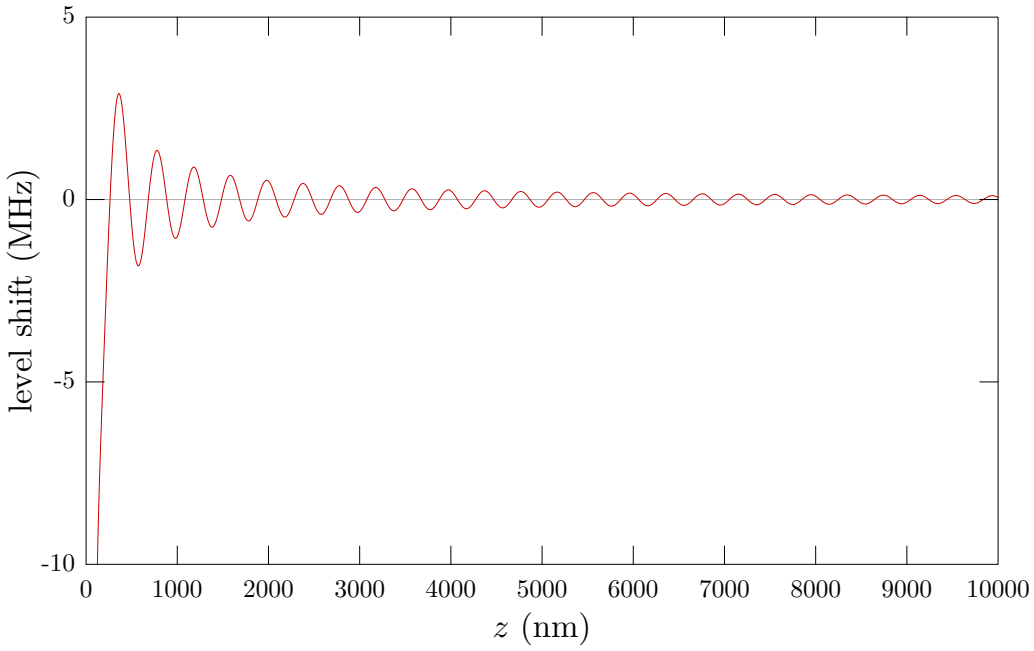
The shift is much smaller at large distances. Here, the general and large-distance results are visually indistinguishable.



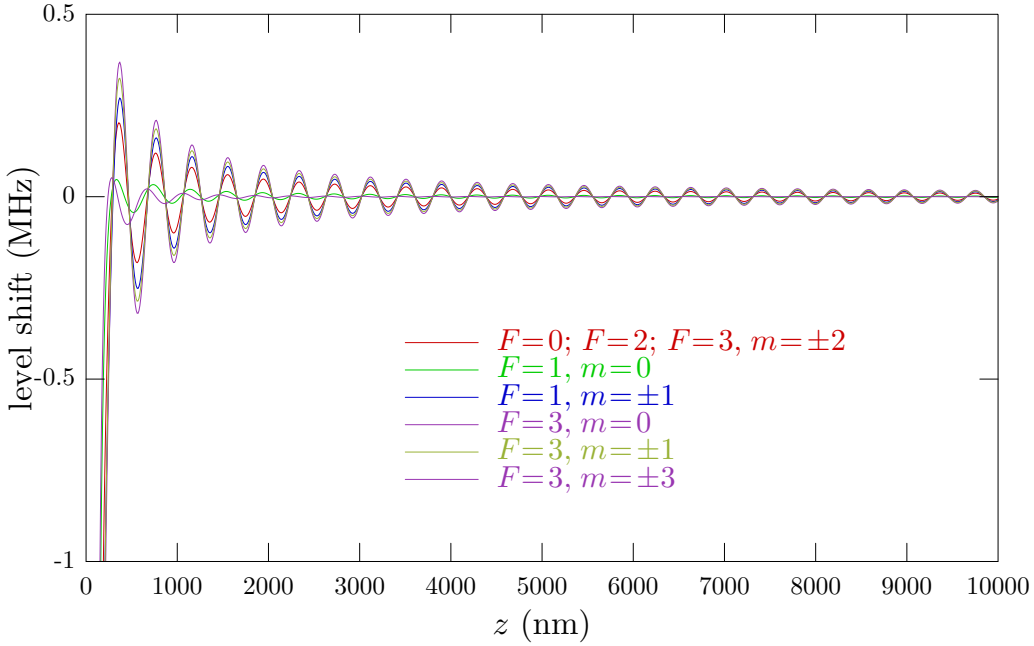
Comparing the general result instead to the small-distance (static-dipole) expression, we also see deviations in the intermediate regime. It appears that the small-distance result is only very accurate in the regime of such short distances that the accuracy of this calculation is questionable.



The shift of the D_1 excited manifold is much stronger and oscillates due to the standing-wave pattern formed by the atomic radiation. The shift here is scalar, so that all hyperfine levels in the manifold are shifted equally.



The shifts of the levels in the D_2 excited manifold depend substantially on the level, but the same general oscillatory behavior is apparent.



13.12 Lamb Shift

The effect that we ignored when renormalizing the divergence in Section 13.3 is the **Lamb shift**,[18] which is the shift of any atomic transition frequency due to the quantum vacuum, even in free space.

13.12.1 Coulomb Gauge

We will now first examine a conventional nonrelativistic calculation of the Lamb shift using the usual Coulomb-gauge interaction Hamiltonian

$$H_{\rm AF}^{(A)} = \frac{e}{m_{\rm e}}\mathbf{p}_{\rm e}\cdot\mathbf{A} + \frac{e^2}{2m_{\rm e}}A^2. \tag{13.129}$$

Again using second-order perturbation theory to compute the shift of level α, we can write

$$\Delta E_\alpha = \langle\alpha|H_{\rm AF}|\alpha\rangle + \sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle\alpha|H_{\rm AF}|j,1_{\mathbf{k},\zeta}\rangle|^2}{E_{\alpha,0} - E_{j,1_{\mathbf{k},\zeta}}}, \tag{13.130}$$

which we will again evaluate to order e^2 (second order in the atomic dipole moment).

We will start by evaluating the first-order perturbation, to which only the A^2 term contributes. Thus, we will write the first-order shift as

$$\Delta E_\alpha^{(1)} = \frac{e^2}{2m_{\rm e}}\langle\alpha|A^2|\alpha\rangle. \tag{13.131}$$

Using the expression

$$\mathbf{A}_{\mathbf{k},\zeta}(\mathbf{r},t) = i\sqrt{\frac{\hbar}{2\omega\epsilon_0 V}}\hat{\varepsilon}_{\mathbf{k},\zeta}e^{i\mathbf{k}\cdot\mathbf{r}}a_{\mathbf{k},\zeta}(t) + \text{H.c.} \tag{13.132}$$

[18] Willis E. Lamb, Jr. and Robert C. Retherford, "Fine Structure of the Hydrogen Atom by a Microwave Method," *Physical Review* **72**, 241 (1947) (doi: 10.1103/PhysRev.72.241); H. A. Bethe, "The Electromagnetic Shift of Energy Levels," *Physical Review* **72**, 339 (1947) (doi: 10.1103/PhysRev.72.339); H. A. Bethe, L. M. Brown, and J. R. Stehn, "Numerical Value of the Lamb Shift," *Physical Review* **77**, 370 (1950) (doi: 10.1103/PhysRev.77.370); Edwin A. Power, "Zero-Point Energy and the Lamb Shift," *American Journal of Physics* **34**, 516 (1966) (doi: 10.1119/1.1973082).

from Eq. (8.68) for the vector potential quantized in free space, we can proceed in the usual way so that the first-order shift becomes

$$\begin{aligned}\Delta E_\alpha^{(1)} &= \frac{e^2\hbar}{4m_\mathrm{e}\epsilon_0 V}\sum_{\mathbf{k},\zeta}\frac{1}{\omega_\mathbf{k}} \\ &= \frac{e^2\hbar}{2m_\mathrm{e}\epsilon_0 V}\sum_{\mathbf{k}}\frac{1}{\omega_\mathbf{k}} \\ &= \frac{e^2\hbar}{2(2\pi)^3 m_\mathrm{e}\epsilon_0 c}\int d^3k\,\frac{1}{k} \\ &= \frac{e^2\hbar}{4\pi^2 m_\mathrm{e}\epsilon_0 c}\int_0^\infty dk\,k.\end{aligned} \tag{13.133}$$

This result is independent of the state $|\alpha\rangle$, and therefore only produces an overall shift of the atomic energy level. Since it does not contribute to the observable shifts of the atomic transition energies, we can neglect it.

In the second-order shift, only the $\mathbf{p}_\mathrm{e}\cdot\mathbf{A}$ term contributes, and thus

$$\begin{aligned}\Delta E_\alpha = \Delta E_\alpha^{(2)} &= \frac{e^2}{m_\mathrm{e}^2}\sum_j\sum_{\mathbf{k},\zeta}\frac{|\langle\alpha|\mathbf{p}_\mathrm{e}\cdot\mathbf{A}|j,1_{\mathbf{k},\zeta}\rangle|^2}{E_{\alpha,0}-E_{j,1_{\mathbf{k},\zeta}}} \\ &= \frac{e^2}{m_\mathrm{e}^2}\sum_j\sum_{\mathbf{k},\zeta}\frac{|\langle\alpha|\mathbf{p}_\mathrm{e}\cdot\mathbf{A}|j,1_{\mathbf{k},\zeta}\rangle|^2}{E_\alpha-E_j-\hbar\omega_\mathbf{k}} \\ &= -\frac{e^2}{m_\mathrm{e}^2}\frac{\hbar}{2\epsilon_0 V}\sum_j\sum_{\mathbf{k},\zeta}\frac{1}{\omega_\mathbf{k}}\frac{|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle\cdot\hat{\varepsilon}_{\mathbf{k},\zeta}|^2}{\hbar(\omega_{j\alpha}+\omega_\mathbf{k})}.\end{aligned} \tag{13.134}$$

Recall here that $\hbar\omega_{j\alpha} := E_j - E_\alpha$ and thus can be positive if level j is an excited state or negative if $|j\rangle$ is a ground state with respect to $|\alpha\rangle$. In the continuum limit, the angular part of mode sum amounts an average over the relative orientation of the dipole and the field direction. We can thus take a uniform average over all dipole orientations and replace the squared dot product by a factor of $1/3$. Continuing with the calculation,

$$\begin{aligned}\Delta E_\alpha = \Delta E_\alpha^{(2)} &= -\frac{e^2}{3m_\mathrm{e}^2\epsilon_0 V}\sum_j\sum_{\mathbf{k}}\frac{|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2}{\omega_\mathbf{k}(\omega_{j\alpha}+\omega_\mathbf{k})} \\ &= -\frac{e^2}{3(2\pi)^3 m_\mathrm{e}^2\epsilon_0 c^2}\sum_j|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2\int d^3k\,\frac{1}{k(k_{j\alpha}+k)} \\ &= -\frac{e^2}{6\pi^2 m_\mathrm{e}^2\epsilon_0 c^2}\sum_j|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2\int_0^\infty dk\,\frac{k}{k_{j\alpha}+k}.\end{aligned} \tag{13.135}$$

Clearly this expression is divergent, and the divergence is asymptotically linear in k. The basic problem here is that high energies are only treated correctly in *relativistic* theory, and this is a nonrelativistic calculation. Nevertheless, Bethe[19] used this nonrelativistic theory in a clever way to produce a finite prediction for the energy-level shift.

The first step in Bethe's argument is to realize that to get a finite energy shift, we will need to cut off the integral. Supposing that we cut off contributions from energies larger than some large energy Λ, we can write

$$\Delta E_\alpha = \Delta E_\alpha^{(2)} = -\frac{e^2}{6\pi^2 m_\mathrm{e}^2\epsilon_0 c^2}\sum_j|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2\int_0^{\Lambda/\hbar c}dk\,\frac{k}{k_{j\alpha}+k}. \tag{13.136}$$

The next step is Bethe's **mass renormalization**. The idea is that the above energy shift contains the energy of the *free* electron, which is unobservable; we can only observe the shifts in *transition frequencies*, which

[19] H. A. Bethe, "The Electromagnetic Shift of Energy Levels," *Physical Review* **72**, 339 (1947) (doi: 10.1103/PhysRev.72.339).

are related to the electron energies only in *bound* states. We can find the free-electron energy by taking the limit $k_{j\alpha} \longrightarrow 0$ in Eq. (13.136), corresponding to a vanishing binding potential, so that

$$\Delta E_\alpha^{\text{free}} = -\frac{e^2}{6\pi^2 m_e^2 \epsilon_0 c^2} \sum_j |\langle\alpha|\mathbf{p}_e|j\rangle|^2 \int_0^{\Lambda/\hbar c} dk. \tag{13.137}$$

Subtracting this from Eq. (13.136), we use the integrand subtraction

$$\frac{k}{k_{j\alpha}+k} - 1 = -\frac{k_{j\alpha}}{k_{j\alpha}+k}, \tag{13.138}$$

and thus we find the renormalized energy

$$\Delta E_\alpha - \Delta E_\alpha^{\text{free}} = \frac{e^2}{6\pi^2 m_e^2 \epsilon_0 c^2} \sum_j k_{j\alpha} |\langle\alpha|\mathbf{p}_e|j\rangle|^2 \int_0^{\Lambda/\hbar c} \frac{dk}{k_{j\alpha}+k}. \tag{13.139}$$

This renormalization is important in that it has reduced the divergence from linear to logarithmic, and thus the result is now relatively insensitive to the value of the cutoff Λ. The interpretation of the renormalization is as follows: in relativistic theory, the free electron energy due to the field coupling shifts the atomic rest mass, but by using the *observed* (or *renormalized*) mass m_e, we have already included this contribution, and should not double-count it. This is the rationale for subtracting it after computing the atom-field coupling energy.[20]

Now we can carry out the integration in Eq. (13.139), using

$$\begin{aligned}
\int_0^t \frac{dx}{x+a} &= \int_0^t \frac{d(x+a)}{x+a} \\
&= \log|x+a|\Big|_0^t \\
&= \log|t+a| - \log|a| \\
&= \log\left|\frac{t}{a}+1\right| \\
&\approx \log\frac{t}{|a|},
\end{aligned} \tag{13.140}$$

where the last equality holds if $t > 0$ and $t \gg |a|$. Thus, we obtain the **Bethe logarithm**

$$\Delta E_\alpha - \Delta E_\alpha^{\text{free}} = \frac{e^2}{6\pi^2 m_e^2 \epsilon_0 c^2} \sum_j k_{j\alpha} |\langle\alpha|\mathbf{p}_e|j\rangle|^2 \log\frac{\Lambda}{\hbar|\omega_{j\alpha}|}. \tag{13.141}$$

The final step is to choose the cutoff Λ. Bethe chose $\Lambda = m_e c^2$ as a reasonable energy beyond which the nonrelativistic theory should fail. Thus,

$$\Delta E_\alpha - \Delta E_\alpha^{\text{free}} = \frac{e^2}{6\pi^2 m_e^2 \epsilon_0 c^2} \sum_j k_{j\alpha} |\langle\alpha|\mathbf{p}_e|j\rangle|^2 \log\frac{m_e c^2}{\hbar|\omega_{j\alpha}|}. \tag{13.142}$$

(Lamb shift)

This is the basic, nonrelativistic result for the Lamb shift of level $|\alpha\rangle$. Notice that for the ground state, $k_{jg} > 0$, and thus the Lamb shift is *positive*. Viewed as a Stark shift, evidently the ultraviolet modes far above the dominant transition frequencies are the most important in determining the shift.

[20] For a much more detailed discussion of this point, as well as a nice historical account and many viewpoints of the Lamb shift, see Peter W. Milonni, *op. cit.*, Sections 3.4-3.9, pp. 82-96.

13.12.1.1 Evaluation

Using this energy shift, Bethe computed a numerical value for the Lamb shift of the 2S $\longrightarrow$ 2P fine-structure transition in atomic hydrogen—a transition between degenerate states according to the solutions to the Dirac equation. As a rough approximation, Bethe noted that the arguments of the logarithms are large and relatively weakly dependent on the transition frequencies. As a rough approximation, he assumed it was constant in the level summation and replaced it by an average excitation energy. Thus, the level sum amounts to the sum

$$\begin{aligned}
\sum_j k_{j\alpha} |\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2 &= \frac{1}{\hbar c}\sum_j (E_j - E_\alpha)\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle \cdot \langle j|\mathbf{p}_\mathrm{e}|\alpha\rangle \\
&= \frac{1}{\hbar c}\sum_j \langle\alpha|[\mathbf{p}_\mathrm{e}, H_\mathrm{A}]|j\rangle \cdot \langle j|\mathbf{p}_\mathrm{e}|\alpha\rangle \\
&= -\frac{i}{c}\langle\alpha|(\nabla_\mathrm{e} V)\cdot\mathbf{p}_\mathrm{e}|\alpha\rangle \\
&= -\frac{i}{2c}\langle\alpha|[\nabla_\mathrm{e} V, \mathbf{p}_\mathrm{e}]|\alpha\rangle \\
&= \frac{\hbar}{2c}\langle\alpha|\nabla_\mathrm{e}^2 V|\alpha\rangle \\
&= \frac{\hbar}{2c}\int d^3 r_\mathrm{e}\, |\psi_\alpha(\mathbf{r}_\mathrm{e})|^2 \nabla_\mathrm{e}^2 V(\mathbf{r}_\mathrm{e}) \\
&= \frac{2\pi\hbar Z e^2}{c}|\psi_\alpha(0)|^2,
\end{aligned} \tag{13.143}$$

where in the last step we used the Coulomb binding potential

$$V(\mathbf{r}_\mathrm{e}) = -\frac{Ze^2}{r_\mathrm{e}}, \tag{13.144}$$

so that

$$\nabla_\mathrm{e}^2 V(\mathbf{r}_\mathrm{e}) = 4\pi Z e^2 \delta^3(\mathbf{r}_\mathrm{e}). \tag{13.145}$$

In this approximation, then only S ($l = 0$) states have a Lamb shift, since their probability densities are nonvanishing at the nucleus. We can then write the observed Lamb shift from (13.142) as

$$\Delta E_\alpha^\mathrm{observed} \approx \frac{4\alpha_0 Z}{3}\left(\frac{e\hbar}{m_\mathrm{e} c}\right)^2 |\psi_\alpha(0)|^2 \log\frac{m_\mathrm{e} c^2}{\hbar|(\omega_{j\alpha})_\mathrm{avg}|}, \tag{13.146}$$

where again $\alpha_0 := e^2/4\pi\epsilon_0\hbar c$ is the fine-structure constant. Then for a hydrogenic S state in orbital n,

$$|\psi_n(0)|^2 = \frac{1}{\pi}\left(\frac{Z}{n a_0}\right)^3, \tag{13.147}$$

where $a_0 = \hbar/m_\mathrm{e} c\alpha_0$ is the Bohr radius, so that

$$\Delta E_n^\mathrm{observed} \approx \frac{8\alpha_0^3 Z^4}{3\pi n^3} R_\infty \log\frac{m_\mathrm{e} c^2}{\hbar|(\omega_{j\alpha})_\mathrm{avg}|}, \tag{13.148}$$

(S-state Lamb shift)

where $R_\infty = e^2/2a_0 \approx 13.6$ eV is the Rydberg energy (hydrogen ionization energy). Bethe used the average value $(\omega_{j\alpha})_\mathrm{avg} \approx 17.8\, R_\infty$, computed by averaging $\log|(\omega_{j\alpha})_\mathrm{avg}|$, weighted by $|\omega_{j\alpha}|\,|\langle\alpha|\mathbf{p}_\mathrm{e}|j\rangle|^2$. The value here is much larger than R_∞ since evidently ionized states make a large contribution to the sum. Using these results, Bethe arrived at a shift of 1040 MHz for the 2S state (with negligible shift for the 2P state). This is in surprisingly good agreement with the modern value of about 1058 MHz, considering the *ad hoc* nature of the calculation.

13.12.2 Electric Dipole Interaction

It is again interesting to see how the Lamb shift arises from the electric-dipole Hamiltonian, which we showed in Eq. (9.38) has the form in free space takes the form

$$H_{\text{AF}} = -\mathbf{d}\cdot\mathbf{E}^{\perp}(0) + \frac{1}{2\epsilon_0}\int d^3r\, P^{\perp 2}(\mathbf{r}), \tag{13.149}$$

where the last term is the dipole self-energy, with

$$P^{\perp}_{\alpha}(\mathbf{r}) := -er_{\text{e},\beta}\delta^{\perp}_{\alpha\beta}(\mathbf{r}). \tag{13.150}$$

Starting with the second-order shift due to the $-\mathbf{d}\cdot\mathbf{E}^{\perp}(0)$ part of the interaction, we have already computed as the position-independent part of the Casimir–Polder expression (13.17), which we can write as a shift for level $|\alpha\rangle$ as

$$\Delta E^{\mathbf{d}\cdot\mathbf{E}}_{\alpha} = -\frac{1}{2\epsilon_0 V}\sum_j\sum_{\mathbf{k}}\frac{\omega_{\mathbf{k}}}{(\omega_{j\alpha}+\omega_{\mathbf{k}})}\left[\left(d^2_{j\alpha,\parallel}/2 + d^2_{j\alpha,z}\right) + \frac{k_z^2}{k^2}\left(d^2_{j\alpha,\parallel}/2 - d^2_{j\alpha,z}\right)\right]. \tag{13.151}$$

For a spherically symmetric atom, the asymmetric part of the dipole vanishes, and so

$$\begin{aligned}\Delta E^{\mathbf{d}\cdot\mathbf{E}}_{\alpha} &= -\frac{1}{3\epsilon_0 V}\sum_j |d_{j\alpha}|^2\sum_{\mathbf{k}}\frac{\omega_{\mathbf{k}}}{\omega_{j\alpha}+\omega_{\mathbf{k}}}\\ &= -\frac{1}{3(2\pi)^3\epsilon_0}\sum_j |d_{j\alpha}|^2\int d^3k\,\frac{k}{k_{j\alpha}+k}\\ &= -\frac{1}{6\pi^2\epsilon_0}\sum_j |d_{j\alpha}|^2\int_0^{\infty} dk\,\frac{k^3}{k_{j\alpha}+k}.\end{aligned} \tag{13.152}$$

Again, the infinite upper limit here is understood to be an appropriate ultraviolet cutoff. Note, however, that the asymptotic scaling of the integrand is now k^2, where it was k^0 in the $\mathbf{p}\cdot\mathbf{A}$ calculation. Obviously we are missing some contributions that will make the scaling correct.

13.12.2.1 Dipole Self-Energy

Now to evaluate the dipole self-energy term

$$H_{\mathbf{P}^{\perp}} = \frac{1}{2\epsilon_0}\int d^3r\, P^{\perp 2}(\mathbf{r}) \tag{13.153}$$

of the interaction Hamiltonian. Since $\partial_\alpha\delta^{\perp}_{\alpha\beta}(\mathbf{r}) = 0$ (Problem 8.7), the transverse delta function is *itself* a transverse vector field for any particular value of β, and thus we may write (see also Problem 8.11)

$$\int d^3r\,\delta^{\perp}_{\alpha\beta}(\mathbf{r})\delta^{\perp}_{\beta\gamma}(\mathbf{r}) = \delta^{\perp}_{\alpha\gamma}(0). \tag{13.154}$$

Thus,

$$\begin{aligned}H_{\mathbf{P}^{\perp}} &= \frac{1}{2\epsilon_0}\int d^3r\, P^{\perp 2}(\mathbf{r})\\ &= \frac{e^2}{2\epsilon_0}\int d^3r\, r_{\text{e},\alpha}\delta^{\perp}_{\alpha\beta}(\mathbf{r})\delta^{\perp}_{\beta\gamma}(\mathbf{r})r_{\text{e},\gamma}\\ &= \frac{e^2}{2\epsilon_0}r_{\text{e},\alpha}r_{\text{e},\beta}\delta^{\perp}_{\alpha\beta}(0)\\ &= \frac{e^2}{2\epsilon_0(2\pi)^3}r_{\text{e},\alpha}r_{\text{e},\beta}\int d^3k\,\left(\delta_{\alpha\beta} - \frac{k_\alpha k_\beta}{k^2}\right),\end{aligned} \tag{13.155}$$

where we have used the momentum-space representation of the transverse delta function from Eq. (8.178).

Again, since we are calculating the Lamb shift to order e^2, it is sufficient to consider the level shift due to the dipole self-energy to *first* order in perturbation theory. In this case, the shift of level α is

$$\begin{aligned}\Delta E_\alpha^{\mathbf{P}^\perp} &= \langle\alpha|H_{\mathbf{P}^\perp}|\alpha\rangle \\ &= \frac{e^2}{2\epsilon_0(2\pi)^3}\langle\alpha|r_{\mathrm{e},\mu}r_{\mathrm{e},\nu}|\alpha\rangle\int d^3k\,\left(\delta_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right) \\ &= \frac{e^2}{2\epsilon_0(2\pi)^3}\sum_j\langle\alpha|r_{\mathrm{e},\mu}|j\rangle\langle j|r_{\mathrm{e},\nu}|\alpha\rangle\int d^3k\,\left(\delta_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right) \\ &= \frac{1}{2\epsilon_0(2\pi)^3}\sum_j(d_{\alpha j})_\mu(d_{\alpha j}^*)_\nu\int d^3k\,\left(\delta_{\mu\nu}-\frac{k_\mu k_\nu}{k^2}\right).\end{aligned} \tag{13.156}$$

Notice that for a spherically symmetric atom, the dipole matrix elements are independent of direction, and thus $(d_{\alpha j})_\mu$ is independent of μ, and we can carry out the sum to write

$$\begin{aligned}\Delta E_\alpha^{\mathbf{P}^\perp} &= \langle\alpha|H_{\mathbf{P}^\perp}|\alpha\rangle \\ &= \frac{1}{2\epsilon_0(2\pi)^3}\sum_j\frac{|d_{\alpha j}|^2}{3}\int d^3k\,\left(3-\frac{k_\mu k_\mu}{k^2}\right) \\ &= \frac{1}{3\epsilon_0(2\pi)^3}\sum_j|d_{\alpha j}|^2\int d^3k \\ &= \frac{1}{6\pi^2\epsilon_0}\sum_j|d_{\alpha j}|^2\int_0^\infty dk\,k^2.\end{aligned} \tag{13.157}$$

We can then write the total shift as

$$\begin{aligned}\Delta E_\alpha &= \Delta E_\alpha^{\mathbf{d}\cdot\mathbf{E}}+\Delta E_\alpha^{\mathbf{P}^\perp} \\ &= \frac{1}{6\pi^2\epsilon_0}\sum_j|d_{j\alpha}|^2\int_0^\infty dk\,\left(-\frac{k^3}{k_{j\alpha}+k}+k^2\right) \\ &= \frac{1}{6\pi^2\epsilon_0}\sum_j k_{j\alpha}|d_{j\alpha}|^2\int_0^\infty dk\,\frac{k^2}{k_{j\alpha}+k}.\end{aligned} \tag{13.158}$$

By accounting for the dipole self-energy, we have reduced the order of the divergence, and we are on track to obtain the correct Lamb shift.[21]

13.12.2.2 Mass Renormalization

Again, we must subtract the free-electron energy, which we will calculate from the Coulomb-gauge Hamiltonian (13.129). We computed the contribution from the A^2 Hamiltonian in Eq. (13.133) and found

$$\Delta E_\alpha^{(1),\,\mathrm{free}} = \frac{e^2\hbar}{4\pi^2 m_\mathrm{e}\epsilon_0 c}\int_0^\infty dk\,k. \tag{13.159}$$

Recall from Eq. (13.41) that the TRK sum rule is

$$\sum_j\omega_{j\alpha}|\langle\alpha|r_{\mathrm{e},\beta}|j\rangle|^2 = \frac{\hbar}{2m_\mathrm{e}}, \tag{13.160}$$

[21] The importance of the dipole self-energy and the following renormalization procedure were pointed out by E. A. Power and S. Zienau, "Coulomb Gauge in Non-Relativistic Quantum Electro-Dynamics and the Shape of Spectral Lines," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **251**, 427 (1959) (doi: 10.1098/rsta.1959.0008).

which we can also write as

$$\sum_j \omega_{j\alpha} |\langle \alpha | \mathbf{r}_\mathrm{e} | j \rangle|^2 = \frac{3\hbar}{2m_\mathrm{e}}. \tag{13.161}$$

This gives

$$\Delta E_\alpha^{(1),\,\mathrm{free}} = \frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha} |d_{j\alpha}|^2 \int_0^\infty dk\, k. \tag{13.162}$$

Subtracting this part of the free-electron energy from the Lamb shift (13.158), we find

$$\begin{aligned} \Delta E_\alpha - \Delta E_\alpha^{(1),\,\mathrm{free}} &= \frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha} |d_{j\alpha}|^2 \int_0^\infty dk \left(\frac{k^2}{k_{j\alpha}+k} - k \right) \\ &= -\frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha}^2 |d_{j\alpha}|^2 \int_0^\infty dk\, \frac{k}{k_{j\alpha}+k}. \end{aligned} \tag{13.163}$$

This result is, in fact, equivalent to the result (13.135) in the Coulomb-gauge Hamiltonian before mass renormalization in that gauge. We can again perform the same renormalization with the $\mathbf{p}\cdot\mathbf{A}$ part of the Hamiltonian, which we have already computed in Eq. (13.137):

$$\begin{aligned} \Delta E_\alpha^{(2),\,\mathrm{free}} &= -\frac{e^2}{6\pi^2 m_\mathrm{e}^2 \epsilon_0 c^2} \sum_j |\langle \alpha | \mathbf{p}_\mathrm{e} | j \rangle|^2 \int_0^\infty dk \\ &= -\frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha}^2 |d_{j\alpha}|^2 \int_0^\infty dk. \end{aligned} \tag{13.164}$$

Subtracting this part (i.e., the rest) of the electron free energy, we find

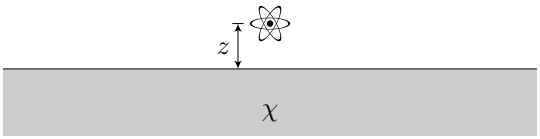
$$\begin{aligned} \Delta E_\alpha - \Delta E_\alpha^\mathrm{free} &= \frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha}^2 |d_{j\alpha}|^2 \int_0^\infty dk \left(-\frac{k}{k_{j\alpha}+k} + 1 \right) \\ &= \frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha}^3 |d_{j\alpha}|^2 \int_0^\infty \frac{dk}{k_{j\alpha}+k} \\ &= \frac{1}{6\pi^2\epsilon_0} \sum_j k_{j\alpha}^3 |d_{j\alpha}|^2 \log \frac{m_\mathrm{e} c^2}{\hbar |\omega_{j\alpha}|}. \end{aligned} \tag{13.165}$$

This result is exactly equivalent to the Coulomb-gauge result (13.142), if we use the conversion between the momentum and dipole matrix elements.

13.13 Casimir–Polder Potential for a Rarefied Dielectric Surface

At this point, we return to the atom–surface potential, but instead of a perfect conductor, as we have been considering, we will consider instead a *dielectric* planar surface, with the dielectric filling half of all space. This problem is better handled with the powerful formalism of electromagnetic Green tensors [Section 14.3.5.6, particularly Eqs. (14.211) and (14.212)]; however, this setup is a nice example of the mode-summation formalism applied to a different set of modes.[22] We will also approach the problem slightly differently, computing the local (renormalized) *energy density* of the electromagnetic field due to the surface, and then convert this into a Casimir–Polder potential. To keep things simple, we will assume a ''rarefied'' dielectric, such that the dielectric susceptibility χ [see Eqs. (14.16) and (14.18)] is small (i.e., we will work to first order in χ). We will also ignore dispersion of the dielectric (i.e., χ is independent of frequency).

[22] for a general approach to the Casimir–Polder potential of an atom near a planar dielectric via mode summation, see Shin-Tza Wu and Claudia Eberlein, ''Quantum Electrodynamics of an Atom in Front of a Non-Dispersive Dielectric Half-Space,'' *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **455**, 2487 (1999) (doi: 10.1098/rspa.1999.0413). See also Nicola Bartolo and Roberto Passante, ''Electromagnetic-field fluctuations near a dielectric-vacuum boundary and surface divergences in the ideal conductor limit,'' *Physical Review A* **86**, 012112 (2012) (doi: 10.1103/PhysRevA.86.012122).



For the geometry, we will assume as before that the atom is a distance z above the dielectric, with the dielectric occupying the half space $z < 0$.

We want to calculate the electromagnetic energy density for the vacuum state of the field, which is the expectation value of the Hamiltonian (8.34),

$$\begin{aligned}\mathscr{E}(\mathbf{r}) &= \mathscr{E}_E(\mathbf{r}) + \mathscr{E}_B(\mathbf{r}) \\ &= \frac{\epsilon_0}{2}\langle \mathbf{E}^2(\mathbf{r},t)\rangle + \frac{\epsilon_0 c^2}{2}\langle \mathbf{B}^2(\mathbf{r},t)\rangle \\ &= \frac{\epsilon_0}{2}\langle 0|\mathbf{E}^2(\mathbf{r},t)|0\rangle + \frac{\epsilon_0 c^2}{2}\langle 0|\mathbf{B}^2(\mathbf{r},t)|0\rangle,\end{aligned} \tag{13.166}$$

where we use ϵ_0 here instead of $\epsilon(\mathbf{r})$ since we are interested in the energy density *outside* the dielectric. Note that the two contributions here, from the electric and magnetic fields, are in fact equal in vacuum, but we will see that their *changes* due to a planar boundary are *not* the same.

We will begin with the electric-field energy, which is the important part for a polarizable, nonmagnetic atom interacting with the dielectric, as it will not "see" the magnetic field (more generally, the interaction via the magnetic dipole moment will be much weaker than via the electric dipole moment). The quantum electric field from Eq. (8.56) is

$$\mathbf{E}(\mathbf{r},t) = -\sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\,\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})\,a_{\mathbf{k},\zeta}(t) + \text{H.c.}, \tag{13.167}$$

where the $a_{\mathbf{k},\zeta}$ are the field annihilation operators, and the $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$ are the unit-normalized mode functions. Noting that in the vacuum expectation value, the only nonvanishing terms have the form of $aa^\dagger$ for some mode, and so

$$\mathscr{E}_E(\mathbf{r}) = \frac{1}{2}\sum_{\mathbf{k},\zeta}\frac{\hbar\omega_{\mathbf{k}}}{2}|\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2\langle 0|a_{\mathbf{k},\zeta}a^\dagger_{\mathbf{k},\zeta}|0\rangle = \frac{1}{2}\sum_{\mathbf{k},\zeta}\frac{\hbar\omega_{\mathbf{k}}}{2}|\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})|^2, \tag{13.168}$$

which is half of the zero-point energy of each mode, spatially distributed via the mode function, and then summed over all modes. We only have a half of the zero-point energy because we are ignoring the half due to the magnetic fields thus far.

To set up the modes, we will need the Fresnel reflection coefficients[23]

$$\begin{aligned}r_{\text{TE}}(\theta) &= \frac{\cos\theta - \sqrt{1+\chi-\sin^2\theta}}{\cos\theta + \sqrt{1+\chi-\sin^2\theta}} = -\frac{1}{4\cos^2\theta}\,\chi + O(\chi^2) \\ r_{\text{TM}}(\theta) &= \frac{\sqrt{1+\chi-\sin^2\theta} - (1+\chi)\cos\theta}{\sqrt{1+\chi-\sin^2\theta} + (1+\chi)\cos\theta} = \frac{1}{4}\left(\frac{1}{\cos^2\theta} - 2\right)\chi + O(\chi^2),\end{aligned} \tag{13.169}$$

which give the amplitude of the reflection of a plane wave from the dielectric surface, normalized to the incident amplitude. Here, θ is the angle of incidence, measured from the normal, and the two polarizations are transverse-electric (TE, where the electric field is parallel to the surface) and transverse-magnetic (TM, where the magnetic field is parallel to the surface). We have already expanded these expressions to lowest order in χ, in anticipation of our perturbative calculation.

[23] Daniel A. Steck, *Classical and Modern Optics* (2006), Chapter 9. Available at `http://steck.us/teaching`. Note that $r_{\text{TE}} \equiv r_{\text{S}}$ and $r_{\text{TM}} \equiv r_{\text{P}}$ in the notation there. See also Eq. (14.175) for further usage information.

13.13.1 TE Energy Density

To proceed with the computation of the energy density, we will begin with the TE modes. Modes incident from above must include incident, reflected, and transmitted plane-wave components, and the modes have the form

$$\mathbf{f}^{\downarrow}_{\mathbf{k},\mathrm{TE}}(\mathbf{r}) = \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\left[e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(z) + r_{\mathrm{TE}}(\theta)e^{i\mathbf{k}^{-}\cdot\mathbf{r}}\Theta(z) + \sqrt{1-|r_{\mathrm{TE}}|^2}e^{i\mathbf{k}_{\mathrm{t}}\cdot\mathbf{r}}\Theta(-z)\right], \tag{13.170}$$

where the arrow superscript indicates the direction of incidence, $\mathbf{k}$ is the incident wave vector (with $k_z < 0$), $\mathbf{k}^{-} = k_x\hat{x} + k_y\hat{y} - k_z\hat{z}$ is the reflected wave vector, and k_{t} is the transmitted wave vector. Notice that the polarizations of all fields match, as required by continuity of the surface-transverse component of the electric field.[24] (Although k_{t} is determined by Snell's Law, its precise form turns out to be irrelevent to the calculation here.) We have written the transmitted field in terms of the reflection coefficient so that the mode is explicitly normalized to a unit integral like the free-space mode functions, which is sufficient for our purposes here (technically, they should be normalized under an ϵ integral measure, but since we are staying outside the dielectric, this will not matter). Notice also that we have adopted unconfined mode functions (Section 8.7) here, so there is no explicit quantization volume. The contribution of these modes to the electric-field energy density (13.168) is

$$\mathscr{E}_{E^{\downarrow}_{\mathrm{TE}}}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^3}\int_{k_z<0} d^3k\, k|\mathbf{f}^{\downarrow}_{\mathbf{k},\mathrm{TE}}(\mathbf{r})|^2, \tag{13.171}$$

where we have written the sum as an integral and written $\omega_{\mathbf{k}} = ck$. Then for $z > 0$,

$$\begin{aligned}\mathscr{E}_{E^{\downarrow}_{\mathrm{TE}}}(\mathbf{r}) &= \frac{\hbar c}{4(2\pi)^3}\int_{k_z<0} d^3k\, k\left|e^{i\mathbf{k}\cdot\mathbf{r}} + r_{\mathrm{TE}}(\theta)e^{i\mathbf{k}^{-}\cdot\mathbf{r}}\right|^2 \\ &= \frac{\hbar c}{4(2\pi)^3}\int_{k_z<0} d^3k\, k\left[1 + r_{\mathrm{TE}}^2 + 2r_{\mathrm{TE}}\cos(2k_z z)\right].\end{aligned} \tag{13.172}$$

Since we only want the renormalized energy, we drop the z-independent terms:

$$\mathscr{E}_{E^{\downarrow}_{\mathrm{TE}}}(\mathbf{r}) = \frac{\hbar c}{2(2\pi)^3}\int_{k_z<0} d^3k\, k r_{\mathrm{TE}}(\theta)\cos(2k_z z). \tag{13.173}$$

Before proceeding, consider the TE modes incident from below, which are of the form

$$\mathbf{f}^{\uparrow}_{\mathbf{k},\mathrm{TE}}(\mathbf{r}) = \hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}\left[e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(-z) - r_{\mathrm{TE}}(\theta)e^{i\mathbf{k}^{-}\cdot\mathbf{r}}\Theta(-z) + \sqrt{1-|r_{\mathrm{TE}}|^2}e^{i\mathbf{k}_{\mathrm{t}}\cdot\mathbf{r}}\Theta(z)\right], \tag{13.174}$$

where the sign of the reflection is changed to respect the unitarity of the interface reflection, and k_{t} is once again the transmitted wave vector, which we will leave unspecified. The associated energy density is

$$\mathscr{E}_{E^{\uparrow}_{\mathrm{TE}}}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^3}\int_{k_z>0} d^3k\, k|\mathbf{f}^{\uparrow}_{\mathbf{k},\mathrm{TE}}(\mathbf{r})|^2 = \frac{\hbar c}{4(2\pi)^3}\int_{k_z>0} d^3k\, k\left(1 - |r_{\mathrm{TE}}|^2\right), \tag{13.175}$$

which is z-independent, and thus should be entirely dropped upon renormalization. Note that the total we have discarded so far is equivalent to the total energy density of the free TE field, and the only contribution that we have kept is the interference term between the incident and reflected fields.

Thus, the total TE energy density after renormalization is

$$\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{\hbar c}{2(2\pi)^3}\int_{k_z<0} d^3k\, k r_{\mathrm{TE}}(\theta)\cos(2k_z z). \tag{13.176}$$

This integral turns out to be somewhat tricky to evaluate. Writing out the integral in spherical coordinates, we have

$$\begin{aligned}\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) &= \frac{\hbar c}{8\pi^2}\int_0^\infty dk \int_0^{\pi/2} d\theta\, \sin\theta\, k^3 r_{\mathrm{TE}}(\theta)\cos(2kz\cos\theta) \\ &= \frac{\hbar c}{8\pi^2}\int_0^\infty dk \int_0^1 d\xi\, k^3 r_{\mathrm{TE}}(\xi)\cos(2kz\xi),\end{aligned} \tag{13.177}$$

[24] Daniel A. Steck, *op. cit.*

where we have chosen the range of θ integration to be compatible with the definition for the reflection-coefficient angle, and

$$\xi = \cos\theta. \tag{13.178}$$

The k integral can be performed most easily by inserting a convergence factor of $\exp(-ak)$ to cut off high frequencies, and the letting $a \longrightarrow 0$ afterwards (recall the discussion of convergence in Section 13.5, with the result

$$\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{3\hbar c}{64\pi^2 z^4}\int_0^1 d\xi\,\frac{r_{\mathrm{TE}}(\xi)}{\xi^4}. \tag{13.179}$$

This integral unfortunately diverges at $\xi = 0$; inserting the reflection coefficient doesn't help, since it is well-behaved at $\xi = 0$ (and diverges like $1/\xi^2$ for small χ). Our little convergence trick isn't valid if the result blows up!

13.13.1.1 Digression: Back to the Perfect Conductor

For inspiration on how to proceed, let's briefly go back to the familiar territory of the perfect conductor, where $r_{\mathrm{TE}} = -1$. We can then evaluate the integral as follows:

$$\begin{aligned}
\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) &= -\frac{\hbar c}{8\pi^2}\int_0^\infty dk\int_0^{\pi/2} d\theta\,\sin\theta\,k^3\cos(2kz\cos\theta)\\
&= -\frac{\hbar c}{8\pi^2}\int_0^\infty dk\,k^3\frac{\sin(2kz)}{2kz}\\
&= -\frac{\hbar c}{8\pi^2(2z)^4}\int_0^\infty dk\,k^2\sin k\\
&= -\lim_{a\to 0}\frac{\hbar c}{8\pi^2(2z)^4}\int_0^\infty dk\,k^2 e^{-ak}\sin k\\
&= -\frac{\hbar c}{8\pi^2(2z)^4}(-2)\\
&= \frac{\hbar c}{64\pi^2 z^4},
\end{aligned} \tag{13.180}$$

where we have first carried out the angular integral, and then carried out the k integral by inserting a convergence factor. Finally, the result is

$$\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{1}{6}\right). \tag{13.181}$$

(TE energy density, perfect conductor)

This is the energy density in the perfect-conductor limit, and we will comment on this after we complete the TE and TM calculations for the rarefied dielectric.

13.13.1.2 Integral for the Rarefied Dielectric

Evidently, the key to the success in the perfect-conductor case is in first carrying out the angular integration, and we will be more careful about the order of integration. Writing out the integral (13.176) again in spherical coordinates, we have

$$\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{\hbar c}{8\pi^2}\int_0^1 d\xi\,r_{\mathrm{TE}}(\xi)\int_0^\infty dk\,k^3\cos(2kz\xi). \tag{13.182}$$

The problem here is that while $r_{\mathrm{TE}}(\xi)$ is a well-behaved function, it is complex enough that it is difficult to perform the integration with the cos factor. And we can simplify the expression by using the lowest-order expression in χ from Eqs. (13.169), but then the integral diverges at $\xi = 0$. We will handle this difficulty by

first carrying out the k integral by parts to build up powers of ξ that will regularize the integral at $\xi = 0$. Performing the k integral by parts four times, we have

$$\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{\hbar c}{8\pi^2}\frac{(2z)^4}{840}\int_0^1 d\xi\, r_{\mathrm{TE}}(\xi)\,\xi^4 \int_0^\infty dk\, k^7 \cos(2kz\xi). \tag{13.183}$$

We have ignored boundary terms here. At $k = 0$, the integrand always involves a positive power of k, so this is obvious. At $k = \infty$, we are again forcing the integral to converge by cutting it off with an exponential convergence factor, or equivalently by regarding the cosine integration as being shifted in the complex plane off of the real axis as

$$\cos k = \frac{1}{2}\left(e^{i(k+i0^+)} + e^{-i(k-i0^+)}\right), \tag{13.184}$$

so that the value at infinity is suppressed without affecting the value of the integral. We then proceed by using Eqs. (13.169) to set $r_{\mathrm{TE}} = -1/4\xi^2$, and proceed as in the perfect-conductor case,

$$\begin{aligned}
\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) &= -\frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^\infty dk\, k^7 \int_0^1 d\xi\, \xi^2 \cos(2kz\xi) \\
&= -\frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\, k^7 \int_0^1 d\xi\, \xi^2 \cos(k\xi) \\
&= -\frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\, k^7 \left(\frac{2k\cos k + (k^2-2)\sin k}{k^3}\right) \\
&= -\frac{\hbar c}{2^{12}105\pi^2 z^4}\chi(-1008) \\
&= \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{\chi}{40}\right),
\end{aligned} \tag{13.185}$$

where we changed variables $k \longrightarrow k/2z$. The final result is

$$\boxed{\mathscr{E}_{E\,\mathrm{TE}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{\chi}{40}\right).}$$

(TE electric energy density, rarefied dielectric) (13.186)

Note that this is equivalent to the energy density (13.181) for the perfect conductor, up to a factor $6\chi/40$.

13.13.2 TM Energy Density

Now we can repeat the calculation for the TM modes. The polarization of the modes here works out to be slightly more complicated, because the polarization is not parallel to the surface, but only the parallel component is continuous across the dielectric interface. The resulting mode is[25]

$$\mathbf{f}^{\downarrow}_{\mathbf{k},\mathrm{TM}}(\mathbf{r}) = \hat{\varepsilon}_{\mathbf{k},\mathrm{TM}} e^{i\mathbf{k}\cdot\mathbf{r}}\Theta(z) + \hat{\varepsilon}^{-}_{\mathbf{k},\mathrm{TM}} r_{\mathrm{TM}}(\theta) e^{i\mathbf{k}^{-}\cdot\mathbf{r}}\Theta(z) + \hat{\varepsilon}^{\mathrm{t}}_{\mathbf{k},\mathrm{TM}}\sqrt{1-|r_{\mathrm{TM}}|^2}\, e^{i\mathbf{k}_{\mathrm{t}}\cdot\mathbf{r}}\Theta(-z), \tag{13.187}$$

where the reflected polarization vector $\hat{\varepsilon}^{-}_{\mathbf{k},\mathrm{TM}}$ is the same as the incident vector $\hat{\varepsilon}_{\mathbf{k},\mathrm{TM}}$, but with the z component reversed, and $\hat{\varepsilon}^{\mathrm{t}}_{\mathbf{k},\mathrm{TM}}$ is the polarization vector of the transmitted wave, whose precise form we will not need. However, note that both polarization vectors are orthogonal to $\mathbf{k}$ and each other, so that

$$\hat{\varepsilon}_{\mathbf{k},\mathrm{TM}} = \hat{k}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}, \qquad \hat{\varepsilon}^{-}_{\mathbf{k},\mathrm{TM}} = \hat{k}^{-}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}, \qquad \hat{\varepsilon}^{\mathrm{t}}_{\mathbf{k},\mathrm{TM}} = \hat{k}_{\mathrm{t}}\times\hat{\varepsilon}_{\mathbf{k},\mathrm{TE}}, \tag{13.188}$$

where $\hat{k}^{-}$ also has its z-component reversed compared to $\hat{k}$. Considering only the $z > 0$ region, the only z-dependent part of the mode envelope that we will need is the cross-term between the incident and reflected fields:

$$|\mathbf{f}^{\downarrow}_{\mathbf{k},\mathrm{TM}}(\mathbf{r})|^2 = 2r_{\mathrm{TE}}(\theta)\cos 2\theta\, \cos 2k_z z, \tag{13.189}$$

[25] Daniel A. Steck, *op. cit.*

where the angular factor comes from the inner product of the incident and reflected polarization vectors. As in the TE case, the modes incident from the dielectric side do not contribute any z-dependent terms. Thus, the TM contribution to the electric energy density (13.168) is

$$\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) = \frac{\hbar c}{2(2\pi)^3}\int_{k_z<0} d^3k\, k r_{\mathrm{TM}}(\theta)\left[2\cos^2\theta - 1\right]\cos(2k_z z), \tag{13.190}$$

after renormalizing away the vacuum contribution, changing the mode sum to an integral, and using the double-angle formula $\cos 2\theta = 2\cos^2\theta - 1$.

We will adopt the same method of evaluation as in the TE case. Writing out the integral in spherical coordinates,

$$\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) = \frac{\hbar c}{8\pi^2}\int_0^1 d\xi\,(2\xi^2-1)\, r_{\mathrm{TM}}(\xi)\int_0^\infty dk\, k^3\cos(2kz\xi), \tag{13.191}$$

where again $\xi = \cos\theta$. Integrating by parts in k four times,

$$\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) = \frac{\hbar c}{8\pi^2}\frac{(2z)^4}{840}\int_0^1 d\xi\,\xi^4(2\xi^2-1)\, r_{\mathrm{TM}}(\xi)\int_0^\infty dk\, k^7\cos(2kz\xi), \tag{13.192}$$

where we dropped boundary terms as before. Then using Eqs. (13.169) to set $r_{\mathrm{TM}} = (1/\xi^2 - 2)/4$, and proceeding as in the perfect-conductor case,

$$\begin{aligned}
\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) &= \frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^1 d\xi\,\xi^4(2\xi^2-1)\left(\frac{1}{\xi^2}-2\right)\int_0^\infty dk\, k^7\cos(2kz\xi)\\
&= \frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^\infty dk\, k^7\int_0^1 d\xi\left(-\xi^2+4\xi^4-4\xi^6\right)\cos(2kz\xi)\\
&= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\, k^7\int_0^1 d\xi\left(-\xi^2+4\xi^4-4\xi^6\right)\cos(k\xi),
\end{aligned} \tag{13.193}$$

where we have again changed variables $k \longrightarrow k/2z$. Carrying out the ξ integral,

$$\begin{aligned}
\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) &= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\left(-2880+1344k^2-74k^4+k^6\right)\sin k - 2k\left(1440-192k^2+5k^4\right)\cos k\\
&= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi(14448)\\
&= \frac{43\hbar c}{1280\pi^2 z^4}\chi,
\end{aligned} \tag{13.194}$$

where we have again used a convergence factor to perform the k integration.

$$\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{43\chi}{120}\right), \tag{13.195}$$

(TM electric energy density, rarefied dielectric)

Note that this is equivalent to the TE energy density (13.186) up to an overall constant.

13.13.2.1 Digression: TM Energy Density for the Perfect Conductor

For completeness, we can also compute the TM energy for the perfect-conductor case. Putting $r_{\text{TM}} = -1$ into Eq. (13.191), we have

$$\begin{aligned}
\mathscr{E}_{E_{\text{TM}}}(\mathbf{r}) &= -\frac{\hbar c}{8\pi^2}\int_0^\infty dk\, k^3 \int_0^1 d\xi\, (2\xi^2 - 1)\cos(2kz\xi) \\
&= -\frac{\hbar c}{8\pi^2 (2z)^4}\int_0^\infty dk\, k^3 \int_0^1 d\xi\, (2\xi^2 - 1)\cos(k\xi) \\
&= -\frac{\hbar c}{8\pi^2 (2z)^4}\int_0^\infty dk\, \left[4k\cos k + (k^2 - 4)\sin k\right] \\
&= -\frac{\hbar c}{8\pi^2 (2z)^4}(-10),
\end{aligned} \tag{13.196}$$

with the same procedure as for the dielectric. Finally, the result is

$$\mathscr{E}_{E_{\text{TM}}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{5}{6}\right). \tag{13.197}$$

(TM electric energy density, perfect conductor)

This is the same as the TE energy density for the perfect conductor (13.181), but with an extra factor of 5.

13.13.3 Total Casimir–Polder Potential for a Polarizable Atom

Summing the two polarization contributions (13.186) and (13.195) for a rarefied dielectric, the total electric-field energy density

$$\mathscr{E}_E(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{23\chi}{60}\right). \tag{13.198}$$

(electric energy density, rarefied dielectric)

Notice that the contribution of the TM mode to the energy density is much larger than the contribution of the TE mode (43/120 vs. 1/40). Similarly, summing the two contributions (13.181) and (13.197) for the perfect conductor gives

$$\mathscr{E}_E(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}. \tag{13.199}$$

(electric energy density, perfect conductor)

Note that we have written all of the energy densities as a fraction of this one, which is the largest of all the cases. Again, the contribution of the TM mode to the energy density is much larger than the contribution of the TE mode (5/6 vs. 1/6). This may be somewhat surprising, as in the dielectric case, $|r_{\text{TM}}| < |r_{\text{TE}}|$ (recall that TM light is transmitted perfectly at Brewster's angle, but not TE light), and in both cases, the interference is imperfect due to polarization mismatching in the TM case.

To relate the electric energy density to the Casimir–Polder potential for an atom, recall [see Eq. (1.60)] that the dipole potential for an atom interacting via its induced dipole moment with an electric field has the form $V_{\text{dipole}} = -(1/2)\alpha_0 E^2$, where α_0 is the dc polarizability. As noted above, we are ignoring dispersion here, so we are implicitly assuming that the atomic polarizability satisfies $\alpha(\omega) = \alpha_0$—this is a far-field limit, where only the low-frequency modes contribute to the potential (the equivalent of the limit of large atomic resonance frequencies). There is no dc field, but we have exactly computed the vacuum expectation value of E^2 (with a factor of $\epsilon_0/2$), so we wish to consider

$$V_{\text{dipole}} = -\frac{1}{2}\alpha_0 \langle E^2 \rangle. \tag{13.200}$$

Combining this with Eq. (13.166), we can relate the atomic potential directly to the energy density via

$$V_{\text{dipole}} = -\frac{\alpha_0}{\epsilon_0}\mathscr{E}_E(\mathbf{r}). \tag{13.201}$$

(far-field Casimir–Polder potential)

Using this result, the perfect-conductor Casimir–Polder potential becomes

$$V_{\text{dipole}} = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}$$

(Casimir–Polder potential, perfect conductor) (13.202)

in the far field, which agrees with Eq. (13.60). For the rarified dielectric, the Casimir–Polder potential is

$$V_{\text{dipole}} = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{23\chi}{60}\right),$$

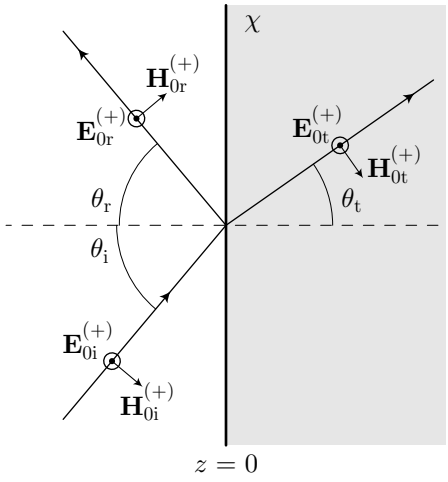
(Casimir–Polder potential, rarefied dielectric) (13.203)

or a factor of $23\chi/60$ times the perfect-conductor result, valid for $\chi \ll 1$. The results here agree with the limiting cases of the general treatment in Section 14.3.5.6.

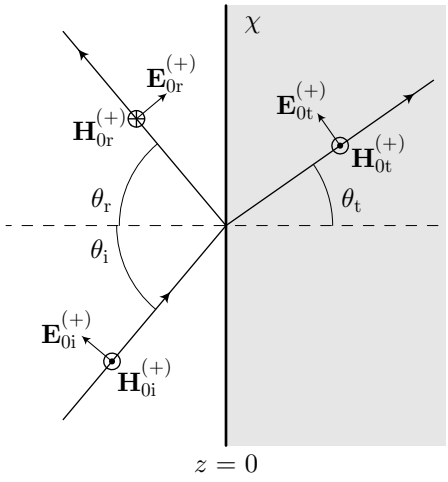
13.13.4 Magnetic-Field Energies

While the magnetic fields do not contribute to the Casimir–Polder potential of an electrically polarizable atom, it is nonetheless interesting to compute the magnetic-field energies, and sum these to find the total energies. While the electric and magnetic field energies are equivalent in vacuum, the changes in the energy densities due to the dielectric are not the same. The derivation will be approximately the same as for the electric-field case, but with two modifications, due to the magnetic fields are orthogonal to both the electric fields and wave vector (i.e., the magnetic-field modes involve $\nabla \times \mathbf{f}$ rather than $\mathbf{f}$. The first modification is that the interference terms from the incident and reflected fields in both polarizations gain a relative minus sign compared to the electric-field case, because of this orthogonality condition (in particular, $\mathbf{E} \times \mathbf{B}$ must point in the direction of $\mathbf{k}$. So for example, for normal incidence, if the incident and reflected waves are assumed to have parallel (electric-field) polarizations, the magnetic-field vectors will be antiparallel. The second modification is that the geometry factor of $\cos 2\theta$ in the TM expression (13.189), which later becomes a factor of $2\xi^2 - 1$, moves from the TM case to the TE case for the magnetic fields.

These considerations are best visualized by comparing a diagram of the electric and magnetic fields at a dielectric interface for TE polarization,



and for TM polarization,



noting the relative orientations of the incident and reflected vectors in each case.

13.13.4.1 Magnetic TE Energy Densities

Starting with the expression (13.182) and making the sign and geometric-factor changes, we have

$$\mathscr{E}_{B\,\mathrm{TE}}(\mathbf{r}) = -\frac{\hbar c}{8\pi^2}\int_0^1 d\xi\,(2\xi^2-1)\,r_{\mathrm{TE}}(\xi)\int_0^\infty dk\,k^3\cos(2kz\xi). \tag{13.204}$$

The procedure to evaluate this is to integrate by parts four times,

$$\mathscr{E}_{B\,\mathrm{TE}}(\mathbf{r}) = -\frac{\hbar c}{8\pi^2}\frac{(2z)^4}{840}\int_0^1 d\xi\,\xi^4(2\xi^2-1)\,r_{\mathrm{TE}}(\xi)\int_0^\infty dk\,k^7\cos(2kz\xi), \tag{13.205}$$

where we dropped boundary terms as before. Then using Eqs. (13.169) to set $r_{\mathrm{TE}} = -1/4\xi^2$, and proceeding as in the perfect-conductor case,

$$\begin{aligned}
\mathscr{E}_{B\,\mathrm{TE}}(\mathbf{r}) &= \frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^1 d\xi\,\xi^4(2\xi^2-1)\left(\frac{1}{\xi^2}\right)\int_0^\infty dk\,k^7\cos(2kz\xi)\\
&= \frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^\infty dk\,k^7\int_0^1 d\xi\,\xi^2\left(2\xi^2-1\right)\cos(2kz\xi)\\
&= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\,k^7\int_0^1 d\xi\,\xi^2\left(2\xi^2-1\right)\cos(k\xi)\\
&= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\,k^2\left[6k(k^2-8)\cos k+(48-22k^2+k^4)\sin k\right]\\
&= \frac{\hbar c}{2^{12}105\pi^2 z^4}\chi(-2352),
\end{aligned} \tag{13.206}$$

or simplifying the result,

$$\mathscr{E}_{B\,\mathrm{TE}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(-\frac{7\chi}{120}\right).$$

(TE magnetic energy density, rarefied dielectric) (13.207)

For the perfect conductor case, we instead set $r_{\mathrm{TE}} = -1$. This is the same as the TM electric energy density for the perfect conductor, so Eq. (13.197) can be adapted here as

$$\mathscr{E}_{E\,\mathrm{TM}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(-\frac{5}{6}\right),$$

(TE magnetic energy density, perfect conductor) (13.208)

where we have only had to change the overall minus sign.

13.13.4.2 Magnetic TM Energy Densities

For the TM polarization, we take the electric TE expression, remove the geometric factor, and put in an overall minus sign, so Eq. (13.191) becomes

$$\mathscr{E}_{B\,\mathrm{TM}}(\mathbf{r}) = -\frac{\hbar c}{8\pi^2}\int_0^1 d\xi\, r_{\mathrm{TM}}(\xi)\int_0^\infty dk\, k^3\cos(2kz\xi), \tag{13.209}$$

Then using Eqs. (13.169) to set $r_{\mathrm{TM}} = (1/\xi^2 - 2)/4$, and proceeding as usual to integrate by parts.

$$\begin{aligned}
\mathscr{E}_{B\,\mathrm{TM}}(\mathbf{r}) &= -\frac{\hbar c}{8\pi^2}\frac{(2z)^4}{840}\int_0^1 d\xi\,\xi^4\, r_{\mathrm{TM}}(\xi)\int_0^\infty dk\, k^7\cos(2kz\xi)\\
&= -\frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^1 d\xi\,\xi^4\left(\frac{1}{\xi^2}-2\right)\int_0^\infty dk\, k^7\cos(2kz\xi)\\
&= -\frac{\hbar c}{8\pi^2}\frac{z^4}{210}\chi\int_0^\infty dk\, k^7\int_0^1 d\xi\,\xi^2\left(1-2\xi^2\right)\cos(2kz\xi)\\
&= -\frac{\hbar c}{2^{12}105\pi^2 z^4}\chi\int_0^\infty dk\, k^7\int_0^1 d\xi\,\xi^2\left(1-2\xi^2\right)\cos(k\xi).
\end{aligned} \tag{13.210}$$

Remarkably, this is exactly the same integral as in the TE case, essentially because $r_{\mathrm{TM}} = r_{\mathrm{TE}}(2\xi^2 - 1)$ for small χ. Thus, we have the same contribution as in Eq. (13.207).

$$\mathscr{E}_{B\,\mathrm{TM}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(-\frac{7\chi}{120}\right).$$

(TM magnetic energy density, rarefied dielectric) (13.211)

For the perfect conductor case, we again instead set $r_{\mathrm{TM}} = -1$. This is the same as the TE electric energy density for the perfect conductor, so Eq. (13.181) can be adapted here as

$$\mathscr{E}_{B\,\mathrm{TM}}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(-\frac{1}{6}\right),$$

(TM magnetic energy density, perfect conductor) (13.212)

where we have again only changed the overall minus sign.

13.13.4.3 Total Magnetic and Electromagnetic Energy Densities

The total magnetic energy density for the rarefied dielectric is the sum of Eqs. (13.207) and (13.211) or

$$\mathscr{E}_{B}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(-\frac{7\chi}{60}\right).$$

(magnetic energy density, rarefied dielectric) (13.213)

Then in the case of the rarefied dielectric, the total electromagnetic energy is given by the sum of Eqs. (13.198) and (13.213), which is

$$\mathscr{E}(\mathbf{r}) = \frac{3\hbar c}{32\pi^2 z^4}\left(\frac{4\chi}{15}\right).$$

(total electromagnetic energy density, rarefied dielectric) (13.214)

Thus, the magnetic energy density cancels some of the electric energy density (after renormalization against vacuum).

For the perfectly conducting plane, the total magnetic energy density is the sum of Eqs. (13.208) and (13.211)

$$\mathscr{E}_{B}(\mathbf{r}) = -\frac{3\hbar c}{32\pi^2 z^4},$$

(magnetic energy density, perfect conductor) (13.215)

which is exactly the opposite of the electric-field contribution (13.199), so that the total energy density is

$$\mathscr{E}(\mathbf{r}) = 0.$$

(total electromagnetic energy density, perfect conductor) (13.216)

Remarkably, the change in the total energy density vanishes for the perfect conductor. While the dielectric boundary increases the fluctuations (and thus the energy) of the electric field, it tends to suppress those of the magnetic field, at least outside the dielectric. This means that a hypothetical atom that was not electrically polarizable, but purely magnetic, would experience a *repulsive* Casimir–Polder force near a dielectric boundary, at least in the limits of a weak and strong dielectric.

13.14 Exercises

Problem 13.1

The sine integral is defined as

$$\mathrm{Si}(x) := \frac{\pi}{2} - \int_x^\infty \frac{\sin t}{t}\, dt, \tag{13.217}$$

and the cosine integral is defined as

$$\mathrm{Ci}(x) := -\int_x^\infty \frac{\cos t}{t}\, dt. \tag{13.218}$$

The auxiliary functions are also defined as

$$\begin{aligned} f(z) &= \sin z\, \mathrm{Ci}(z) + \cos z \left[\frac{\pi}{2} - \mathrm{Si}(z)\right] \\ g(z) &= -\cos z\, \mathrm{Ci}(z) + \sin z \left[\frac{\pi}{2} - \mathrm{Si}(z)\right]. \end{aligned} \tag{13.219}$$

(a) Show that

$$\begin{aligned} f(z) &= \int_0^\infty \frac{\sin t}{t+z}\, dt \qquad (|\arg z| < \pi) \\ g(z) &= \int_0^\infty \frac{\cos t}{t+z}\, dt \qquad (|\arg z| < \pi). \end{aligned} \tag{13.220}$$

(b) Show that

$$\begin{aligned} f(z) &= \int_0^\infty \frac{e^{-zt}}{1+t^2}\, dt \qquad (\mathrm{Re}[z] > 0) \\ g(z) &= \int_0^\infty \frac{t\, e^{-zt}}{1+t^2}\, dt \qquad (\mathrm{Re}[z] > 0). \end{aligned} \tag{13.221}$$

You can do this by considering the combination $g(z) + if(z)$, writing out its integral expression, and then changing to an integral along the positive imaginary axis.

Note that simple changes of variable justify the more general integral formulae[26]

$$\begin{aligned} \int_0^\infty \frac{\sin(ax)}{x+\beta}\, dx &= f(a\beta) \qquad (|\arg\beta| < \pi,\ a > 0) \\ \int_0^\infty \frac{\cos(ax)}{x+\beta}\, dx &= g(a\beta) \qquad (|\arg\beta| < \pi,\ a > 0) \\ \int_0^\infty \frac{e^{-\mu x}}{\beta^2+x^2}\, dx &= \frac{f(\beta\mu)}{\beta} \qquad (\mathrm{Re}[\mu] > 0, \mathrm{Re}[\beta] > 0) \\ \int_0^\infty \frac{x\, e^{-\mu x}}{\beta^2+x^2}\, dx &= g(\beta\mu) \qquad (\mathrm{Re}[\mu] > 0, \mathrm{Re}[\beta] > 0) \end{aligned} \tag{13.222}$$

that are useful for evaluating the Casimir–Polder potential near perfectly conducting planes.

Problem 13.2

Prove the "branch-cut formulae" for the cosine integral,

$$\begin{aligned} \mathrm{Ci}(-z) &= \mathrm{Ci}(z) - i\pi \qquad (0 < \arg z < \pi) \\ \mathrm{Ci}(-z) &= \mathrm{Ci}(z) + i\pi \qquad (-\pi < \arg z < 0), \end{aligned} \tag{13.223}$$

[26] See I. S. Gradstein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 6th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2000), Integrals 3.722.1, 3.722.3, 3.354.1, and 3.354.2.

where recall that we define the cosine integral by

$$\mathrm{Ci}(x) := -\int_x^\infty \frac{\cos t}{t}\,dt. \tag{13.224}$$

You can do this as follows.

(a) Defining the exponential integral

$$E_1(x) := \int_x^\infty \frac{e^{-t}}{t}\,dt, \tag{13.225}$$

show that

$$E_1(x) = -\gamma - \log x - \sum_{j=1}^{\infty} \frac{(-1)^j x^j}{jj!}, \tag{13.226}$$

where $\gamma \approx 0.577\,215\,664\,901\,532\,860\,607$ is the Euler–Mascheroni constant, defined to be the asymptotic difference between the partial sums of the harmonic series and the logarithmic function:

$$\gamma := \lim_{n\to\infty} \left[1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log n\right]. \tag{13.227}$$

Hint: you may find it helpful to use the integral relation

$$\gamma = -\int_0^\infty e^{-t} \log t\,dt, \tag{13.228}$$

but if you use it, you should *prove* it. Whether or not you use it, it will help to get started by proving that

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \int_0^1 \frac{1-(1-t)^n}{t}\,dt. \tag{13.229}$$

You should prove this by *induction* (if you don't know what that means, then make sure you find out). Then recalling that the exponential function may be defined by

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n, \tag{13.230}$$

you can, for example, establish Eq. (13.228). In general, you should be integrating by parts like crazy in this problem.

(b) Let $t \longrightarrow xt$ in Eq. (13.225) to remove x from the integration limit, and then show that

$$\mathrm{Ci}(x) = -\frac{1}{2}\Big[E_1(ix) + E_1(-ix)\Big]. \tag{13.231}$$

(c) Use the result of (a) to show that

$$\mathrm{Ci}(x) = \gamma + \log x + \int_0^x \frac{1-\cos t}{t}\,dt. \tag{13.232}$$

(d) Now recall that the log function has a branch cut along the negative real axis, and is otherwise defined by $\log z = \log r + i\theta$ for $z = re^{i\theta}$ ($r > 0$, $-\pi < \theta < \pi$). Use this property of the logarithm to prove Eqs. (13.223).

Problem 13.3

Show that the integrals

$$\begin{aligned} I_1 &= \int d^3k \, \frac{k}{(k+k_0)} \cos(2k_z z) \\ I_2 &= \int d^3k \, \frac{k_z^2}{k(k+k_0)} \cos(2k_z z). \end{aligned} \tag{13.233}$$

can also be performed in cylindrical coordinates, and agree with the results from spherical-coordinate integration in the region $z > 0$. Be careful to avoid infinities!

Problem 13.4

Show that the integrals

$$\begin{aligned} I_1' &= k_0^2 \int d^3k \, \frac{1}{k(k+k_0)} \cos(2k_z z) \\ I_2' &= k_0^2 \int d^3k \, \frac{k_z^2}{k^3(k+k_0)} \cos(2k_z z). \end{aligned} \tag{13.234}$$

can also be performed in cylindrical coordinates, and agree with the results from spherical-coordinate integration in the region $z > 0$.

Problem 13.5

(a) Show that expression (13.46) for the Casimir–Polder potential

$$V_{\text{CP}} = \frac{1}{(4\pi\epsilon_0)8\pi} \sum_j \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2 \right) \frac{4k_{j0}^2}{z} - \left(d_{j,\parallel}^2/2 + d_{j,z}^2 \right) \left(\partial_z^2 \frac{1}{z} \right) \right] f(2k_{j0}z) \tag{13.235}$$

reduces in the long-distance limit to

$$V_{\text{CP}} = -\sum_j \frac{c}{(4\pi\epsilon_0)4\pi\omega_{j0}} \left(d_{j,\parallel}^2 + d_{j,z}^2 \right) \frac{1}{z^4} = -\sum_j \frac{c d_j^2}{(4\pi\epsilon_0)4\pi\omega_{j0} z^4}, \tag{13.236}$$

so that even for an anisotropic molecule, the far-field potential is independent of the molecular orientation and the potential still scales as z^{-4}.

(b) Defining the normal and parallel static polarizabilities by

$$\begin{aligned} \alpha_{z0} &:= \alpha_z(0) = \sum_j \frac{2d_{j,z}^2}{\hbar\omega_{j0}} \\ \alpha_{\parallel 0} &:= \alpha_\parallel(0) = \sum_j \frac{d_{j,\parallel}^2}{\hbar\omega_{j0}}, \end{aligned} \tag{13.237}$$

show that the far-field Casimir–Polder potential can then be written

$$V_{\text{CP}} = -\frac{\hbar c}{(4\pi\epsilon_0)8\pi} \left(\alpha_{z0} + 2\alpha_{\parallel 0} \right) \frac{1}{z^4}. \tag{13.238}$$

Justify the above definitions of the static polarizability on the basis of the usual expression (13.61). Also show that for an isotropic atom, the Casimir–Polder potential here reduces to the usual far-field formula, Eq. (13.60).

(c) Argue that the anisotropic correction—the term involving $(d_{j,\parallel}^2/2 - d_{j,z}^2)$—to the near-field potential (van der Waals regime) should be negligible compared to the usual component for the spherically symmetric part of the atom.

Problem 13.6

Starting with the integral expression (13.23) for the Casimir–Polder potential,

$$V_{\text{CP}} = \frac{1}{16\pi^3\epsilon_0}\sum_j \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \int_{-\infty}^{\infty} dk_z \frac{k}{(k_{j0}+k)} \left[\left(d_{j,\parallel}^2/2 - d_{j,z}^2\right) + \frac{k_z^2}{k^2}\left(d_{j,\parallel}^2/2 + d_{j,z}^2\right)\right] \cos(2k_z z), \tag{13.239}$$

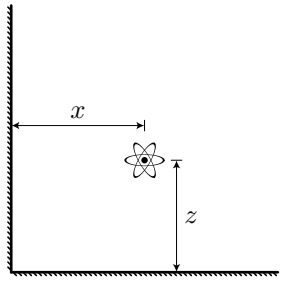
obtain *directly* the asymptotic form for large z for a spherically symmetric atom (that is, do not use the full expression in terms of the auxiliary function $f(z)$ as an intermediate step). To do this, note that for large z, $\cos(2k_z z)$ is a rapidly oscillating function of k_z (and thus of k), and thus the integrand only gives a nonvanishing contribution where k_z (and thus k) is close to zero.

Problem 13.7

Work through the entire derivation of the Casimir–Polder potential near a planar conductor, but with the following simplifications, which you should implement as early as possible in the derivation: assume a spherically symmetric atom from the beginning, and work in the far field, where only low frequencies are important (see the comment in Problem 13.6). Also, start with the mode functions inside a rectangular cavity.

Problem 13.8

Compute the Casimir–Polder potential for a spherically symmetric atom near an L-shaped conductor, where the conductor occupies the region $x < 0$ and $z < 0$, with the atom at $(x > 0, y, z > 0)$. Use the setup of Problem 13.7, including the assumption of being in the far-field regime.



In particular, you should find a potential of the form

$$V_{\text{CP}} = V_{\text{CP}}^{(x)} + V_{\text{CP}}^{(z)} + V_{\text{CP}}^{(xz)}, \tag{13.240}$$

where $V_{\text{CP}}^{(x)}$ and $V_{\text{CP}}^{(z)}$ have the form for atom-plane Casimir–Polder energies, and $V_{\text{CP}}^{(xz)}$ is a nonadditive correction for the atom interacting with both surfaces. You should be able to give a simple interpretation for $V_{\text{CP}}^{(xz)}$ once simplified (in fact it is the Casimir–Polder potential for an atom–surface distance $\sqrt{x^2+z^2}$.

Problem 13.9

Work out an integral expression for the Casimir–Polder potential, in analogy to Eq. (13.23), for the electromagnetic field in *two* dimensions (that is, two-dimensional space with a line conductor, which of course does not correspond to reality; instead, consider the physical problem of an atom constrained to be halfway between two planar, parallel conductors, spaced by the very small distance ℓ, and the atom is a distance z from a third planar conductor that intersects the other two conductors perpendicularly). For simplicity, just work out the case of a spherically symmetric atom. Show that the potential scales asymptotically as z^{-3} at *large* z.

Problem 13.10

Recompute the Casimir–Polder force for a spherically symmetric atom near a perfectly conducting

plane, but using only the TE modes [Eq. (8.76)]

$$\mathbf{f}_{\mathbf{k}}^{(\mathrm{TE})}(\mathbf{r}) = \sqrt{\frac{2}{V}}\left(\hat{k}_{\parallel} \times \hat{z} \sin k_z z\right) e^{i\mathbf{k}_{\parallel}\cdot\mathbf{r}} \tag{13.241}$$

(i.e., ignoring the contribution of the TM modes). Show that in the far-field regime, the TE modes contribute only 1/6 of the total potential.

Problem 13.11

Derive an expression for the ground-state Casimir–Polder potential of an atom near a perfectly conducting, infinite surface, generalized to the case where the electromagnetic field is at temperature T.

Chapter 14

QED with Dielectric Media

14.1 Classical Electrodynamics in Dielectric Media

Our starting point will be Maxwell's equations for the electromagnetic fields in a medium:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{H} &= \partial_t \mathbf{D} + \mathbf{j}. \end{aligned} \tag{14.1}$$

(Maxwell's equations)

Here, $\mathbf{D}$ is the **electric flux density** or **electric displacement**, $\mathbf{B}$ is the **magnetic flux density**, $\mathbf{E}$ and $\mathbf{H}$ are the usual electric and magnetic fields, respectively, ρ is the source charge density, and $\mathbf{j}$ is the source current density. We will ignore magnetic effects, so that

$$\mathbf{B} = \mu_0 \mathbf{H}, \tag{14.2}$$

and the electric fields are related by

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}, \tag{14.3}$$

where $\mathbf{P}$ is the polarization density of the medium (i.e., dipole moment per unit volume).

14.1.1 Effective Sources

There are multiple ways to treat the polarization field here. One is to treat it in terms of equivalent, effective sources. To see this, we can put Eq. (14.3) into the first Maxwell equation, with the result

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \left(\rho - \nabla \cdot \mathbf{P} \right). \tag{14.4}$$

Defining the effective ''bound charge density'' for the polarization field by

$$\rho_{\text{P}} := -\nabla \cdot \mathbf{P}, \tag{14.5}$$

(bound charge density)

we see that the first Maxwell equation becomes

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \left(\rho + \rho_{\text{P}} \right). \tag{14.6}$$

This is the form for the free-space Maxwell equation, with an extra source charge density.

Similarly, from the continuity equation

$$\nabla \cdot \mathbf{j} = -\partial_t \rho, \tag{14.7}$$

the existence of the bound charge density implies a ''bound current density:''

$$\nabla \cdot (\partial_t \mathbf{P}) = -\partial_t \rho_{\text{P}} = \nabla \cdot \mathbf{j}_{\text{P}}. \tag{14.8}$$

Thus, we can define

$$\mathbf{j}_{\text{P}} := \partial_t \mathbf{P} \tag{14.9}$$
(bound polarization current density)

as the effective polarization current density. We then see that the last Maxwell equation,

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{j}, \tag{14.10}$$

can be written

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \partial_t \mathbf{E} + \mu_0 (\mathbf{j} + \mathbf{j}_{\text{P}}). \tag{14.11}$$

This is the corresponding free-space Maxwell equation, with an extra source current density $\mathbf{j}_{\text{P}}$. If we were not ignoring magnetic media, there would also be a second bound current associated with the magnetization.

These effective sources are useful ways to think about the medium response. In particular, one route to quantizing the field in absorptive media is to realize that in quantum mechanics, absorption (dissipation) is always accompanied by noise (fluctuations) to preserve commutation relations at all times. The noise can be explicitly put in via noise source fields. However, for now, we will specialize to a linear, dispersive medium, and instead treat the medium response by frequency-dependent response functions.

14.1.2 Linear, Dispersive Media

14.1.2.1 Frequency Domain

To continue, we will consider the Fourier transforms of the Maxwell equations (14.1)

$$\begin{aligned}
\nabla \cdot \mathbf{D}(\mathbf{r}, \omega) &= \rho(\mathbf{r}, \omega) \\
\nabla \cdot \mathbf{B}(\mathbf{r}, \omega) &= 0 \\
\nabla \times \mathbf{E}(\mathbf{r}, \omega) &= i\omega \mathbf{B}(\mathbf{r}, \omega) \\
\nabla \times \mathbf{H}(\mathbf{r}, \omega) &= -i\omega \mathbf{D}(\mathbf{r}, \omega) + \mathbf{j}(\mathbf{r}, \omega).
\end{aligned} \tag{14.12}$$
(Maxwell's equations, frequency domain)

We are now explicitly marking the dependence of each field on space and frequency, with frequency ω of course representing an implicit time dependence of the form $e^{-i\omega t}$. Note that in our notation here, we are writing fields such as $\mathbf{E}(t)$ and $\mathbf{E}(\omega)$ as different functions comprising a Fourier-transform pair:

$$\begin{aligned}
\mathbf{E}(\mathbf{r}, \omega) &= \int_{-\infty}^{\infty} dt\, \mathbf{E}(\mathbf{r}, t)\, e^{i\omega t} \\
\mathbf{E}(\mathbf{r}, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \mathbf{E}(\mathbf{r}, \omega)\, e^{-i\omega t}.
\end{aligned} \tag{14.13}$$

Again ignoring magnetic effects, we can write

$$\mathbf{B}(\mathbf{r}, \omega) = \mu_0 \mathbf{H}(\mathbf{r}, \omega), \tag{14.14}$$

while the electric fields are still related by

$$\mathbf{D}(\mathbf{r}, \omega) = \epsilon_0 \mathbf{E}(\mathbf{r}, \omega) + \mathbf{P}(\mathbf{r}, \omega). \tag{14.15}$$

A **linear, dispersive medium** is defined such that the medium polarization is given by the relation

$$\mathbf{P}(\mathbf{r},\omega) = \epsilon_0 \chi(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega), \tag{14.16}$$

where $\chi(\mathbf{r},\omega)$ is the dimensionless, linear, frequency dependent **susceptibility** of the medium. In this context, ''linear'' means that the susceptibility is independent of the electric-field amplitude, and ''dispersive'' means that the polarization response is a simple proportionality at each frequency. Thus, the electric fields are related by

$$\mathbf{D}(\mathbf{r},\omega) = \epsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega), \tag{14.17}$$

where

$$\epsilon(\mathbf{r},\omega) := \epsilon_0 \left[1 + \chi(\mathbf{r},\omega)\right], \tag{14.18}$$

is the (linear) dielectric permittivity (the dimensionless ratio ϵ/ϵ_0 is the **dielectric constant**).

14.1.2.2 Time Domain

The description of dispersive media in the time domain is more complicated. According to the convolution theorem (Section 17.1.2), the time-domain version of Eq. (14.16) is the convolution

$$\mathbf{P}(\mathbf{r},t) = \epsilon_0 \int_0^\infty dt'\, g_\chi(\mathbf{r},t')\, \mathbf{E}(\mathbf{r},t-t'), \tag{14.19}$$

where the susceptibility $\chi(\mathbf{r},\omega)$ is the Fourier transform of the correlation function $g_\chi(\mathbf{r},t)$:

$$\chi(\mathbf{r},\omega) = \int_0^\infty dt\, g_\chi(\mathbf{r},t)\, e^{i\omega t}. \tag{14.20}$$

Note that in writing down Eqs. (14.19) and (14.20), we are only integrating over positive times, where normally the integration should extend over the entire real axis. We do this to avoid an unphysical feature: otherwise, the *medium response* $\mathbf{P}(\mathbf{r},t)$ would depend on the *input field* $\mathbf{E}(\mathbf{r},t')$, even for *future* times $t' > t$. It is thus physically reasonable to impose the *causality* requirement that the polarization field only depends on the electric field in the present or in the past. We can also just as well write the causality requirement as

$$g_\chi(\mathbf{r},t) = g_\chi(\mathbf{r},t)\,\Theta(t), \tag{14.21}$$

where $\Theta(t)$ is the Heaviside step function. Based on these relations, we see that the electric fields are related by

$$\mathbf{D}(\mathbf{r},t) = \epsilon_0 \mathbf{E}(\mathbf{r},t) + \epsilon_0 \int_0^\infty dt'\, g_\chi(\mathbf{r},t')\, \mathbf{E}(\mathbf{r},t-t'). \tag{14.22}$$

Of course, we could also write this relation down in terms of a permittivity kernel

$$\epsilon(\mathbf{r},t) = \epsilon_0 \delta(t) + \epsilon_0 g_\chi(\mathbf{r},t'), \tag{14.23}$$

so that

$$\mathbf{D}(\mathbf{r},t) = \int_{0^-}^\infty dt'\, \epsilon(\mathbf{r},t')\, \mathbf{E}(\mathbf{r},t-t'). \tag{14.24}$$

Here, the integration limit 0^- denotes a limit of δ, where the limit $\delta \longrightarrow 0$ is taken from below.

14.1.3 Classical Green Tensor

Now we can ask, what are the decoupled wave equations corresponding to the above Maxwell equations in linear, dispersive media? It is most convenient to stick to the frequency domain, as we have seen above. First, taking the curl of the third Maxwell equation in (14.12), and using $\mathbf{B} = \mu_0 \mathbf{H}$,

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r},\omega)] = i\omega\mu_0 \nabla \times \mathbf{H}(\mathbf{r},\omega). \tag{14.25}$$

Then using Eq. (14.17) and the last Maxwell equation, we arrive at the wave equation for the electric field:

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r},\omega)] - \omega^2 \mu_0 \epsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega) = i\mu_0\omega \mathbf{j}(\mathbf{r},\omega).$$

(wave equation in dielectric media) (14.26)

In free space, it is conventional to continue by using the vector identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}$ to replace the iterated curl in terms of the Laplacian. However, this doesn't help much here because $\nabla \cdot \mathbf{E} \neq 0$ in general for a dielectric, even without a source charge. Note that this wave equation in this form hides the source charge density ρ. In general, this is not a concern because any time-dependent charge (such as a dipole) will generate a field through the current density $\mathbf{j}$, which is tied to the charge density by the continuity constraint. Unfortunately the magnetic field $\mathbf{H}$ does not decouple as cleanly into its own wave equation, so we will stick to analyzing the electric field $\mathbf{E}$ as the primary object.

Now as a general solution of the wave equation (14.26), we will introduce the classical **Green tensor** $\mathbf{G}(\mathbf{r},\mathbf{r}',\omega)$. Specifically, we will define the Green tensor component $G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)$ to be the solution $E_\alpha(\mathbf{r})$, given a localized source current

$$i\mu_0\omega j_\mu(\mathbf{r},\omega) \longrightarrow \mu_0\omega^2 \delta^3(\mathbf{r}-\mathbf{r}')\,\delta_{\mu\beta}. \tag{14.27}$$

That is, it is the α component of the electric-field solution to Eq. (14.26), assuming we replace the right-hand side by a delta function, with an orientation along the β direction (the factor of $\mu_0\omega^2$ is arbitrary, but we will see the reason for this choice later). The Green tensor is thus the solution of the impulse-driven wave equation

$$\nabla \times [\nabla \times \mathbf{G}(\mathbf{r},\mathbf{r}',\omega)] - \omega^2 \mu_0 \epsilon(\mathbf{r},\omega)\,\mathbf{G}(\mathbf{r},\mathbf{r}',\omega) = \mu_0\omega^2\delta^3(\mathbf{r}-\mathbf{r}'),$$

(Green tensor wave equation) (14.28)

where the delta function on the right-hand side, being set equal to a tensor object, is understood to be proportional to the identity tensor. We can write this same equation in components as

$$\Box_{\alpha\beta} G^{(0)}_{\beta\gamma}(\mathbf{r},\mathbf{r}',\omega) = \mu_0\omega^2\delta_{\alpha\gamma}\delta^3(\mathbf{r}-\mathbf{r}'), \tag{14.29}$$

where the box operator is defined by

$$\Box := (\nabla \times \nabla\times) - \omega^2\mu_0\epsilon(\mathbf{r},\omega), \tag{14.30}$$

or in components,

$$\begin{aligned} \Box_{\alpha\beta} &:= \varepsilon_{\alpha\mu\gamma}\varepsilon_{\gamma\nu\beta}\partial_\mu\partial_\nu - \omega^2\mu_0\epsilon(\mathbf{r},\omega)\delta_{\alpha\beta} \\ &= (\partial_\alpha\partial_\beta - \nabla^2\delta_{\alpha\beta}) - \omega^2\mu_0\epsilon(\mathbf{r},\omega)\delta_{\alpha\beta}, \end{aligned} \tag{14.31}$$

where $\varepsilon_{\alpha\beta\gamma}$ is the **Levi-Civita permutation symbol** [which is zero if any two indices have the same value, or 1 or -1 if $(\alpha\beta\gamma)$ is respectively an even or odd permutation of (xyz)]. The first form follows from the component representation of the vector curl as

$$(\nabla \times \mathbf{A})_\alpha = \varepsilon_{\alpha\beta\gamma}\partial_\beta A_\gamma, \tag{14.32}$$

while the second follows from the identity $\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A}$. Note that by assuming a scalar field $\epsilon(\mathbf{r},\omega)$, we are implicitly assuming an *isotropic* medium, without "preferred" directions. In general, to handle something such as a dielectric medium, we would have to make the generalization to a permittivity *tensor*

$$\epsilon(\mathbf{r},\omega)\,\delta_{\alpha\beta} \longrightarrow \epsilon_{\alpha\beta}(\mathbf{r},\omega), \tag{14.33}$$

to give the most general linear response (in terms of direction) of the medium to the field.

As with any Green function, the Green tensor here is convenient to know, as it in principle represents a very general solution to the wave equation. Because the wave equation is linear, the solution for an *arbitrary* source $i\mu_0\omega\mathbf{j}(\mathbf{r},\omega)$ can then be written as the integral over the Green tensor and the source current,

$$\mathbf{E}(\mathbf{r},\omega) = \frac{i}{\omega}\int d^3r'\,\mathbf{G}(\mathbf{r},\mathbf{r}',\omega)\cdot\mathbf{j}(\mathbf{r}',\omega), \tag{14.34}$$

or to disambiguate the tensor product,

$$E_\alpha(\mathbf{r},\omega) = \frac{i}{\omega}\int d^3r'\, G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) j_\beta(\mathbf{r}',\omega). \tag{14.35}$$

This solution follows from the fact that the source can be viewed as a weighted sum over delta functions:

$$i\mu_0\omega j_\alpha(\mathbf{r},\omega) = \frac{i}{\omega}\int d^3r' \left[\mu_0\omega^2\delta_{\alpha\beta}\delta(\mathbf{r}-\mathbf{r}')\right] j_\beta(\mathbf{r}',\omega). \tag{14.36}$$

The Green tensor gives the solution in the case of each of the component delta functions $\mu_0\omega^2\delta_{\alpha\beta}\delta(\mathbf{r}-\mathbf{r}')$, and the full solution is just a sum over the Green tensors, since the wave equation is linear.

14.1.3.1 Example: Green Tensor in Free Space

As it turns out, the electromagnetic Green tensor is quite simple in the case of free space: essentially, it is just the electric field due to a single dipole. To see this, consider what is the meaning of an oscillating current density localized to a single point. A simple model of an oscillating dipole is two opposite but equal charges, one fixed (as in an atomic nucleus), and one oscillating in space. The moving charge, however, implies an oscillating current density. Furthermore, the dipole idealization involves the contraction of the distance $\mathbf{r}$ between the two charges to zero, while increasing the charge q such that the dipole moment $\mathbf{d} = q\mathbf{r}$ remains fixed. In this limit, the current density likewise becomes localized to a single point.

To see this more formally, the multipole expansion (cf. Section 9.5.4) about the origin $\mathbf{r} = 0$ comes from the identity

$$\rho(\mathbf{r},\omega) = \int d^3r'\, \delta^3(\mathbf{r}'-\mathbf{r})\, \rho(\mathbf{r}',\omega), \tag{14.37}$$

along with the expansion [Eq. (9.110)]

$$\delta^3(\mathbf{r}-\mathbf{r}') = \delta^3(\mathbf{r}) - (\mathbf{r}'\cdot\nabla)\,\delta^3(\mathbf{r}) + \frac{1}{2}(\mathbf{r}'\cdot\nabla)^2\,\delta^3(\mathbf{r}) + \cdots, \tag{14.38}$$

so that

$$\rho(\mathbf{r},\omega) = \left[\int d^3r'\,\rho(\mathbf{r}',\omega)\right]\delta^3(\mathbf{r}) - \left[\int d^3r'\, r'_\alpha\rho(\mathbf{r}',\omega)\right]\partial_\alpha\delta^3(\mathbf{r}) + \left[\frac{1}{2}\int d^3r'\, r'_\alpha r'_\beta\rho(\mathbf{r}',\omega)\right]\partial_\alpha\partial_\beta\delta^3(\mathbf{r}) + \cdots. \tag{14.39}$$

Repeated indices imply summations as usual here. The quantities in square brackets, from left to right, are the monopole moment (total charge), the dipole moment vector, and the quadrupole moment tensor. When the charge density is multiplied by another function under an integral, the moments act as coefficients of a Taylor expansion, weighting each of the derivatives of the other function.

Now that we have defined the multipoles, we can compute the multipole expansion corresponding to the localized current density

$$i\mu_0\omega\mathbf{j} = \mu_0\omega^2\delta^3(\mathbf{r}-\mathbf{r}')\,\hat{\varepsilon}, \tag{14.40}$$

where ε is a unit (polarization) vector. To simplify notation, we will take $\mathbf{r}' = 0$, so the expansion about the origin is a multipole expansion about $\mathbf{r}'$. The Fourier transform of the continuity constraint (14.7) is

$$\nabla\cdot\mathbf{j}(\mathbf{r},\omega) = i\omega\rho(\mathbf{r},\omega), \tag{14.41}$$

so that the localized current implies the charge density

$$\rho(\mathbf{r},\omega) = -\hat{\varepsilon}\cdot\nabla\delta^3(\mathbf{r}). \tag{14.42}$$

Clearly, the monopole moment of this charge density vanishes, since the integral of the derivative of the delta function vanishes. This is perhaps more evident by examining the identity

$$\int_{-\infty}^{\infty} dz\,\delta'(z)f(z) = -f'(0), \tag{14.43}$$

which is easy to prove via integration by parts.

Now to compute the dipole moment:

$$\begin{aligned}\mathbf{d}(\omega) &= \int d^3r\, \mathbf{r}\rho(\mathbf{r},\omega) \\ &= -\int d^3r\, \mathbf{r}\hat{\varepsilon}\cdot\nabla\delta^3(\mathbf{r}) \\ &= \hat{\varepsilon}\int d^3r\, \delta^3(\mathbf{r}) \\ &= \hat{\varepsilon}.\end{aligned} \tag{14.44}$$

To get rid of the gradient here, we simply integrated by parts. Thus, the localized current density implied by the Green tensor is simply a dipole, oriented in the same direction as the source delta-function vector, of *unit magnitude* (now we see the reason for the factor of $\mu_0\omega^2$ in front of the delta function in the Green-tensor equation (14.28); without this factor, the dipole here would have amplitude $1/\mu_0\omega^2$). A calculation analogous to this one shows that all higher-order multipoles vanish, because they involve higher powers of $\mathbf{r}$. After integration by parts, they will leave a final integral of the form $\int d^3r\, r_\alpha r_\beta \cdots r_\gamma \delta^3(\mathbf{r})$, which vanishes.

Thus, the Green-tensor elements $G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)$ in free space are constructed simply as follows: it is the α component of the electric field vector at position $\mathbf{r}$, due to an oscillating dipole located at $\mathbf{r}'$, which is oriented along $\hat{r}_\beta$ and oscillates with unit "amplitude." Note that the effective dipole amplitude, instead of having the expected SI units of (C m), is in fact *dimensionless*, so that the resulting electric field, normally of units (N/C), has units of $\mathrm{N/C^2m}$. Thus, the Green tensor has units of $\mathrm{N/C^2m}$, which agrees with what we find by examining the defining relation (14.28), or by noting that the product with a dipole moment $G_{\alpha\beta}\, d_\beta$ should have the dimension of an electric field. The units we have chosen here are rather odd, but they will yield a particularly simple form for the Green tensor (14.129) that will parallel other, familiar Green functions that we have already used but not specifically identified as such.

We have written down the electric field for a single dipole, located at the origin and oscillating along an arbitrary direction $\hat{\varepsilon}$ before in Eqs. (1.42) and (1.115):

$$\mathbf{E}^{(+)}(\mathbf{r},\omega) = \frac{1}{4\pi\epsilon_0}\left\{[3(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}]\left[\frac{1}{r^3}-i\frac{k}{r^2}\right]-[(\hat{\varepsilon}\cdot\hat{r})\hat{r}-\hat{\varepsilon}]\,\frac{k^2}{r}\right\} d^{(+)}(\omega)\, e^{ikr}. \tag{14.45}$$

Recall that $\hat{r}$ is a unit vector in the $\mathbf{r}$ direction, and $\hat{\varepsilon}$ is a unit vector marking the dipole orientation. This completely determines the form of the present Green tensor, and evidently from our discussion above, we may write the free-space Green tensor as[1]

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\left\{[3\hat{r}_\alpha\hat{r}_\beta-\delta_{\alpha\beta}]\left[\frac{1}{r^3}-i\frac{k}{r^2}\right]-[\hat{r}_\alpha\hat{r}_\beta-\delta_{\alpha\beta}]\,\frac{k^2}{r}\right\} e^{ikr},$$

(free-space Green tensor) (14.46)

where $k=\omega/c$, $\hat{r}_\alpha \equiv r_\alpha/r$ is the component of the unit vector $\hat{r}$ along the α direction, and the naught superscript indicates that this is the Green tensor for free space. To obtain this expression, we have set $d^{(+)}(\omega)\longrightarrow 1$, associated the index α with the field orientation, and associated the index β with the dipole orientation $\hat{\varepsilon}$. To obtain the general case of the Green function for a dipole located at $\mathbf{r}'$ instead of the origin, we may clearly write

$$G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = G^{(0)}_{\alpha\beta}(\mathbf{r}-\mathbf{r}',0,\omega), \tag{14.47}$$

since the Green function here only depends on the *difference* between the two coordinates (i.e., free space is invariant under translations).

In the presence of a material of some shape, the Green tensor is modified by reflections of fields at interfaces, changes of wavelength due to dispersion, and attenuation via absorption. In general, it will be

[1] Technically, this expression, along with the radiation field (14.45), are only valid for $\mathbf{r}\neq 0$. For these to be valid everywhere, there should be an additional "contact" term, corresponding for example to $-(1/3\epsilon_0)\delta_{\alpha\beta}\,\delta^3(\mathbf{r})$ in the Green tensor (14.46). However, in the following we will not write this term, because it is typically not useful for Casimir–Polder calculations.

quite a bit more complicated than in the simple free-space case. However, under the restrictions of our assumptions about the media, the Green tensor *completely characterizes the response of the medium to external perturbations, even within quantum electrodynamics.*

14.1.3.2 Green Tensor in Free Space: Alternate Forms

A couple of other forms of the free-space Green tensor (14.46) will be useful later on, and they provide a bit more insight now, so we'll go ahead and derive them. The first alternate form is

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\Big(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\Big)\frac{e^{ikr}}{r}. \tag{14.48}$$
(free-space Green tensor)

The equivalence of this expression to Eq. (14.46) can be verified directly by evaluating the derivatives in this expression (see Problem 8.8). The factor here of $(\delta_{\alpha\beta} + \partial_\alpha\partial_\beta/k^2)$ is characteristic of a *transverse* field, as we see by comparison with the momentum-representation expression for the transverse delta function $\delta^\perp_{\alpha\beta}(\mathbf{r})$, Eq. (8.178). In fact, this is essentially the projection operator for a transverse wave at frequency ck, so from this we see that the dipole field is just the transverse part of the scalar spherical wave. This form for the Green tensor will be useful in deriving the Green tensor for a planar material interface (Section 14.3.5).

To arrive at the other alternate form, we can use $\nabla r = \hat{r}$ and $\nabla\cdot\hat{r} = 2/r$ to write

$$\begin{aligned}
\nabla^2\frac{e^{ikr}}{r} &= -4\pi\delta^3(\mathbf{r}) + \frac{1}{r}\nabla^2 e^{ikr} + 2\left(\nabla\frac{1}{r}\right)\cdot\left(\nabla e^{ikr}\right)\\
&= -4\pi\delta^3(\mathbf{r}) + \frac{ik}{r}\nabla\cdot\left(\hat{r}e^{ikr}\right) + 2\left(-\frac{\hat{r}}{r^2}\right)\cdot\left(ik\hat{r}e^{ikr}\right)\\
&= -4\pi\delta^3(\mathbf{r}) + \left(-k^2\frac{e^{ikr}}{r} + 2ik\frac{e^{ikr}}{r^2}\right) + \left(-2ik\frac{e^{ikr}}{r^2}\right)\\
&= -4\pi\delta^3(\mathbf{r}) - k^2\frac{e^{ikr}}{r},
\end{aligned} \tag{14.49}$$

where we used $\nabla^2(1/r) = -4\pi\delta^3(\mathbf{r})$. Thus, we can replace the k^2 by $-\nabla^2$, if we also introduce a delta-function term, so that we arrive at the representation

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\Big(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\Big)\frac{e^{ikr}}{r} - \frac{1}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}). \tag{14.50}$$
(free-space Green tensor)

We will use this form in deriving the Green tensor in the presence of a second atom (Section 14.3.7).

14.1.3.3 Derivation of the Formula for the Dipole Radiation Field

We derived the above forms for the free-space Green tensor based on the knowledge of the dipole radiation field (14.45), which we simply wrote down. Now we are in a position to justify the above forms for $\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega)$ and thus for the dipole radiation field. We start with the wave equation Eq. (14.28) with $\epsilon(\mathbf{r},\omega) = \epsilon_0$, which defines the Green tensor in free space. We also use $\mu_0\epsilon_0 = 1/c^2$, and $\omega = ck$, obtaining

$$\nabla\times[\nabla\times\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega)] - k^2\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega) = \frac{k^2}{\epsilon_0}\delta^3(\mathbf{r}-\mathbf{r}'). \tag{14.51}$$

Then using the vector identity $\nabla\times(\nabla\times\mathbf{A}) = \nabla(\nabla\cdot\mathbf{A}) - \nabla^2\mathbf{A}$, we rearrange terms to find

$$\left(\nabla^2 + k^2\right)\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega) = \nabla[\nabla\cdot\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega)] - \frac{k^2}{\epsilon_0}\delta^3(\mathbf{r}-\mathbf{r}'). \tag{14.52}$$

In components, this relation reads

$$\left(\nabla^2 + k^2\right)G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = \partial_\alpha\partial_\gamma G^{(0)}_{\gamma\beta}(\mathbf{r},\mathbf{r}',\omega) - \frac{k^2}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}'). \tag{14.53}$$

Guided by our comments above, we make the *ansatz* that the dipole radiation field is the transverse part of a scalar field, and thus we assume

$$G^{(0)}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = \Big(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\Big)\psi(\mathbf{r}-\mathbf{r}'), \tag{14.54}$$

where $\psi(\mathbf{r})$ is a scalar function to be determined. Putting this *ansatz* into the wave equation for the Green tensor, we find

$$\begin{aligned}\Big(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\Big)\left(\nabla^2 + k^2\right)\psi(\mathbf{r}-\mathbf{r}') &= \partial_\alpha\partial_\gamma\Big(k^2\delta_{\gamma\beta} + \partial_\gamma\partial_\beta\Big)\psi(\mathbf{r}-\mathbf{r}') - \frac{k^2}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}') \\ &= \partial_\alpha\partial_\beta\Big(k^2 + \nabla^2\Big)\psi(\mathbf{r}-\mathbf{r}') - \frac{k^2}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}').\end{aligned} \tag{14.55}$$

Cancelling terms, we find

$$k^2\delta_{\alpha\beta}\left(\nabla^2 + k^2\right)\psi(\mathbf{r}-\mathbf{r}') = -\frac{k^2}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}'). \tag{14.56}$$

For nonzero k, we then find

$$\left(\nabla^2 + k^2\right)\psi(\mathbf{r}) = -\frac{1}{\epsilon_0}\delta^3(\mathbf{r}). \tag{14.57}$$

Evidently, $\psi(\mathbf{r})$ is essentially the Green function for the scalar wave equation (Helmholtz equation). Note that we already showed in Eq. (14.49) that

$$\left(\nabla^2 + k^2\right)\frac{e^{ikr}}{r} = -4\pi\delta^3(\mathbf{r}), \tag{14.58}$$

and thus it follows that

$$\psi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0}\frac{e^{ikr}}{r}. \tag{14.59}$$

That is, $\psi(\mathbf{r})$ is a spherical wave. This result, together with the *ansatz* (14.54), establishes the validity of the form (14.48) for the Green tensor, and thus of the other associated forms (14.46) and (14.50) and the dipole radiation field (14.45). Note that the scalar Green function (14.59), and thus the Green tensor (14.48), can correspond to *outgoing* or *incoming* waves, if k is respectively positive or negative.

14.1.4 Permittivity Properties

The medium itself is described solely by the function $\epsilon(\mathbf{r},\omega)$ (or equivalently, the susceptibility), within the assumptions of a medium that is: linear, dispersive, isotropic, spatially nondispersive, nonmagnetic, and describable in terms of macroscopic variables (i.e., where one can coarse-grain away any microscopic structure). It is thus worth spending some time looking at this function more carefully.[2] First, since $g_\chi(t)$, as in Eq. (14.19), relates the real fields $\mathbf{P}(\mathrm{r},t)$ and $\mathbf{E}(\mathrm{r},t)$, $g_\chi(t)$ itself is a real function. This implies that

$$\chi(-\omega) = \chi^*(\omega^*). \tag{14.60}$$

This result is implied by the relation (14.16), because for any real field, $\mathbf{A}^*(\omega) = \mathbf{A}(-\omega^*)$, since the frequency components at ω and $-\omega^*$, being complex conjugates, must be present with the same amplitude. From the definition (14.18) of the permittivity, it has the same frequency constraint as the susceptibility:

$$\epsilon(-\omega) = \epsilon^*(\omega^*). \tag{14.61}$$

In particular, this implies that $\mathrm{Re}[\epsilon(\omega)]$ is an *even* function of $\omega \in \mathbb{R}$, while $\mathrm{Im}[\epsilon(\omega)]$ is an *odd* function.

Furthermore, from the Fourier-transform relation (14.20), we can see that under the physically reasonable assumption that $g_\chi(t)$ is bounded, $\chi(\mathbf{r},\omega)$ is analytic in the upper half of the complex-frequency plane,

[2] For further reading, see John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), section 7.10; or L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, 1960), §58-62.

$\mathrm{Im}[\omega] > 0$. This is because the Fourier integral (14.20) is always cut off exponentially, and thus guaranteed to converge. Thus, $\epsilon(\omega)$ is also analytic in the upper half-plane. But what about right on the real axis? For dielectrics, $\epsilon(\omega)$ is analytic on the real axis, if we make the reasonable assumption that $g_\chi(t) \longrightarrow 0$ as $t \longrightarrow \infty$, so that the medium response "forgets" the input field in the distant past. In this case the oscillatory integral (14.20) converges because of this cutoff. However, for conductors, which are also modeled by a (complex) permittivity $\epsilon(\omega)$, the situation is slightly more complicated. Starting with the assumption of linear, dispersive conduction,

$$\mathbf{j}(\omega) = \sigma(\omega)\mathbf{E}(\omega), \tag{14.62}$$

where $\sigma(\omega)$ is the **conductivity**, the wave equation (14.26) becomes

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r},\omega)] - \omega^2\mu_0\epsilon(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega) = i\mu_0\omega\sigma(\mathbf{r},\omega)\mathbf{E}(\mathbf{r},\omega). \tag{14.63}$$

The source term can then be combined with the permittivity term as

$$\nabla \times [\nabla \times \mathbf{E}(\mathbf{r},\omega)] - \omega^2\mu_0 \left[\epsilon(\mathbf{r},\omega) + i\frac{\sigma(\mathbf{r},\omega)}{\omega}\right]\mathbf{E}(\mathbf{r},\omega) = 0. \tag{14.64}$$

The bracketed quantity in the second term can then be interpreted as a complex permittivity,

$$\tilde{\epsilon}(\mathbf{r},\omega) = \left[\epsilon(\mathbf{r},\omega) + i\frac{\sigma(\mathbf{r},\omega)}{\omega}\right], \tag{14.65}$$
(permittivity for a conductor)

and we can henceforth regard $\epsilon(\omega)$ as modeling conductors as well as dielectrics. Since the conductivity tends to a finite constant at zero frequency, we see that for a conductor, the permittivity $\epsilon(\omega)$ (including conduction effects) has a simple pole at the origin. Thus, except for a possible simple pole at the origin, $\epsilon(\omega)$ is analytic for $\mathrm{Im}[\omega] \geq 0$.[3]

14.1.4.1 Energy Loss and Poynting's Theorem

Before, we argued that the total energy of the field in free space was [Eq. (8.34)]

$$U = \frac{1}{2}\int d^3r \, \left[\epsilon_0 E^2 + \mu_0 H^2\right], \tag{14.66}$$

so that the field energy density is

$$u = \frac{1}{2}\left[\epsilon_0 E^2 + \mu_0 H^2\right]. \tag{14.67}$$

As we recall, in considering loss of transported energy, we need only consider *transverse* fields here, though we will drop the usual superscript denoting this. To generalize this expression to the dispersive-dielectric case, we must decompose the fields into frequency components, and then make the replacement $\epsilon_0 \longrightarrow \epsilon(\omega)$ at each frequency ω. We can thus write the dielectric energy spectral density as

$$u(\omega) = \frac{1}{2}\int d^3r \, \left[\mathbf{E}^{(+)}(\mathbf{r},\omega)\cdot\mathbf{D}^{(-)}(\mathbf{r},\omega) + \mathbf{H}^{(+)}(\mathbf{r},\omega)\cdot\mathbf{B}^{(-)}(\mathbf{r},\omega)\right] + \text{c.c.}, \tag{14.68}$$

such that the total energy density is

$$u = \int_0^\infty d\omega \, u(\omega). \tag{14.69}$$

This comes from writing the real fields at frequency ω as, e.g., $\mathbf{E}^{(+)}(\mathbf{r},\omega)e^{-i\omega t} + \mathbf{E}^{(-)}(\mathbf{r},\omega)e^{i\omega t}$, and then discarding the fast-rotating terms to effect a time average over short times corresponding to optical oscillations. Note that the electric-field part of the energy, represented by $\mathbf{E}\cdot\mathbf{D} = \epsilon_0 E^2 + \mathbf{P}\cdot\mathbf{E}$ explicitly contains both the free-field energy as well as the coupling of the medium polarization to the field.

[3] In fact, under the assumption that $\mathrm{Im}[\epsilon(\omega)] > 0$ on the real axis except at $\omega = 0$, $\epsilon(\omega) - \epsilon_0$ also turns out to not have any *zeroes* in the upper half-plane. See L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Pergamon, 1980), §123.

In fact, for a dispersive dielectric, this energy is not constant. However, differentiating this expression is difficult, as in the time domain we recall from Eq. (14.24) that such a differentiation will not be simple. Thus, we will start from a different direction. The force on a charge q due to the electric field is $\mathbf{F} = q\mathbf{E}$, and thus the rate at which the field does work on the charge is $\mathbf{F}\cdot\mathbf{v} = q\mathbf{v}\cdot\mathbf{E}$ (no work is done by the magnetic field, since the force is normal to the velocity $\mathbf{v}$). For a charge density ρ, the rate per unit volume at which the electromagnetic field does work on the charge is $\rho\mathbf{v}\cdot\mathbf{E} = \mathbf{j}\cdot\mathbf{E}$. This is the rate at which field energy is converted to heat by the electromagnetic force, and thus represents an energy *dissipation*. Using the Maxwell equations $\nabla\times\mathbf{H} = \partial_t\mathbf{D}+\mathbf{j}$ and $\nabla\times\mathbf{E} = -\partial_t\mathbf{B}$, along with the vector identity $\nabla\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\nabla\times\mathbf{A}) - \mathbf{A}\cdot(\nabla\times\mathbf{B})$, we may write

$$\begin{aligned}\mathbf{j}\cdot\mathbf{E} &= \mathbf{E}\cdot(\nabla\times\mathbf{H}) - \mathbf{E}\cdot\partial_t\mathbf{D} \\ &= \mathbf{H}\cdot\nabla\times\mathbf{E} - \nabla\cdot(\mathbf{E}\times\mathbf{H}) - \mathbf{E}\cdot\partial_t\mathbf{D} \\ &= -\nabla\cdot(\mathbf{E}\times\mathbf{H}) - [\mathbf{E}\cdot\partial_t\mathbf{D} + \mathbf{H}\cdot\partial_t\mathbf{B}]\,.\end{aligned} \tag{14.70}$$

Rearranging terms and using the definition $\mathbf{S} := \mathbf{E}\times\mathbf{H}$ of the Poynting vector, we arrive at **Poynting's theorem**:

$$-\nabla\cdot\mathbf{S} = \mathbf{j}\cdot\mathbf{E} + [\mathbf{E}\cdot\partial_t\mathbf{D} + \mathbf{H}\cdot\partial_t\mathbf{B}]\,. \tag{14.71}$$
(Poynting's theorem)

To see what this means, we can integrate over a volume and use the divergence theorem to write Poynting's theorem as

$$-\oint\mathbf{S}\cdot d\mathbf{a} = \int d^3r\;\mathbf{j}\cdot\mathbf{E} + \int d^3r\;[\mathbf{E}\cdot\partial_t\mathbf{D} + \mathbf{H}\cdot\partial_t\mathbf{B}]\,. \tag{14.72}$$
(Poynting's theorem)

Recalling that the Poynting vector represents the energy flux density of the field, the surface integral on the left represents the rate of energy *entering* the volume. The stuff on the right-hand side must therefore tell us where this incoming energy is going. We have already decided that the first term on the right-hand side represents loss of energy due to motion of free charges, or equivalently conversion of electromagnetic energy to heat. Now the important interpretation here is that the second term *also* represents field energy loss, but due to *bound* charges, in the form of the polarization of the dielectric medium (and the magnetization in the magnetodielectric case). Thus we may define

$$R := \mathbf{E}\cdot\partial_t\mathbf{D} + \mathbf{H}\cdot\partial_t\mathbf{B}.$$
(dielectric dissipation rate per unit volume) (14.73)

as the rate of energy loss per unit volume due to dielectric absorption.

For a monochromatic field, we have, using $\epsilon(-\omega) = \epsilon^*(\omega)$ for $\omega\in\mathbb{R}$,

$$\begin{aligned}\mathbf{E}(\mathbf{r},\omega)\cdot\partial_t\mathbf{D}(\mathbf{r},\omega) &= \left(\mathbf{E}^{(+)}e^{-i\omega t} + \mathbf{E}^{(-)}e^{i\omega t}\right)\cdot\left(-i\omega\epsilon(\omega)\mathbf{E}^{(+)}e^{-i\omega t} + i\omega\epsilon^*(\omega)\mathbf{E}^{(-)}e^{i\omega t}\right) \\ &= [-i\omega\epsilon(\omega) + i\omega\epsilon^*(\omega)]\left|E^{(+)}\right|^2 \\ &= 2\omega\mathrm{Im}[\epsilon(\omega)]\left|E^{(+)}\right|^2.\end{aligned} \tag{14.74}$$

Note that in the second step, we discarded terms rotating at optical frequencies, replacing them by their zero average values. The same calculation for the magnetic-field term gives

$$\mathbf{H}(\mathbf{r},\omega)\cdot\partial_t\mathbf{B}(\mathbf{r},\omega) = 2\omega\mathrm{Im}[\mu_0]\left|H^{(+)}\right|^2 = 0 \tag{14.75}$$

for a nonmagnetic medium. Thus, the dissipation rate for a field at frequency ω is

$$R(\omega) = 2\omega\mathrm{Im}[\epsilon(\omega)]\left|E^{(+)}\right|^2.$$
(dielectric dissipation rate per unit volume) (14.76)

Thus, we see that $\mathrm{Im}[\epsilon(\omega)]$ controls the dielectric dissipation of energy. Assuming a *passive* medium (i.e., one with no gain), $R(\omega) \geq 0$, and thus it follows that for positive frequencies,

$$\mathrm{Im}[\epsilon(\omega)] \geq 0 \qquad (\omega > 0, \text{ passive medium}). \tag{14.77}$$

Similarly, a passive medium has $\mathrm{Im}[\epsilon(\omega)] \geq 0$ for $\omega < 0$, which is consistent with the requirement from Eq. (14.61) that $\mathrm{Im}[\epsilon(\omega)]$ be an odd function of ω.

14.1.4.2 Kramers–Kronig Relations

At this point we will need a couple of results from complex analysis. Consider a closed, simply connected contour in the complex plane, and let $f(z)$ be a function analytic everywhere inside the contour. Then an integral around the contour vanishes,

$$\oint f(z)\, dz = 0, \tag{14.78}$$

and **Cauchy's integral formula** states that on the same contour,

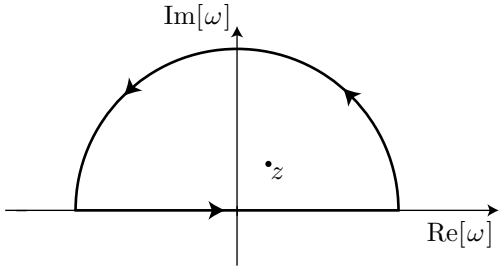
$$f(a) = \frac{1}{2\pi i} \oint \frac{f(z)}{z-a}\, dz \tag{14.79}$$

if a is a point interior to the contour and the integration proceeds counterclockwise around the contour (a clockwise integration implies an extra minus sign). The common terminology is that the function $f(z)/(z-a)$ has a **simple pole** at $z = a$, and $f(z)$ is the **residue** of the function $f(z)/(z-a)$ at the pole. Remarkably, both results are *independent of the contour*, so long as the above conditions hold.

We can rewrite Cauchy's integral formula as

$$\epsilon(z) - \epsilon_0 = \frac{1}{2\pi i} \oint \frac{\epsilon(\omega') - \epsilon_0}{\omega' - z}\, d\omega', \tag{14.80}$$

where z is a point in the upper half-plane, where the contour runs along the real axis and along a semicircle in the upper half-plane, with the radius of the semicircle expanding to infinity.



Our goal will be to keep only the part of the contour on the real axis, and get rid of the semicircular part. If we integrate the Fourier-transform relation for $\chi(\omega)$ by parts, we find

$$\begin{aligned} \frac{\epsilon(\omega) - \epsilon_0}{\epsilon_0} = \chi(\omega) &= \int_0^\infty dt\, g_\chi(t)\, e^{i\omega t} \\ &= \frac{i g_\chi(0^+)}{\omega} - \frac{g'_\chi(0^+)}{\omega^2} - \frac{i g''_\chi(0^+)}{\omega^3} + \cdots . \end{aligned} \tag{14.81}$$

The arguments are 0^+ because the integral extends over $t \geq 0$. But under the assumption that $g_\chi(t)$ is continuous,

$$g_\chi(0^+) = g_\chi(0^-) = 0, \tag{14.82}$$

so that the first term in the asymptotic expansion vanishes. The continuity requirement is physically reasonable because $g_\chi(t)$ represents a temporal response of the medium polarization to the input field. This

signal involves the displacement of electrons, which cannot happen instantaneously in response to a sudden change in the field. Thus, we have the asymptotic behavior

$$\begin{aligned} \mathrm{Re}[\epsilon(\omega) - \epsilon_0] &= O(\omega^{-2}) \\ \mathrm{Im}[\epsilon(\omega) - \epsilon_0] &= O(\omega^{-3}). \end{aligned} \tag{14.83}$$

It is evident as the semicircle is taken out to infinity, its contribution to the contour integral vanishes. Thus, we may write the contour integral (14.80) as

$$\epsilon(z) - \epsilon_0 = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\epsilon(\omega') - \epsilon_0}{\omega' - z}\, d\omega'. \tag{14.84}$$

Now if we take the point z and move it to the real axis, $z \longrightarrow \omega + i0^+$, so that

$$\epsilon(\omega) - \epsilon_0 = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\epsilon(\omega') - \epsilon_0}{\omega' - \omega - i0^+}\, d\omega'. \tag{14.85}$$

Notice that by our contour construction, the 0^+ says that the integration contour must be deformed below the real line to avoid the pole at the real frequency ω. If we instead deform the contour *above* the real line, the integral *vanishes*, because the semicircular contour no longer contains a pole:

$$0 = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\epsilon(\omega') - \epsilon_0}{\omega' - \omega - i0^-}\, d\omega'. \tag{14.86}$$

Adding together Eqs. (14.85) and (14.86), we can write

$$\epsilon(\omega) - \epsilon_0 = -\frac{i}{\pi} \int_{-\infty}^{\infty} \frac{\epsilon(\omega') - \epsilon_0}{\omega' - \omega}\, d\omega', \tag{14.87}$$

where the cut integral sign represents the **Cauchy principal value** integral,

$$\int_{-\infty}^{\infty} \frac{f(z')}{z' - z}\, dz' := \frac{1}{2} \left[\int_{-\infty}^{\infty} \frac{f(z')}{z' - z - i0^+}\, dz' + \int_{-\infty}^{\infty} \frac{f(z')}{z' - z - i0^-}\, dz' \right], \tag{14.88}$$

which averages the results of avoiding the pole by deforming the contour on either side of it. Note that the Cauchy principal value can be alternately defined as a symmetric excision of the pole from the real-axis integral,

$$\int_{-\infty}^{\infty} \frac{f(z')}{z' - z}\, dz' = \lim_{\delta \longrightarrow 0} \left[\int_{-\infty}^{z-\delta} \frac{f(z')}{z' - z}\, dz' + \int_{z+\delta}^{\infty} \frac{f(z')}{z' - z}\, dz' \right], \tag{14.89}$$

as can be seen from the contour definition by considering semicircular deformations of the contour around the pole. In either case, we may write out the real and imaginary parts of the complex relation (14.87) to obtain the **Kramers–Kronig relations**

$$\begin{aligned} \mathrm{Re}[\epsilon(\omega) - \epsilon_0] &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Im}[\epsilon(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega' \\ \mathrm{Im}[\epsilon(\omega) - \epsilon_0] &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Re}[\epsilon(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega', \end{aligned} \tag{14.90}$$

(Kramers–Kronig relations)

relating the real and imaginary parts of $\epsilon(\omega) - \epsilon_0$. These relations can be written more compactly in terms of the **Hilbert transform**, defined as

$$\mathscr{H}[f(x)] := \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(x')}{x' - x}\, dx', \tag{14.91}$$

so that

$$\epsilon(\omega) - \epsilon_0 = -i\mathscr{H}[\epsilon(\omega) - \epsilon_0]\,, \tag{14.92}$$

or separated into real and imaginary parts,

$$\begin{aligned} \mathrm{Re}\,[\epsilon(\omega) - \epsilon_0] &= \quad \mathscr{H}\{\mathrm{Im}\,[\epsilon(\omega) - \epsilon_0]\} \\ \mathrm{Im}\,[\epsilon(\omega) - \epsilon_0] &= -\mathscr{H}\{\mathrm{Re}\,[\epsilon(\omega) - \epsilon_0]\}\,. \end{aligned} \tag{14.93}$$

Note that the we derived the Kramers–Kronig relations just on the assumption that $\epsilon(\omega) - \epsilon_0$ represents a causal response of the medium (polarization field) to an input stimulus (electric field). *Any* causal response function, such as the gain spectrum of an amplifier circuit, must satisfy the same relations.

From the Kramers–Kronig relations, some features of the permittivity are immediately apparent. First, notice that the Hilbert transform acts ''something like'' a derivative. To see this, note that the Hilbert transform of a constant vanishes:

$$\mathscr{H}[c] = \frac{c}{\pi}\int_{-\infty}^{\infty} \frac{dx'}{x' - x} = 0. \tag{14.94}$$

This is because the convolution kernel $1/(x' - x)$ is antisymmetric about x, and even though the integrand is singular the contributions to the integral on either side of x cancel. This is precisely how the Cauchy principal value avoids problems with the singular integrand. However, as soon as the integrand has a slope, the antisymmetry of the kernel ''picks it out,'' because it upsets the cancellation in the above integral. In terms of the Kramers–Kronig relations, this implies that any variation in the real part of ϵ implies a nonzero imaginary part, and vice versa. We can rephrase this statement as: *the existence of dispersion implies the existence of absorption* (though possibly at some *other* frequency). This is quite a general result, being only due to the causal nature of the temporal response function $g_\chi(t) = g_\chi(t)\Theta(t)$, which implied that $\epsilon(\omega)$ is analytic in the upper half-plane, which allowed us to the whole contour-integral calculation. Of course, these conclusions apply to *any* causal, dispersive response function.

The other obvious implication of the Kramers–Kronig relations is that for the integrals in Eqs. (14.93) to converge, the integrands must vanish at large frequencies, so that $\epsilon(\omega) \longrightarrow \epsilon_0$ as $\omega \longrightarrow \infty$. Physically, this means that any medium must behave like the vacuum (i.e., become perfectly transparent) for very large frequencies. This is reasonable in the context of Lorentz-type models for media, since the atom will refuse to respond to fields with frequencies well over the atomic resonance frequencies. But we see here that this property is also implied by causality.

Finally, we note that in the above derivation, we ignored any possible pole at $\omega = 0$. In fact, the Kramers–Kronig relations written above are valid only for dielectric media, and must be adapted to the case of conductors (Problem 14.1).

14.1.4.3 Imaginary Frequencies

It turns out to be useful to know the behavior of the permittivity $\epsilon(\omega)$ on the positive imaginary axis, $\epsilon(is)$ for $s > 0$. First, from the reality constraint (14.61), we can let $\omega \longrightarrow -is$ to find

$$\epsilon(is) = \epsilon^*(is), \tag{14.95}$$

and thus $\epsilon(is) \in \mathbb{R}$ for $s > 0$. We can go even further along this line. Using the first of the Kramers–Kronig relations (14.93) in the form (Problem 14.2),

$$\mathrm{Re}[\epsilon(\omega)] = \epsilon_0 + \frac{2}{\pi}\int_0^{\infty} \frac{\omega'\mathrm{Im}[\epsilon(\omega')]}{\omega'^2 - \omega^2}\, d\omega', \tag{14.96}$$

we can let $\omega \longrightarrow is$ so that

$$\epsilon(is) = \mathrm{Re}[\epsilon(is)] = \epsilon_0 + \frac{2}{\pi}\int_0^{\infty} \frac{\omega'\mathrm{Im}[\epsilon(\omega')]}{\omega'^2 + s^2}\, d\omega'. \tag{14.97}$$

(permittivity for imaginary frequencies)

Since $\mathrm{Im}[\epsilon(\omega)] \geq 0$ for positive (real) frequencies in the absence of gain, as we showed above, clearly $\epsilon(is) \geq 0$, where the equality only holds in vacuum. Furthermore, the nonnegative integrand decreases monotonically with s, and thus $\epsilon(is)$ is a *monotonically decreasing function* of s, so long as $s > 0$. Finally, from this expression it is apparent that for any imaginary frequency is, $\epsilon(is)$ contains information about $\epsilon(\omega)$ over the whole real axis. Specifically, the imaginary part of $\epsilon(\omega)$ is involved here, although of course information about the real part is implicitly included as well in view of the Kramers–Kronig relations.

As an example, consider the general permittivity for an atomic vapor corresponding to the Lorentz model (1.32):

$$\epsilon(\omega) = \epsilon_0 + \frac{Ne^2}{m} \sum_j \frac{f_{0j}}{\omega_{j0}^2 - \omega^2 - i\gamma_j \omega}. \tag{14.98}$$

For the imaginary frequency, this expression becomes

$$\epsilon(is) = \epsilon_0 + \frac{Ne^2}{m} \sum_j \frac{f_{0j}}{\omega_{j0}^2 + s^2 + \gamma_j s}, \tag{14.99}$$

which is obviously positive, real, and monotonically decreasing with s. These same results obviously hold for closely related functions such as the polarizability $\alpha(is)$ and the susceptibility $\chi(is)$.

Another way of thinking about the imaginary-axis permittivity comes from Eq. (14.20):

$$\frac{\epsilon(is) - \epsilon_0}{\epsilon_0} = \chi(is) = \int_0^\infty dt\, g_\chi(t)\, e^{-st}. \tag{14.100}$$

Hence $\epsilon(is)$ [and of course $\chi(is)$] is related to the *Laplace* transform of $g_\chi(t)$. This is of course because $\chi(\omega)$ is the *Fourier* transform of $g_\chi(t)$, but the Laplace integral is properly cut off at zero here because $g_\chi(t)$ is a causal response function.

14.2 Generalized Susceptibility and Linear-Response Theory

Recall that the usual electromagnetic susceptibility $\chi(\mathbf{r}, \omega)$ gives the dispersive response of the dielectric medium polarization $\mathbf{P}(\mathbf{r}, \omega)$ to an applied field $\mathbf{E}(\mathbf{r}, \omega)$:

$$\mathbf{P}(\mathbf{r}, \omega) = \epsilon_0 \chi(\mathbf{r}, \omega) \mathbf{E}(\mathbf{r}, \omega). \tag{14.101}$$

Of course, this is only the *linear* response, valid for small applied fields. Now we want to generalize this notion of the susceptibility to the quantum case, and to more general input-response situations.[4]

Consider the interaction Hamiltonian

$$H_{\mathrm{int}}(t) = -x_\alpha F_\alpha(t), \tag{14.102}$$

where $\mathbf{x}$ is an operator coupling to a time-dependent "force" function $\mathbf{F}(t)$. We can write the linear response of the expectation value $\langle x_\alpha(t) \rangle$ in general as

$$\delta\langle x_\alpha(t) \rangle = \int_{-\infty}^\infty dt'\, g_{\alpha\beta}(t')\, F_\beta(t - t'), \tag{14.103}$$

assuming that the perturbation has been in effect since the distant past (corresponding to the $+\infty$ integration limit). Again, the integration limits ensure a causal response to the perturbation. The Fourier transform of this relation is, by the convolution theorem,

$$\delta\langle x_\alpha(\omega) \rangle = \chi_{\alpha\beta}(\omega)\, F_\beta(\omega), \tag{14.104}$$

[4] Ryogo Kubo and Kazuhisa Tomita, "A General Theory of Magnetic Resonance Absorption," *Journal of the Physical Society of Japan* **8**, 888 (1954); R. B. Stinchcombe, "Kubo and Zubarev Formulations of Response Theory," in *Correlation Functions and Quasiparticle Interactions in Condensed Matter*, J. Woods Halley, Ed. (Plenum, 1978), p. 3; A. D. McLachlan, "Retarded dispersion forces between molecules," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **271**, 387 (1963).

where

$$\begin{aligned}\chi_{\alpha\beta}(\omega) &= \int_0^\infty d\tau\, g_{\alpha\beta}(\tau)\, e^{i\omega\tau} \\ \delta\langle x_\alpha(\omega)\rangle &= \int_{-\infty}^\infty dt\, \delta\langle x_\alpha(t)\rangle\ e^{i\omega t} \\ F_\alpha(\omega) &= \int_{-\infty}^\infty dt\, F_\alpha(t)\, e^{i\omega t}.\end{aligned} \tag{14.105}$$

What we want to show is that the temporal response can be written as the correlation function

$$g_{\alpha\beta}(\tau) = \frac{i}{\hbar}\big\langle [x_\alpha(\tau), x_\beta(0)]\big\rangle\, \Theta(\tau), \tag{14.106}$$

where $x_\alpha(\tau)$ is in the interaction picture. Note that the Heaviside function enforces a causal response in Eq. (14.103). In what follows, we will want to show that the generalized susceptibility can be written

$$\chi_{\alpha\beta}(\omega) = \frac{i}{\hbar}\int_0^\infty d\tau \big\langle [x_\alpha(\tau), x_\beta(0)]\big\rangle\, e^{i\omega\tau}. \tag{14.107}$$

(generalized susceptibility)

That is, in the linear regime, the generalized susceptibility can be written as a Fourier transform of a quantum correlation function.

14.2.1 Proof

To prove these results, we will start by representing the quantum state by the density operator

$$\rho = \rho_0 + \delta\rho, \tag{14.108}$$

where ρ_0 represents the quantum state in the presence of only the background, time-independent Hamiltonian H_0, and $\delta\rho$ is the linear-order correction due to $H_{\rm int}$. Then the equation of motion for the quantum state is

$$\partial_t \rho = -\frac{i}{\hbar}\left[H_0 + H_{\rm int}, \rho_0 + \delta\rho\right]. \tag{14.109}$$

Note that the use of Hamiltonian evolution here is generally valid even with dissipation, since we can always extend the system to a larger Hilbert space where the evolution is Hamiltonian. Then canceling the zeroth-order parts and discarding the second-order term $[H_{\rm int}, \delta\rho]$, the equation of motion becomes

$$\begin{aligned}\partial_t \delta\rho &= -\frac{i}{\hbar}\left[H_0, \delta\rho\right] - \frac{i}{\hbar}\left[H_{\rm int}, \rho_0\right] \\ &= -\frac{i}{\hbar}\left[H_0, \delta\rho\right] + \frac{i}{\hbar}\left[x_\alpha, \rho_0\right] F_\alpha(t).\end{aligned} \tag{14.110}$$

Transforming to the interaction picture by defining

$$\delta\tilde\rho(t) = e^{iH_0 t/\hbar}\, \delta\rho(t)\, e^{-iH_0 t/\hbar}, \tag{14.111}$$

this equation of motion becomes

$$\partial_t \delta\tilde\rho = \frac{i}{\hbar} e^{iH_0 t/\hbar}\left[x_\alpha, \rho_0\right] F_\alpha(t)\, e^{-iH_0 t/\hbar}. \tag{14.112}$$

Integrating this equation gives

$$\delta\tilde\rho(t) = \frac{i}{\hbar}\int_{-\infty}^t dt'\, e^{iH_0 t'/\hbar}\left[x_\alpha, \rho_0\right] F_\alpha(t')\, e^{-iH_0 t'/\hbar}, \tag{14.113}$$

and then transforming out of the interaction picture, the formal solution for the perturbed state is

$$\delta\rho(t) = \frac{i}{\hbar}\int_{-\infty}^{t} dt'\, e^{iH_0(t'-t)/\hbar}\,[x_\alpha,\rho_0]\,F_\alpha(t')\,e^{-iH_0(t'-t)/\hbar}. \tag{14.114}$$

Note that we have assumed $\delta\rho(t\longrightarrow-\infty)=0$, corresponding to a perturbation turned on adiabatically in the distant past.

Now the perturbation to the mean response it

$$\begin{aligned}
\delta\langle x_\alpha(t)\rangle &= \mathrm{Tr}\,[x_\alpha\,\delta\rho(t)] \\
&= \frac{i}{\hbar}\int_{-\infty}^{t} dt'\,\mathrm{Tr}\left\{x_\alpha\,e^{iH_0(t'-t)/\hbar}\,[x_\beta,\rho_0]\,e^{-iH_0(t'-t)/\hbar}\right\}F_\beta(t') \\
&= \frac{i}{\hbar}\int_{-\infty}^{t} dt'\,\mathrm{Tr}\left\{x_\alpha(t-t')\,[x_\beta,\rho_0]\right\}F_\beta(t') \\
&= \frac{i}{\hbar}\int_{-\infty}^{t} dt'\,\langle[x_\alpha(t-t'),x_\beta]\rangle\,F_\beta(t') \\
&= \frac{i}{\hbar}\int_{-\infty}^{0} d\tau\,\langle[x_\alpha(-\tau),x_\beta]\rangle\,F_\beta(t+\tau) \\
&= \frac{i}{\hbar}\int_{0}^{\infty} d\tau\,\langle[x_\alpha(\tau),x_\beta]\rangle\,F_\beta(t-\tau).
\end{aligned} \tag{14.115}$$

Here we have used $t'=\tau+t$, we have defined the interaction-picture response operator

$$x_\alpha(t) = e^{iH_0t/\hbar}\,x_\alpha\,e^{-iH_0t/\hbar}, \tag{14.116}$$

and, to linear order, we see that the expectation value is taken with respect to the background state ρ_0. Comparison to Eq. (14.103) thus allows us to identify

$$g_{\alpha\beta}(\tau) = \frac{i}{\hbar}\langle[x_\alpha(\tau),x_\beta]\rangle = \frac{i}{\hbar}\langle[x_\alpha(\tau),x_\beta(0)]\rangle\,\Theta(\tau), \tag{14.117}$$

which is our desired result.

14.2.2 Atom and Field Susceptibilities

For the atom–field interaction, corresponding to the interaction Hamiltonian $H_{\mathrm{int}}=-\mathbf{d}\cdot\mathbf{E}$, we can work out two important generalized-susceptibility formulae. First thinking about the dipole moment as an operator coupled to a classical electric field, we can consider the dipole response due to the field. In fact, we already have treated this in detail, as the mean dipole induced by a classical field is the atomic polarizability $\alpha(\omega)$. In the general case, we can write the polarizability as a tensor,

$$d_\mu(\omega) = \alpha_{\mu\nu}(\omega)\,E_\nu(\omega), \tag{14.118}$$

so that we do not necessarily assume a spherically symmetric atom (molecule). Thus, Eq. (14.107) implies that

$$\alpha_{\mu\nu}(\omega) = \frac{i}{\hbar}\int_0^\infty d\tau\,\langle[d_\mu(\tau),d_\nu(0)]\rangle\,e^{i\omega\tau}.$$

(polarizability in terms of dipole correlation) (14.119)

It is clear from this expression that if d_μ and d_ν represent completely independent degrees of freedom, then the commutator vanishes and thus the particular component $\alpha_{\mu\nu}$ likewise vanishes. This happens, for example, when μ and ν refer to different principal axes of the atom. In fact, if the coordinate system

is aligned with the principal atomic axes, then it follows that $\alpha_{\mu\nu}$ is a diagonal tensor. For a spherically symmetric atom, *all* orthogonal coordinate systems form principal axes, and hence

$$\alpha_{\mu\nu}(\omega) = \alpha(\omega)\,\delta_{\mu\nu}, \tag{14.120}$$
(scalar polarizatibility)

where $\alpha(\omega)$ is the **scalar polarizability**.

It is also useful to invert the Fourier transform here, so that

$$\langle [d_\mu(\tau), d_\nu(0)] \rangle\, \Theta(\tau) = \frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega\, \alpha_{\mu\nu}(\omega)\, e^{-i\omega\tau}. \tag{14.121}$$

To eliminate the step function here, we can write down the complex conjugate of this equation, using $[A, B]^\dagger = -[A, B]$ if A and B are Hermitian operators, with the result

$$\langle [d_\mu(\tau), d_\nu(0)] \rangle\, \Theta(\tau) = \frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega\, \alpha^*_{\mu\nu}(\omega)\, e^{i\omega\tau}. \tag{14.122}$$

Now, letting $\tau \longrightarrow -\tau$, exchanging the subscripts $\mu \longleftrightarrow \nu$, and using $\alpha_{\mu\nu} = \alpha_{\nu\mu}$ (we'll show this in a bit),

$$\langle [d_\nu(-\tau), d_\mu(0)] \rangle\, \Theta(-\tau) = \frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega\, \alpha^*_{\mu\nu}(\omega)\, e^{-i\omega\tau}. \tag{14.123}$$

Now reversing the operators in the commutator and advancing the time of both operators in the commutator by τ (recall the correlation function only depends on time *differences*, since we aren't considering transients),

$$\langle [d_\mu(\tau), d_\nu(0)] \rangle\, \Theta(-\tau) = -\frac{\hbar}{2\pi i} \int_{-\infty}^{\infty} d\omega\, \alpha^*_{\mu\nu}(\omega)\, e^{-i\omega\tau}. \tag{14.124}$$

Now adding this equation to Eq. (14.121), we find

$$\langle [d_\mu(\tau), d_\nu(0)] \rangle = \frac{\hbar}{\pi} \int_{-\infty}^{\infty} d\omega\, \mathrm{Im}[\alpha_{\mu\nu}(\omega)]\, e^{-i\omega\tau}.$$
(dipole correlation in terms of polarizability) (14.125)

Now, what about the symmetry of the polarization tensor? Recall that according to the Hamiltonian $H_{\text{int}} = -\mathbf{d}\cdot\mathbf{E}$, the energy is the work required to adiabatically turn on the field in the presence of the induced dipole:

$$U = -\int_0^{\mathbf{E}} \mathbf{d}(\mathbf{E}')\cdot d\mathbf{E}' = -\frac{1}{2}\alpha_{\mu\nu}E_\mu E_\nu \tag{14.126}$$

This is valid at any for amplitudes of any component of frequency ω, though we will suppress the frequency dependence. It is also valid for linear polarization (i.e., it omits circular polarization), so this argument is not completely general. The change in energy due to a small field change is

$$\delta U = -\mathbf{d}(\mathbf{E})\cdot\delta\mathbf{E} = E_\mu \alpha_{\mu\nu}\delta E_\nu. \tag{14.127}$$

Of course, we can equivalently write this as

$$\delta U = -\delta\mathbf{E}\cdot\mathbf{d}(\mathbf{E}) = \delta E_\mu \alpha_{\mu\nu} E_\nu = E_\mu \alpha_{\nu\mu}\delta E_\nu. \tag{14.128}$$

For both these relations to hold for every field E_μ and every change δE_ν, we must identify $\alpha_{\mu\nu}(\omega) = \alpha_{\nu\mu}(\omega)$, and thus the polarizability tensor is symmetric.

Similarly, thinking of the field at $\mathbf{r}$ as responding to a classical dipole at position $\mathbf{r}'$, we may write the generalized susceptibility for the field. Recall that the classical Green tensor was precisely this response function for a dipole of unit amplitude, and thus we may write the Green tensor in terms of the correlation function as

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{\hbar} \int_0^{\infty} d\tau\, \langle [E_\alpha(\mathbf{r}, \tau), E_\beta(\mathbf{r}', 0)] \rangle\, e^{i\omega\tau}.$$
(Green tensor in terms of field correlation) (14.129)

The inverse relation is correspondingly

$$\langle [E_\mu(\mathbf{r},\tau), E_\nu(\mathbf{r}',0)] \rangle = \frac{\hbar}{\pi} \int_{-\infty}^{\infty} d\omega \, \mathrm{Im}[G_{\mu\nu}(\mathbf{r},\mathbf{r}',\omega)] \, e^{-i\omega\tau}.$$

(field correlation in terms of Green tensor) (14.130)

As it turns out, the expressions here are very useful in treating interactions between an atom and the quantum electromagnetic field.

One more note is in order here. The permittivity $\epsilon(\omega) - \epsilon_0$ is another generalized susceptibility. Thus, the properties derived in Section 14.1.4 that were not specific to the electromagnetic field apply also to $\alpha_{\mu\nu}(\omega)$ and $G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)$. To summarize, $G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)$ is analytic in the upper half-plane,

$$\begin{aligned} G_{\alpha\beta}(\mathbf{r},\mathbf{r}',-\omega) &= G^*_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega^*) \\ \mathrm{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)]}{\omega' - \omega} \, d\omega' \\ \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)]}{\omega' - \omega} \, d\omega', \end{aligned} \tag{14.131}$$

and $G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is)$ is real and monotonically decreasing with $s > 0$. Of course, all these properties apply to $\alpha_{\mu\nu}(\omega)$ as well.

14.3 Atom–Surface Potentials Near Dielectric Media

Now we can use the above formalism to compute the energy shift for a ground-state atom due to the electromagnetic vacuum, modified by a dielectric medium. The potential shift is given by second-order perturbation theory as

$$V_{\mathrm{CP}} = \langle \mathrm{g}|H_0|\mathrm{g}\rangle + \langle \mathrm{g}|H_{\mathrm{AF}}|\mathrm{g}\rangle + \sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle \mathrm{g}|H_{\mathrm{AF}}|\mathrm{e}_j, 1_{\mathbf{k},\zeta}\rangle|^2}{E_{\mathrm{g},0} - E_{\mathrm{e}_j,1_{\mathbf{k},\zeta}}}, \tag{14.132}$$

just as in our treatment of the Casimir–Polder potential in the conducting-plane case of Chapter 13, where $|\mathrm{g}\rangle$ is the ground state, the excited states are $|\mathrm{e}_j\rangle$, and the field modes are labeled by the wave vector $\mathbf{k}$ and the polarization index ξ. We are again considering the usual dipole interaction Hamiltonian $H_{\mathrm{AF}} = -\mathbf{d}\cdot\mathbf{E}$. Both the zeroth- and first-order terms vanish, leaving

$$V_{\mathrm{CP}} = -\sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle \cdot \langle 0|\mathbf{E}|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar(\omega_{j0} + \omega_{\mathbf{k}})}, \tag{14.133}$$

(second-order level shift)

where $\omega_{j0} := (E_j - E_0)/\hbar$, with $E_0 \equiv E_{\mathrm{g}}$ the energy of the ground state. Before proceeding we will now need to derive expressions for the atom and field susceptibilities up to a consistent order in perturbation theory.

14.3.1 Kramers–Heisenberg Formula

Recall that the atomic dipole susceptibility (polarizability) gives the mean induced dipole from the interaction

$$\begin{aligned} H_{\mathrm{AF}} &= -\mathbf{d}\cdot\mathbf{E} \\ &= -\sum_j \left[\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle \sigma_j + \langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle \sigma_j^\dagger \right] \cdot \left[\hat{\epsilon} E_0^{(+)} e^{-i\omega t} + \hat{\epsilon}^* E_0^{(-)} e^{i\omega t} \right] \end{aligned} \tag{14.134}$$

with the *classical* electric field $\mathbf{E}(t)$, which is of the same form as the linear-response interaction (14.102). Here, $\sigma_j := |\mathrm{g}\rangle\langle \mathrm{e}_j|$ is the usual atomic lowering operator for the $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}_j\rangle$ transition. Now we wish to

compute the generalized susceptibility (atomic polarizability) for the mean dipole response to the applied classical field. We can treat this interaction in time-dependent perturbation theory by deriving the perturbed ground state $|\mathrm{g}\rangle + \delta|\mathrm{g}\rangle$, where the state response

$$\delta|\mathrm{g}\rangle = \sum_j \left(a_j^{(+)} e^{-i\omega t} + a_j^{(-)} e^{i\omega t}\right) e^{-iE_0 t/\hbar}|\mathrm{e}_j\rangle, \tag{14.135}$$

to the perturbation represents the mixing in of the excited states by the interaction (E_0 is the ground-state energy). The coefficients $a_j^{(\pm)}$ remain to be determined. The idea is to then put this *ansatz* into the Schrödinger equation

$$\partial_t\Big[|\mathrm{g}\rangle + \delta|\mathrm{g}\rangle\Big] = -\frac{i}{\hbar}(H_0 + H_{\mathrm{AF}})\Big[|\mathrm{g}\rangle + \delta|\mathrm{g}\rangle\Big], \tag{14.136}$$

so that to first order

$$\partial_t\,\delta|\mathrm{g}\rangle = -\frac{i}{\hbar}H_0\,\delta|\mathrm{g}\rangle - \frac{i}{\hbar}H_{\mathrm{AF}}|\mathrm{g}\rangle, \tag{14.137}$$

where $|\mathrm{g}\rangle = |\mathrm{g}(t)\rangle = |\mathrm{g}(0)\rangle e^{-iE_0 t/\hbar}$. The result is

$$\begin{aligned}\sum_j &\Big[-i(\omega + E_0/\hbar)a_j^{(+)}e^{-i\omega t} + i(\omega - E_0/\hbar)a_j^{(-)}e^{i\omega t}\Big]\, e^{-iE_0 t/\hbar}|\mathrm{e}_j\rangle \\ &= -\frac{i}{\hbar}\sum_j E_j\left(a_j^{(+)}e^{-i\omega t} + a_j^{(-)}e^{i\omega t}\right)e^{-iE_0 t/\hbar}|\mathrm{e}_j\rangle \\ &\quad + \frac{i}{\hbar}\sum_j \langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\Big[\hat{\epsilon}E_0^{(+)}e^{-i\omega t} + \hat{\epsilon}^* E_0^{(-)}e^{-i\omega t}\Big]\, e^{-iE_0 t/\hbar}|\mathrm{e}_j\rangle.\end{aligned} \tag{14.138}$$

Matching coefficients of $e^{\pm i\omega t}|\mathrm{e}_j\rangle$,

$$\begin{aligned}-i(\omega + E_0/\hbar)a_j^{(+)} &= -\frac{i}{\hbar}E_j a_j^{(+)} + \frac{i}{\hbar}\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\hat{\epsilon}E_0^{(+)} \\ i(\omega - E_0/\hbar)a_j^{(-)} &= -\frac{i}{\hbar}E_j a_j^{(-)} + \frac{i}{\hbar}\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\hat{\epsilon}^* E_0^{(-)},\end{aligned} \tag{14.139}$$

and thus the perturbation coefficients are

$$\begin{aligned}a_j^{(+)} &= \frac{\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\hat{\epsilon}E_0^{(+)}}{\hbar(\omega_{j0} - \omega)} \\ a_j^{(-)} &= \frac{\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\hat{\epsilon}^* E_0^{(-)}}{\hbar(\omega_{j0} + \omega)},\end{aligned} \tag{14.140}$$

where again $\omega_{j0} := (E_j - E_0)/\hbar$, and $E_0 \equiv E_{\mathrm{g}}$ is the ground-state energy. Now the mean dipole in the new ground state due to the perturbation is

$$\langle \mathbf{d}(t)\rangle = \langle \mathrm{g}(t)|\mathbf{d}|\delta\mathrm{g}(t)\rangle + \langle \delta\mathrm{g}(t)|\mathbf{d}|\mathrm{g}(t)\rangle, \tag{14.141}$$

to first order in the perturbation, assuming that $\langle \mathbf{d}\rangle = 0$ in the unperturbed ground state. Thus,

$$\begin{aligned}\langle \mathbf{d}(t)\rangle = \sum_j &\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle\cdot\left(\frac{\hat{\epsilon}E_0^{(+)}}{\hbar(\omega_{j0} - \omega)}e^{-i\omega t} + \frac{\hat{\epsilon}^* E_0^{(-)}}{\hbar(\omega_{j0} + \omega)}e^{i\omega t}\right) \\ &+ \sum_j\left(\frac{\hat{\epsilon}^* E_0^{(-)}}{\hbar(\omega_{j0} - \omega)}e^{i\omega t} + \frac{\hat{\epsilon}E_0^{(+)}}{\hbar(\omega_{j0} + \omega)}e^{-i\omega t}\right)\cdot\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle.\end{aligned} \tag{14.142}$$

Given that there is no circular polarization, the polarization vector $\hat{\epsilon}$ is real, and thus

$$\langle \mathbf{d}(t)\rangle = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2 - \omega^2)}\cdot\hat{\epsilon}\Big[E_0^{(+)}e^{-i\omega t} + E_0^{(-)}e^{i\omega t}\Big]. \tag{14.143}$$

Writing out only the positive-frequency amplitude,

$$\left\langle \mathbf{d}^{(+)}(\omega) \right\rangle = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2 - \omega^2)} \cdot \hat{\epsilon} E_0^{(+)} \tag{14.144}$$

and then using tensor-component notation,

$$\left\langle d_\mu^{(+)}(\omega) \right\rangle = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle (E_0^{(+)})_\nu}{\hbar(\omega_{j0}^2 - \omega^2)}, \tag{14.145}$$

we can compare this expression to the defining relation

$$\left\langle d_\mu^{(+)}(\omega) \right\rangle = \alpha_{\mu\nu}(\omega)(E_0^{(+)})_\nu \tag{14.146}$$

for the polarizability tensor to arrive at the following expression for the polarizability in terms of the dipole matrix elements:

$$\alpha_{\mu\nu}(\omega) = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2 - \omega^2)}. \tag{14.147}$$

(Kramers–Heisenberg formula)

For a spherically symmetric atom, all dipole-operator components d_μ are identical, and to be consistent with Eq. (14.120), which says $\alpha_{\mu\nu} = \alpha\,\delta_{\mu\nu}$, we find the scalar polarizability

$$\alpha(\omega) = \sum_j \frac{2\omega_{j0}|\langle \mathrm{g}|d_z|\mathrm{e}_j\rangle|^2}{\hbar(\omega_{j0}^2 - \omega^2)}. \tag{14.148}$$

(scalar Kramers–Heisenberg formula)

Either expression is referred to as a **Kramers–Heisenberg formula** for the atomic polarizability. We derived this expression for the atomic ground state, but of course it is valid for *any* atomic state $|q\rangle$, by letting $0 \longrightarrow q$ and $\mathrm{g} \longrightarrow q$ in the above expression, being careful to pay attention to the sign of $\omega_{jq} = (E_j - E_q)/\hbar$, which is negative for states of lower energy than $|q\rangle$.

Actually, we obtained the polarizability by calculating the *physical* dipole moment, which only involves the real part of $\alpha_{\mu\nu}(\omega)$. Thus, we have so far only computed the real part:

$$\mathrm{Re}[\alpha_{\mu\nu}(\omega)] = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2 - \omega^2)}. \tag{14.149}$$

According to the Kramers–Kronig relations, variation in $\mathrm{Re}[\alpha_{\mu\nu}(\omega)]$ leads to nonzero values of $\mathrm{Im}[\alpha_{\mu\nu}(\omega)]$.

$$\mathrm{Im}[\alpha_{\mu\nu}(\omega)] = -\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{Re}[\alpha_{\mu\nu}(\omega')]}{\omega' - \omega}\, d\omega'. \tag{14.150}$$

However, it turns out to be easier to guess $\mathrm{Im}[\alpha_{\mu\nu}(\omega)]$ from the inverse relation,

$$\mathrm{Re}[\alpha_{\mu\nu}(\omega)] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{Im}[\alpha_{\mu\nu}(\omega')]}{\omega' - \omega}\, d\omega', \tag{14.151}$$

if we compare it to Eq. (14.149) in the form

$$\mathrm{Re}[\alpha_{\mu\nu}(\omega)] = \sum_j \frac{\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar}\left(\frac{1}{\omega_{j0} - \omega} + \frac{1}{\omega_{j0} + \omega}\right). \tag{14.152}$$

The two expressions are consistent if

$$\mathrm{Im}[\alpha_{\mu\nu}(\omega)] = \frac{\pi}{\hbar}\sum_j \langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle\Big[\delta(\omega - \omega_{j0}) - \delta(\omega + \omega_{j0})\Big]. \tag{14.153}$$

These expressions are equivalent to the *classical* polarizability expression (from Chapter 1, if we write the oscillator strength in terms of the dipole matrix elements)

$$\alpha_{\mu\nu}(\omega) = \lim_{\Gamma_j \to 0} \sum_j \frac{\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar\left[(\omega_{j0}^2-\omega^2)^2+\Gamma_j^2\omega^2\right]}\left[(\omega_{j0}^2-\omega^2)+i\Gamma_j\omega\right], \tag{14.154}$$

if we take the limit of zero damping, being careful to note the emergence of the delta function in the imaginary part.

What happens if we *don't* assume that the field is linearly polarized? Going back to Eq. (14.147), this is still the correct formula for the tensor and scalar [Eq. (14.148)] parts. For the vector part, going back to Eq. (14.142), we can use the fact that the dipole combination is antisymmetric and the polarization conjugate $\hat{\epsilon}^* = -\hat{\epsilon}$ to write

$$\begin{aligned}\langle \mathbf{d}(t)\rangle = &\sum_j \langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle \cdot \left(\frac{\hat{\epsilon}E_0^{(+)}}{\hbar(\omega_{j0}-\omega)}e^{-i\omega t} - \frac{\hat{\epsilon}E_0^{(-)}}{\hbar(\omega_{j0}+\omega)}e^{i\omega t}\right)\\ &- \sum_j \langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle \cdot \left(-\frac{\hat{\epsilon}E_0^{(-)}}{\hbar(\omega_{j0}-\omega)}e^{i\omega t} + \frac{\hat{\epsilon}E_0^{(+)}}{\hbar(\omega_{j0}+\omega)}e^{-i\omega t}\right),\end{aligned} \tag{14.155}$$

so that (note the ω instead of the ω_{j0} in the numerator)

$$\langle \mathbf{d}(t)\rangle = \sum_j \frac{2\omega\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\langle \mathrm{e}_j|\mathbf{d}|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2-\omega^2)} \cdot \hat{\epsilon}\Big[E_0^{(+)}e^{-i\omega t} + E_0^{(-)}e^{i\omega t}\Big]. \tag{14.156}$$

The same line of reasoning gives

$$\alpha_{\mu\nu}^{(1)}(\omega) = \sum_j \frac{2\omega\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2-\omega^2)}$$

(Kramers–Heisenberg formula, vector part) (14.157)

for the vector part of the polarizability tensor (in the case of circular polarization).

14.3.2 Green Tensor

The interaction is symmetric with respect to the dipole and field operators. Thus, thinking of the mean, linear quantum response at $\mathbf{r}$ of all the electromagnetic field modes (labeled by $\mathbf{k},\zeta$) to a dipole at $\mathbf{r}'$, we may write down the analogous Kramers–Heisenberg formula for the Green tensor as

$$\mathrm{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] = \sum_n \sum_{\mathbf{k},\zeta} \frac{2n\omega_{\mathbf{k}}\langle 0|E_\alpha(\mathbf{r},\omega_{\mathbf{k}})|n_{\mathbf{k},\zeta}\rangle\langle n_{\mathbf{k},\zeta}|E_\beta(\mathbf{r}',\omega_{\mathbf{k}})|0\rangle}{\hbar\,(n^2\omega_{\mathbf{k}}^2-\omega^2)}. \tag{14.158}$$

Again, the numbers here denote the number of photons in the labeled mode, and the energy of $|n_{\mathbf{k},\zeta}\rangle$ is $n\hbar\omega_{\mathbf{k}}$. Since the field is a harmonic oscillator, the electric field operators are effectively a sum over creation and annihilation operators. Thus, when considering the response of the vacuum, only the one-photon states contribute to this sum:

$$\mathrm{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] = \sum_{\mathbf{k},\zeta} \frac{2\omega_{\mathbf{k}}\langle 0|E_\alpha(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle\langle 1_{\mathbf{k},\zeta}|E_\beta(\mathbf{r}',\omega_{\mathbf{k}})|0\rangle}{\hbar\,(\omega_{\mathbf{k}}^2-\omega^2)}.$$

(Kramers–Heisenberg formula for Green tensor) (14.159)

Correspondingly, the imaginary part of the Green tensor is

$$\mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] = \frac{\pi}{\hbar}\sum_{\mathbf{k},\zeta}\langle 0|E_\alpha(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle\langle 1_{\mathbf{k},\zeta}|E_\beta(\mathbf{r}',\omega_{\mathbf{k}})|0\rangle\Big[\delta(\omega-\omega_{\mathbf{k}})-\delta(\omega+\omega_{\mathbf{k}})\Big], \tag{14.160}$$

for consistency with causality.

14.3.2.1 Mode Expansion of the Green Tensor

A particular application of the Kramers–Heisenberg formula is to write down a Green tensor in terms of electromagnetic field modes, assuming such modes exist. This happens, for example, when quantizing in an enclosed domain, as in a cavity with perfectly conducting walls. Recall that in such cases, the electric-field operator has the form [Eq. (8.61)]

$$\mathbf{E}(\mathbf{r},t) = \sum_{\mathbf{k},\zeta} -\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})a_{\mathbf{k},\zeta}(t) + \text{H.c.}, \tag{14.161}$$

where the $\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})$ are the normalized classical mode functions for wave vector $\mathbf{k}$ and polarization ζ. Putting this field operator into Eq. (14.159), we find the general expression

$$\text{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] = \frac{1}{\epsilon_0}\sum_{\mathbf{k},\zeta}\frac{\omega_{\mathbf{k}}^2}{\omega_{\mathbf{k}}^2-\omega^2}f_{\mathbf{k},\zeta,\alpha}(\mathbf{r})f^*_{\mathbf{k},\zeta,\beta}(\mathbf{r}'). \tag{14.162}$$

(mode expansion of Green tensor)

The Green tensor is in principle a more general object, however, and simple mode functions may not exist in particular when the Green tensor represents dispersive (and thus absorbing) media. However, this general approach below to the interaction energy doesn't depend on the existence of mode functions in this simple sense; in principle, when absorption is present, the system can be extended to include extra, effective ''absorbing'' fields that account for absorption, while making the total, extended system Hamiltonian. Mode functions that span multiple fields exist in this case, and though possibly difficult to write down, their existence allows our arguments below to go through.

14.3.3 Interaction Energy

We will now claim that the second-order interaction energy (14.133) may be written in the form[5]

$$\begin{aligned} V_{\text{CP}}(\mathbf{r}) &= -\frac{\hbar}{2\pi}\int_0^\infty ds\,\text{Tr}\Big[\boldsymbol{\alpha}(is)\cdot\mathbf{G}(\mathbf{r},\mathbf{r},is)\Big] \\ &= -\frac{\hbar}{2\pi}\int_0^\infty ds\,\alpha_{\mu\nu}(is)\,G_{\nu\mu}(\mathbf{r},\mathbf{r},is), \end{aligned} \tag{14.163}$$

(Casimir–Polder potential)

where $\mathbf{r}$ denotes the location of the atom. This is a rather remarkable result, as although it assumes that both the atom and the field are quantized, it does not require us to state exactly *how* the field was quantized. This simplifies things dramatically, as quantization with dissipation is not simple. This result is also remarkable in that it only involves the classical susceptibilities, although to derive this we restricted ourselves to a linear (and thus lowest-order) interaction Hamiltonian between the atom and the quantum electromagnetic field.

To confirm this result, first note that $\text{Im}[\alpha_{\mu\nu}(is)]$ and $\text{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is)]$ will vanish for $s\in\mathbb{R}$, because they involve delta functions of the form $\delta(is-\omega_j)$, which will always vanish for real ω_j. Thus, on the imaginary axis, both susceptibilities are given only by their real parts, which again for any generalized susceptibility (Section 14.1.4.3) turn out to be real, monotonically decreasing functions of s. Now, we can

[5] A. D. McLachlan, ''Retarded dispersion forces between molecules,'' *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **271**, 387 (1963); A. D. McLachlan, ''Van der Waals forces between an atom and a surface,'' *Molecular Physics* **7**, 381 (1963) (doi: 10.1080/00268976300101141); J. M. Wylie and J. E. Sipe, ''Quantum electrodynamics near an interface,'' *Physical Review A* **30**, 1185 (1984) (doi: 10.1103/PhysRevA.30.1185).

substitute the above expressions for the susceptibilities into (14.163) to obtain

$$\begin{aligned} V_{\rm CP} &= -\frac{\hbar}{2\pi}\int_0^\infty ds \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2+s^2)} \sum_{\mathbf{k},\zeta} \frac{2\omega_{\mathbf{k}}\langle 0|E_\nu(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle\langle 1_{\mathbf{k},\zeta}|E_\mu(\mathbf{r},\omega_{\mathbf{k}})|0\rangle}{\hbar\,(\omega_{\mathbf{k}}^2+s^2)} \\ &= -\frac{2}{\pi\hbar}\sum_j\sum_{\mathbf{k},\zeta} |\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\cdot\langle 0|\mathbf{E}(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle|^2 \int_0^\infty ds\, \frac{\omega_{j0}\,\omega_{\mathbf{k}}}{(\omega_{j0}^2+s^2)(\omega_{\mathbf{k}}^2+s^2)} \\ &= -\sum_j\sum_{\mathbf{k},\zeta} \frac{|\langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle\cdot\langle 0|\mathbf{E}(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar(\omega_{j0}+\omega_{\mathbf{k}})}. \end{aligned} \tag{14.164}$$

Here, we have used the identity

$$\int_0^\infty dx\, \frac{ab}{(a^2+x^2)(b^2+x^2)} = \frac{\pi}{2(a+b)}, \qquad (a,b>0) \tag{14.165}$$

which follows by noting the integrand is an even function of x, changing to a contour integral over the great semicircle in the upper half-plane, and using the Cauchy integral formula, since the contour encloses the poles at $x = ia$ and $x = ib$. Note that the last expression in Eqs. (14.164) is precisely the second-order perturbation expression (14.133) that we started off with, so we see that Eq. (14.163) gives the correct atom–material interaction energy. The only result that we really needed to derive Eq. (14.163), beyond the contour integral, was the Kramers–Heisenberg formula (applied to both the atom and field susceptibilities).

14.3.4 Renormalization

We have already considered the form for the free-space Green tensor in Section 14.1.3.1. Thus, there is an interaction energy associated with an atom in free space,

$$V^{(0)} = -\frac{\hbar}{2\pi}\int_0^\infty ds\, \mathrm{Tr}\Big[\boldsymbol{\alpha}(is)\cdot\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r},is)\Big], \tag{14.166}$$

where $\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r},\omega)$ is the free-space Green tensor, Eq. (14.46). We have already seen (Section 13.12) that this energy diverges and after renormalization gives the Lamb shift. To investigate the vacuum interaction energy of an atom with a dielectric body, we will need to subtract this contribution to obtain the energy difference in the presence vs. the absence of the body. The correct (observed) renormalized interaction potential is thus $V_{\rm CP}^{(\mathbf{r})} = V_{\rm CP} - V^{(0)}$, or

$$V_{\rm CP} = -\frac{\hbar}{2\pi}\int_0^\infty ds\, \mathrm{Tr}\Big[\boldsymbol{\alpha}(is)\cdot\mathbf{G}^{(\mathrm{s})}(\mathbf{r},\mathbf{r},is)\Big], \tag{14.167}$$

(renormalized Casimir–Polder potential)

after dropping the superscript, where

$$\mathbf{G}^{(\mathrm{s})}(\mathbf{r},\mathbf{r}',\omega) := \mathbf{G}(\mathbf{r},\mathbf{r}',\omega) - \mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega) \tag{14.168}$$

(scattering Green tensor)

is the **scattering part of the Green tensor**, i.e., the part of the Green tensor due specifically to the presence of the dielectric body.

14.3.5 Planar Interface

To compute the atom–surface interaction with a planar interface, we will first start by expanding the free-space Green function into plane waves.[6] The idea is that in a very general way, we can compute the effect

[6] A. D. McLachlan, "Retarded dispersion forces between molecules," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **271**, 387 (1963); J. E. Sipe, "The Dipole Antenna Problem in Surface Physics: A New Approach" *Surface Science* **105**, 489 (1981) (doi: 10.1016/0039-6028(81)90014-5); T. Setälä, M. Kaivola, and A. T. Friberg, "Decomposition of the point-dipole field into homogeneous and evanescent parts," *Physical Review E* **59**, 1200 (1999) (doi: 10.1103/PhysRevE.59.1200).

of the planar interface on a plane wave, via the usual optical reflection coefficients. Then, summing over all plane-wave contributions, we can obtain the effect of the surface on a dipole field, and thus obtain the scattering Green tensor for the surface. We start by recalling the free-space Green tensor in the form (14.48):

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\Big(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\Big)\frac{e^{ikr}}{r}, \tag{14.169}$$

The spherical-wave factor e^{ikr}/r has the plane-wave expansion[7]

$$\frac{e^{ikr}}{r} = \frac{i}{2\pi}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}e^{i(k_x x + k_y y + k_z|z|)}, \tag{14.170}$$

where the magnitude of the wave vector $\mathbf{k}$ must be simply $k = \omega/c$, so

$$k_z = \begin{cases} \sqrt{k^2 - k_x^2 - k_y^2} & \text{if } k_x^2 + k_y^2 \le k^2 \\ i\sqrt{k_x^2 + k_y^2 - k^2} & \text{if } k_x^2 + k_y^2 > k^2. \end{cases} \tag{14.171}$$

Note that k_x and k_y correspond to *transverse* wave numbers in a wave-vector expansion, while k_z corresponds to the *longitudinal* wave number, thinking about propagation along the z direction. Small values of k_x and k_y (the upper expression for k_z) thus correspond to propagating waves (with a longitudinal phase factor of $e^{ik_z z}$), while, large values (the lower expression for k_z) correspond to *evanescent waves*[8] (with longitudinal exponential decay of $e^{-k_z z}$). Thus, we may write the expansion for the free-space Green function as

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}\Big(k^2\delta_{\alpha\beta} + \partial_\alpha\partial_\beta\Big)e^{i(k_x x + k_y y + k_z|z|)}.$$

(free-space Green tensor, plane-wave decomposition) (14.172)

We will generally imagine the interface to be *below* the dipole (i.e., smaller values of z), so that we will need the Green tensor in the region $z < 0$

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}\Big(k^2\delta_{\alpha\beta} - k^-_\alpha k^-_\beta\Big)e^{i(k_x x + k_y y - k_z z)}, \tag{14.173}$$

where $\mathbf{k}^- := k_x\hat{x} + k_y\hat{y} - k_z\hat{z}$, to determine the scattering from the surface. Further, we will ultimately need a Green tensor of the form $G_{\alpha\beta}(0,0,\omega)$, so we can discard the transverse spatial dependence by setting $x = y = 0$, since the planar surface is invariant under transverse displacement. Further restoring the explicit coordinate z' for the dipole,

$$G^{(0)}_{\alpha\beta}(z,z',\omega) = \frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}\Big(k^2\delta_{\alpha\beta} - k^-_\alpha k^-_\beta\Big)e^{-ik_z(z-z')}, \tag{14.174}$$

where again this expression is valid for $z < z'$.

14.3.5.1 Reflection Coefficients

Now we will introduce the reflection coefficients for the plane-wave field amplitudes $r_\perp(\theta,\omega)$ and $r_\parallel(\theta,\omega)$. These are defined such that in incoming and reflected field amplitudes are related by $E_{\text{out}} = rE_{\text{in}}$. The subscripts denote whether the incoming field polarization is perpendicular or parallel to the plane of incidence (the plane containing both the surface-normal vector and the incoming wave vector), and θ is the angle between the surface normal and the wave vector ($\theta = 0$ corresponds to normal incidence). The reflection coefficients are in general complex, reflecting phase shifts from the surface reflection. Note that these can be readily calculated by matching electromagnetic boundary conditions for bodies with planar-type symmetry,

[7] This is the **Weyl representation** for the spherical wave; see Leonard Mandel and Emil Wolf, *Optical Coherence and Quantum Optics* (Cambridge, 1995), Eq. (3.2-61).

[8] The evanescent modes are **plasmonic modes** in the case of a metal and **polaritonic modes** in the case of a dielectric.

including simple interfaces or stacks of thin films, including conducting films.[9] For example, for plane waves incident from vacuum onto a planar interface to a medium of dispersive permittivity $\epsilon(\omega)$, the reflection coefficients are

$$r_\perp(\theta,\omega) = \frac{\cos\theta - \sqrt{n^2(\omega) - \sin^2\theta}}{\cos\theta + \sqrt{n^2(\omega) - \sin^2\theta}}$$
$$r_\parallel(\theta,\omega) = \frac{\sqrt{n^2(\omega) - \sin^2\theta} - n^2(\omega)\cos\theta}{\sqrt{n^2(\omega) - \sin^2\theta} + n^2(\omega)\cos\theta}, \quad \text{(Fresnel reflection coefficients)} \tag{14.175}$$

where the refractive index $n(\omega)$ is defined by $n^2(\omega) = \epsilon(\omega)/\epsilon_0$. Note that the convention here is such that at normal incidence,

$$r_\perp(\theta = 0,\omega) = r_\parallel(\theta = 0,\omega) = \frac{1 - n(\omega)}{1 + n(\omega)}, \tag{14.176}$$

whereas a common alternate convention is to flip the sign of $r_\parallel$ compared to what we have here. These coefficients are then sufficient to determine the scattering part of the Green tensor in the presence of the interface, because we can now calculate how each plane-wave component of the free-space Green tensor is reflected by it.

14.3.5.2 Scattering Green Tensor Due to the Planar Interface

Now let us define our coordinates such that $\hat{z}$ is normal to the planar interface at $z = 0$, and we may assume the atom to be located at $x = y = 0$. We want the scattering Green tensor at the same location as the source dipole, $G^{(\mathrm{s})}_{\alpha\beta}(z,z,\omega)$. In this particular case, the *off-diagonal* components of the scattering Green tensor vanish: the reflection of an oscillating dipole in the r_α-direction will be polarized in the r_α-direction, from the point of view of the original dipole. This is just due to the symmetry of the surface, since the effect of the reflection on a wave vector $\mathbf{k}$ is the transformation $\mathbf{k} \longrightarrow \mathbf{k}^-$. We can start with the dipole oscillating in the z-direction. The scattering part of the Green tensor is then given by multiplying the dipole field by $-r_\parallel$, since the z-axis is always in the plane of incidence, but the extra minus sign is present because the z-component of the polarization vector is always reversed upon reflection from the surface. The phase factor is given by that of a wave traveling the distance z down to the surface and back, and the result is

$$G^{(\mathrm{s})}_{zz}(z,z,\omega) = -\frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y \, \frac{1}{k_z}\left(k^2 - k_z^2\right) r_\parallel(\theta,\omega)\, e^{2ik_z z}. \tag{14.177}$$

Switching the integration to cylindrical coordinates, we can carry out the angular integral and write

$$G^{(\mathrm{s})}_{zz}(z,z,\omega) = -\frac{i}{4\pi\epsilon_0}\int_0^{\infty} dk_\mathrm{T} \, \frac{k_\mathrm{T}^3}{k_z}\, r_\parallel(\theta,\omega)\, e^{2ik_z z}, \tag{14.178}$$

where the transverse part of the wave vector is

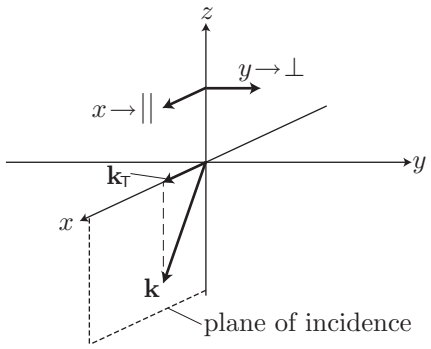
$$\mathbf{k}_\mathrm{T} = k_x\hat{x} + k_y\hat{y}, \tag{14.179}$$

so that

$$k_\mathrm{T} = \sqrt{k_x^2 + k_y^2} = \sqrt{k^2 - k_z^2}, \tag{14.180}$$

and now the incidence angle θ can be defined in the integration coordinates by $\tan\theta = k_\mathrm{T}/k_z$. To evaluate the transverse components of the Green tensor, first pretend that $\mathbf{k}_\mathrm{T} = k_\mathrm{T}\hat{x}$. That is, choose the transverse part of the wave vector to point in the x-direction.

[9] Daniel A. Steck, *Classical and Modern Optics* (2006), Chapters 9-10, available online at `http://steck.us/teaching/`. Note that $r_\perp \equiv r_\mathrm{S}$ and $r_\parallel \equiv r_\mathrm{P}$ in the notation there.



Then a dipole oscillating about x has a field polarized parallel to the incidence plane, and a dipole in the y-direction produces a perpendicular polarization. Thus, the appropriate scattering Green tensor components are

$$\begin{aligned} G^{(s)}_{xx}(z,z,\omega) &= \frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}\Big(k^2-k_\intercal^2\Big)\,r_\parallel(\theta,\omega)\,e^{2ik_z z} \\ G^{(s)}_{yy}(z,z,\omega) &= \frac{i}{8\pi^2\epsilon_0}\int_{-\infty}^{\infty}dk_x\int_{-\infty}^{\infty}dk_y\,\frac{1}{k_z}\Big(k^2\Big)\,r_\perp(\theta,\omega)\,e^{2ik_z z}. \end{aligned} \tag{14.181}$$

Switching to cylindrical components and carrying out the angular integral amounts to simply replacing the xx and yy components by their average:

$$G^{(s)}_{xx}(z,z,\omega) = G^{(s)}_{yy}(z,z,\omega) = \frac{i}{8\pi\epsilon_0}\int_0^{\infty}dk_\intercal\,\frac{k_\intercal}{k_z}\Big(k_z^2 r_\parallel(\theta,\omega)+k^2 r_\perp(\theta,\omega)\Big)\,e^{2ik_z z}. \tag{14.182}$$

These Green-tensor components are all we need to characterize the surface in terms of the atom–surface potential, which we now evaluate.

14.3.5.3 Explicit Expressions for the Atom–Surface Potential

Putting these Green-tensor components into the expression (14.167) for the atomic potential due to the planar surface, we find

$$\begin{aligned} V_{\rm CP} &= -\frac{\hbar}{2\pi}\int_0^{\infty}ds\,\alpha_{\mu\nu}(is)\,G^{(\rm s)}_{\nu\mu}(\mathbf{r},\mathbf{r},is) \\ &= -\frac{i\hbar}{8\pi^2\epsilon_0}\int_0^{\infty}ds\int_0^{\infty}dk_\intercal\,\frac{k_\intercal}{k_z}\left[\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}\Big(k_z^2 r_\parallel(\theta,is)+k^2 r_\perp(\theta,is)\Big)-\alpha_{zz}(is)k_\intercal^2 r_\parallel(\theta,is)\right]e^{2ik_z z}, \end{aligned} \tag{14.183}$$

where $k = is/c$, and $k_z = i\sqrt{s^2/c^2+k_\intercal^2}$. Defining the real quantity

$$\kappa(s,k_\intercal) := -ik_z = \sqrt{s^2/c^2+k_\intercal^2}, \tag{14.184}$$

we may rewrite this expression as

$$V_{\rm CP} = \frac{\hbar}{8\pi^2\epsilon_0}\int_0^{\infty}ds\int_0^{\infty}dk_\intercal\,\frac{k_\intercal}{\kappa}\left[\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}\Big(\kappa^2 r_\parallel(\theta,is)+\frac{s^2}{c^2}r_\perp(\theta,is)\Big)+\alpha_{zz}(is)k_\intercal^2 r_\parallel(\theta,is)\right]e^{-2\kappa z}, \tag{14.185}$$

(Casimir–Polder potential in terms of reflection coefficients)

so that now the potential is written entirely in terms of the reflection coefficients, the polarizability, real (non-negative) integration variables, and real functions of the integration variables. Notice that for a medium with $n > 1$, the reflection coefficients are in fact negative at normal incidence. So long as they stay predominantly negative, the overall potential is negative, and thus attractive (as the contributions decay exponentially with distance).

Suppose now that the atom is spherically symmetric, as is appropriate for typical ground-state atoms. Then the polarizability tensor becomes a scalar, and so[10]

$$V_{\text{CP}} = \frac{\hbar}{8\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\,\alpha(is)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}\left[r_\perp(\theta,is)+\left(1+\frac{2k_\text{T}^2c^2}{s^2}\right)r_\parallel(\theta,is)\right]e^{-2\kappa z}.$$

(Casimir–Polder potential, spherically symmetric atom) (14.186)

Note that the integrand here is rather smooth, and not oscillatory. As a function of the imaginary part s of the frequency, $\alpha(is)$ decreases monotonically owing to its function as a susceptibility. The same can be said of $r_{\perp,\parallel}(\theta,is)$, which can also be regarded as a susceptibility (linear response) of the reflected-field amplitude to the input field. Similarly κ increases smoothly with s, and any explicit dependence on s is smooth.

14.3.5.4 Perfect-Conductor Limit

To gain some more insight into the basic result, we can examine the potential (14.186) in various limits. The first limit we can examine is that of a perfect conductor, corresponding to perfect reflections:

$$r_{\perp,\parallel}(\theta,is)\longrightarrow -1. \tag{14.187}$$

In this case, the potential (14.186) becomes

$$V_{\text{CP}} = -\frac{\hbar}{4\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\,\alpha(is)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}\left(1+\frac{k_\text{T}^2c^2}{s^2}\right)e^{-2\kappa z}. \tag{14.188}$$

Using Eq. (14.148) for the polarizability in the form

$$\alpha(is) = \sum_j \frac{2k_{j0}|\langle \text{g}|d_z|\text{e}_j\rangle|^2}{\hbar c(s^2/c^2+k_{j0}^2)}, \tag{14.189}$$

and putting in the explicit form for κ, we find

$$\begin{aligned} V_{\text{CP}} &= -\frac{1}{2\pi^2\epsilon_0 c}\sum_j d_{j,z}^2 k_{j0}\int_0^\infty ds\int_0^\infty dk_\text{T}\,\frac{k_\text{T}\sqrt{s^2/c^2+k_\text{T}^2}}{s^2/c^2+k_{j0}^2}\,e^{-2z\sqrt{s^2/c^2+k_\text{T}^2}} \\ &= \frac{1}{4\pi^2\epsilon_0 c}\sum_j d_{j,z}^2 k_{j0}\,\partial_z\int_0^\infty \frac{ds}{s^2/c^2+k_{j0}^2}\int_0^\infty dk_\text{T}\,k_\text{T}\,e^{-2z\sqrt{s^2/c^2+k_\text{T}^2}} \\ &= \frac{1}{8\pi^2\epsilon_0 c}\sum_j d_{j,z}^2 k_{j0}\,\partial_z\int_0^\infty \frac{ds}{s^2/c^2+k_{j0}^2}\int_0^\infty d(k_\text{T}^2)\;e^{-2z\sqrt{s^2/c^2+k_\text{T}^2}}, \end{aligned} \tag{14.190}$$

where again $d_{j,z}^2$ is shorthand for $|\langle \text{g}|d_z|\text{e}_j\rangle|^2$. We can evaluate the second integral using

$$\int_0^\infty dx\,e^{-2z\sqrt{s^2/c^2+x}} = \frac{1+2zs/c}{2z^2}e^{-2zs/c}, \tag{14.191}$$

so that

$$\begin{aligned} V_{\text{CP}} &= \frac{1}{16\pi^2\epsilon_0 c}\sum_j d_{j,z}^2 k_{j0}\,\partial_z\frac{1}{z^2}\int_0^\infty ds\,\frac{1+2zs/c}{s^2/c^2+k_{j0}^2}e^{-2zs/c} \\ &= -\frac{1}{16\pi^2\epsilon_0 c}\sum_j d_{j,z}^2 k_{j0}\,\partial_z^2\frac{1}{z}\int_0^\infty \frac{ds}{s^2/c^2+k_{j0}^2}e^{-2zs/c}. \end{aligned} \tag{14.192}$$

[10] cf. Werner Vogel and Dirk-Gunnar Welsch, *Quantum Optics*, 3rd ed. (Wiley, 2006), Eq. (10.84), noting the alternate sign convention for $r_\parallel$. Also see A. D. McLachlan, ''Van der Waals forces between an atom and a surface,'' *Molecular Physics* **7**, 381 (1963) (doi: 10.1080/00268976300101141), Eq (3.9) for the case of a simple dielectric interface.

Now, we can evaluate the last integral, using the integral formula (Problem 13.1)[11]

$$\int_0^\infty dx\, \frac{e^{-\mu x}}{\beta^2 + x^2} = \frac{f(\beta\mu)}{\beta} \qquad (\mathrm{Re}[\beta] > 0, \mathrm{Re}[\mu] > 0), \tag{14.193}$$

where $f(z)$ is an auxiliary function to the sine and cosine integrals [see Eq. (13.26)]. Thus, the Casimir–Polder potential in this case becomes

$$V_{\text{CP}} = -\frac{1}{16\pi^2\epsilon_0} \sum_j d_{j,z}^2\, \partial_z^2 \frac{1}{z} f(2k_{j0}z),$$

(Casimir–Polder potential, perfect conductor) (14.194)

which of course agrees exactly with the result we computed before, Eq. (13.51).

14.3.5.5 Near-Field Limit

In the near-field limit ($2k_{j0}z \ll 1$ for all j), we can simplify the potential (14.186) for a spherically symmetric atom. First, note that due to the exponential factor in Eq. (14.186), only relatively small values of κ,

$$\kappa^2 = \frac{s^2}{c^2} + k_\text{T}^2 \lesssim \frac{1}{4z^2}, \tag{14.195}$$

will contribute in the integral to the potential. However, due to the form of the factor $s^2\alpha(is)/\kappa$, the most important values of s in the integral are of the order of ω_{j0}. But then in the near-field regime, we can conclude that

$$\frac{s}{c} \ll \frac{1}{2z}. \tag{14.196}$$

This means that typically in the integral,

$$\kappa \sim k_\text{T} \sim \frac{1}{2z}, \tag{14.197}$$

and these are all large compared to s/c. Physically, since $|s/c|$ is the optical wave vector, this means that the evanescent-wave modes—modes that propagate along the surface, decaying exponentially away from the surface—dominate the atom–surface potential in this regime. This is sensible, as surface modes *should* be most important in the near field.

We will proceed by considering an atom in the near field of a simple dielectric interface, so that the reflection coefficients are given by the Fresnel expressions (14.175). In the near field, we can use the relations above to simplify the Fresnel coefficients, since we may represent the incidence angle by

$$\begin{aligned} \sin\theta &= \frac{k_\text{T}}{k} = -i\frac{ck_\text{T}}{s} \\ \cos\theta &= \frac{k_z}{k} = \frac{c\kappa}{s}. \end{aligned} \tag{14.198}$$

[11] See I. S. Gradstein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 6th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2000), Formula 3.354.1.

In the case of transverse polarization, then, using $s/c \ll k_\tau, \kappa$,

$$\begin{aligned} r_\perp(\theta, is) &= \frac{\cos\theta - \sqrt{n^2(is) - \sin^2\theta}}{\cos\theta + \sqrt{n^2(is) - \sin^2\theta}} \\ &= \frac{\kappa - \sqrt{s^2n^2/c^2 + k_\tau^2}}{\kappa + \sqrt{s^2n^2/c^2 + k_\tau^2}} \\ &\approx \frac{\kappa - s^2n^2/2c^2k_\tau - k_\tau}{\kappa + s^2n^2/2c^2k_\tau + k_\tau} \\ &\approx \frac{(s^2/2c^2k_\tau)(1-n^2)}{2k_\tau + (s^2/2c^2k_\tau)(1+n^2)} \\ &\approx \frac{s^2}{4c^2k_\tau^2}(1-n^2), \end{aligned} \tag{14.199}$$

where we have used

$$\kappa = \sqrt{\frac{s^2}{c^2} + k_\tau^2} \approx \frac{s^2}{2c^2k_\tau} + k_\tau. \tag{14.200}$$

Thus, in the near field, $r_\perp \ll 1$, and we will in fact neglect it compared to the longitudinal reflection coefficient, which is

$$\begin{aligned} r_\parallel(\theta, is) &= \frac{\sqrt{n^2(is) - \sin^2\theta} - n^2(is)\cos\theta}{\sqrt{n^2(is) - \sin^2\theta} + n^2(is)\cos\theta} \\ &= \frac{\sqrt{n^2 + c^2k_\tau^2/s^2} - n^2c\kappa/s}{\sqrt{n^2 + c^2k_\tau^2/s^2} + n^2c\kappa/s} \\ &\approx \frac{1 - n^2(is)}{1 + n^2(is)} \\ &= \frac{\epsilon_0 - \epsilon(is)}{\epsilon_0 + \epsilon(is)}. \end{aligned} \tag{14.201}$$

Putting this coefficient into Eq. (14.186) and ignoring the transverse reflection,

$$\begin{aligned} V_{\text{CP}} &= -\frac{\hbar}{8\pi^2\epsilon_0c^2}\int_0^\infty ds\, s^2\,\alpha(is)\,\frac{\epsilon(is)-\epsilon_0}{\epsilon(is)+\epsilon_0}\int_0^\infty dk_\tau\,\frac{k_\tau}{\kappa}\left(1 + \frac{2k_\tau^2c^2}{s^2}\right)e^{-2\kappa z} \\ &= -\frac{\hbar}{8\pi^2\epsilon_0}\int_0^\infty ds\,\alpha(is)\,\frac{\epsilon(is)-\epsilon_0}{\epsilon(is)+\epsilon_0}\int_{s/c}^\infty d\kappa\,\left(\frac{s^2}{c^2} + 2k_\tau^2\right)e^{-2\kappa z} \\ &\approx -\frac{\hbar}{4\pi^2\epsilon_0}\int_0^\infty ds\,\alpha(is)\,\frac{\epsilon(is)-\epsilon_0}{\epsilon(is)+\epsilon_0}\int_0^\infty d\kappa\,\kappa^2\,e^{-2\kappa z}. \end{aligned} \tag{14.202}$$

The second integral is easy to carry out, with the result[12]

$$V_{\text{CP}} = -\frac{\hbar}{16\pi^2\epsilon_0z^3}\int_0^\infty ds\,\alpha(is)\,\frac{\epsilon(is)-\epsilon_0}{\epsilon(is)+\epsilon_0} \tag{14.203}$$

(van der Waals potential)

for the near-field (van der Waals) force near a dielectric surface of permittivity $\epsilon(\omega)$. This is the dipole-dipole interaction of an atom with its image, located at distance z inside the dielectric. In the static case, the image charge distribution has a charge of $(\epsilon - \epsilon_0)/(\epsilon + \epsilon_0)$ times the original charge.[13] The expression here is the extension of that image concept to all frequencies, and the contribution to the atomic energy is weighted by the response $\alpha(\omega)$, and then the energy is summed over all frequencies.

[12] A. D. McLachlan, *op. cit.*, Eq. (3.7).

[13] See John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), p. 157.

14.3.5.6 Far-Field Limit

In the far-field limit, where z becomes very large, the values of κ, and thus s and $k_\parallel$, that contribute to the integral are very small. In particular, we can replace $\alpha(is)$ by the static polarizability

$$\alpha_0 = \alpha(0) = \sum_j \frac{2d_{j,z}^2}{\hbar\omega_{j0}}, \tag{14.204}$$

and we can replace the reflection coefficients by their dc values as well. Then the potential becomes

$$\begin{aligned} V_{\text{CP}} &= \frac{\hbar\alpha_0}{8\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2 \int_0^\infty dk_\parallel \frac{k_\parallel}{\kappa}\left[r_\perp(\theta,0) + \left(1 + \frac{2k_\parallel^2 c^2}{s^2}\right) r_\parallel(\theta,0)\right] e^{-2\kappa z} \\ &= \frac{\hbar\alpha_0}{8\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2 \int_{s/c}^\infty d\kappa \left[r_\perp(\theta,0) + \left(\frac{2c^2\kappa^2}{s^2} - 1\right) r_\parallel(\theta,0)\right] e^{-2\kappa z} \\ &= \frac{\hbar\alpha_0}{8\pi^2\epsilon_0 c^3}\int_1^\infty d\xi \, \left[r_\perp(\theta,0) + \left(2\xi^2 - 1\right) r_\parallel(\theta,0)\right] \int_0^\infty ds\, s^3\, e^{-2s\xi z/c}, \end{aligned} \tag{14.205}$$

where we have introduced $\xi = c\kappa/s$. Carrying out the s integral,

$$V_{\text{CP}} = \frac{3\hbar c\alpha_0}{64\pi^2\epsilon_0 z^4}\int_1^\infty \frac{d\xi}{\xi^4}\, \left[r_\perp(\theta,0) + \left(2\xi^2 - 1\right) r_\parallel(\theta,0)\right], \tag{14.206}$$

where again the angle θ is determined by the relation $\cos\theta = \xi$. In the limit of a perfect conductor, the reflection coefficients are replaced by -1, and thus

$$V_{\text{CP}} = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\int_1^\infty \frac{d\xi}{\xi^2}, \tag{14.207}$$

and since the integral evaluates to unity,

$$V_{\text{CP}} = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}. \tag{14.208}$$

(far field, perfect conductor)

This is precisely the far-field (retarded) Casimir–Polder potential that we obtained before in Eq. (13.60). Note, however, that the z^{-4} scaling in this regime is *universal*, as we see from Eq. (14.206): the distance scaling is independent of the material properties.

To be a bit more general, we can write out Eq. (14.207) in the case of a simple dielectric interface, in which case Eqs. (14.175) gives the reflection coefficients as

$$\begin{aligned} r_\perp(\theta,0) &= \frac{\xi - \sqrt{\epsilon/\epsilon_0 - 1 + \xi^2}}{\xi + \sqrt{\epsilon/\epsilon_0 - 1 + \xi^2}} \\ r_\parallel(\theta,0) &= \frac{\sqrt{\epsilon/\epsilon_0 - 1 + \xi^2} - \xi\epsilon/\epsilon_0}{\sqrt{\epsilon/\epsilon_0 - 1 + \xi^2} + \xi\epsilon/\epsilon_0}, \end{aligned} \tag{14.209}$$

where $\epsilon = \epsilon(0)$. Then Eq. (14.207) takes on the rather cumbersome form

$$V_{\text{CP}} = -\frac{3\hbar c\alpha_0}{64\pi^2\epsilon_0 z^4}\int_1^\infty \frac{d\xi}{\xi^4}\left[\frac{\sqrt{\chi+\xi^2} - \xi}{\sqrt{\chi+\xi^2} + \xi} + \left(1 - 2\xi^2\right)\frac{\sqrt{\chi+\xi^2} - \xi(1+\chi)}{\sqrt{\chi+\xi^2} + \xi(1+\chi)}\right], \tag{14.210}$$

where $\chi \equiv \chi(\omega = 0) = \epsilon/\epsilon_0 - 1$ is the dc susceptibility of the material. This is the Lifshitz expression for the Casimir–Polder potential for a dielectric surface, and the integral can be evaluated explicitly, with the

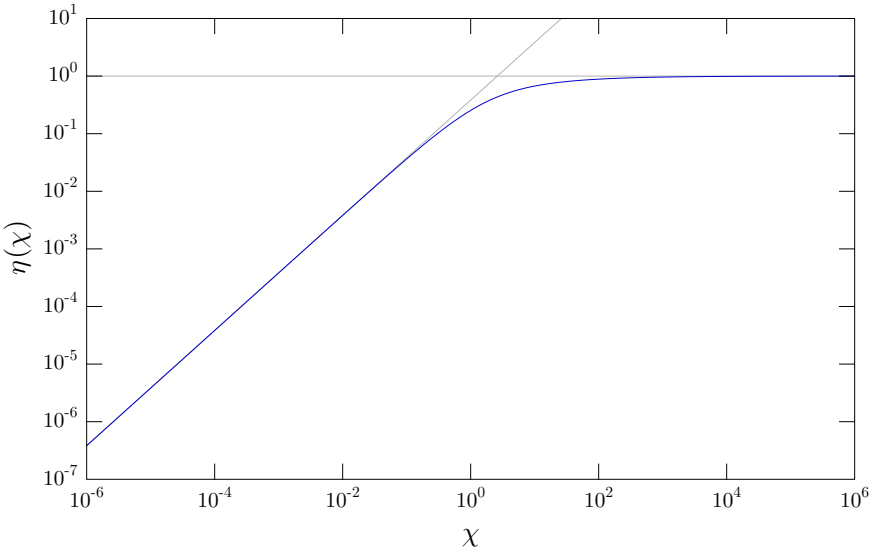
result[14]

$$V_{\text{CP}} = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\eta(\chi), \tag{14.211}$$
(far field, dielectric)

where we have defined the "efficiency"

$$\begin{aligned}\eta(\chi) := \frac{4}{3} + \chi + \frac{4-(2+\chi)\sqrt{1+\chi}}{2\chi} - \frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\left[2+\chi+2(1+\chi)\chi^2\right] \\ + \frac{(1+\chi)^2}{\sqrt{2+\chi}}\left(\sinh^{-1}\sqrt{1+\chi} - \sinh^{-1}\frac{1}{\sqrt{1+\chi}}\right),\end{aligned} \tag{14.212}$$

which ranges from 0 to 1, and measures the strength of the far-field Casimir–Polder potential compared to the perfect-conductor case, Eq. (14.208). This function is plotted below, as the strength of the Casimir–Polder potential for a surface of dc susceptibility χ, compared to a perfectly reflecting surface, $\chi \longrightarrow \infty$.



The function $\eta(\chi)$ varies smoothly with χ, and of course the potential here for finite χ is always weaker than for the perfect conductor. For small χ, $\eta(\chi) \approx (23/60)\chi$, and as $\chi \longrightarrow \infty$, $\eta(\chi) \longrightarrow 1$; both of these asymptotic forms are also plotted in the figure. For example, for synthetic fused silica (pure optical glass), the refractive index n tends to a value near 2.0 at low frequencies, so we may take $\chi \approx 1.0$ and conclude that the amplitude of the long-range Casimir–Polder potential is about 25% of that of the perfect conductor. Schott LaSFN9 glass has a refractive index tending to 4.2 near dc, so that V_{CP} in the far field is about 47% of the perfect-conductor amplitude. Note that in view of the zero-frequency pole in the permittivity for a conductor, Eq. (14.65), the limit $\chi \longrightarrow i\infty$, where $\eta(\chi) \longrightarrow 1$, is appropriate even for an imperfect conductor.

The efficiency (14.212) can also be separated into the relative contributions by the TE ($\perp$) and TM ($\parallel$) polarizations as

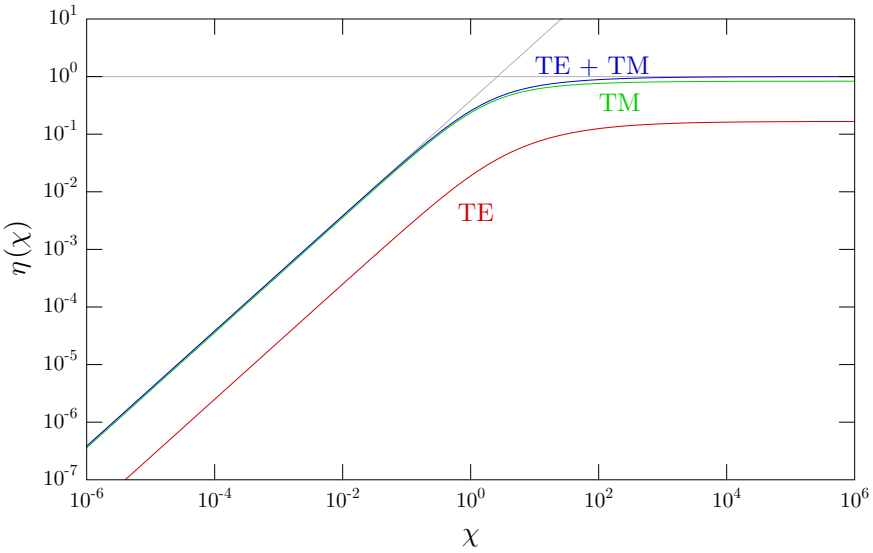
$$\eta(\chi) = \eta_{\text{TE}}(\chi) + \eta_{\text{TM}}(\chi), \tag{14.213}$$

where

$$\begin{aligned}\eta_{\text{TE}}(\chi) &:= \frac{1}{6} + \frac{1}{\chi} - \frac{\sqrt{1+\chi}}{2\chi} - \frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}} \\ \eta_{\text{TM}}(\chi) &:= \frac{7}{6} + \chi + \frac{2-(1+\chi)\sqrt{1+\chi}}{2\chi} - \frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\left[1+\chi+2(1+\chi)\chi^2\right] \\ &\quad + \frac{(1+\chi)^2}{\sqrt{2+\chi}}\left(\sinh^{-1}\sqrt{1+\chi} - \sinh^{-1}\frac{1}{\sqrt{1+\chi}}\right),\end{aligned} \tag{14.214}$$

[14] E. M. Lifshitz, "The Theory of Molecular Attractive Forces between Solids," *Soviet Physics JETP* **2**, 73 (1956); for the explicit form, see I. E. Dzyaloshinskii, E. M. Lifshitz, L. P. Pitaevskii, "The general theory of van der Waals forces," *Advances in Physics* **10**, 165 (1961), Eqs. (4.37)-(4.38).

by integrating separately the two terms in Eq. (14.210). The relative contributions are plotted below, compared to the total.



Interestingly, the contribution from the TM polarization is much larger than from the TE polarization—note that this is true even though the reflection coefficient for TM polarization is always smaller for the same angle than for TE polarization (the reflection coefficient for TM polarization vanishes at Brewster's angle, for example). This is evidently due to the importance of the weighting factor $(1+2k_\tau^2c^2/s^2)$ for the TM contribution in Eq. (14.186). In the perfect-conductor limit $\chi \longrightarrow \infty$, the relative contributions are $\eta_{\text{TE}}(\chi \longrightarrow \infty) = 1/6$ and $\eta_{\text{TM}}(\chi \longrightarrow \infty) = 5/6$. In the rarefied-dielectric limit of small χ, For small χ, $\eta(\chi) \approx (23/60)\chi$, $\eta_{\text{TE}}(\chi) \approx \chi/40$ and $\eta_{\text{TM}}(\chi) \approx (43/120)\chi$.

14.3.5.7 Asymmetric Atoms and Molecules

Going back to Eq. (14.185)

$$V_{\text{CP}} = \frac{\hbar}{8\pi^2\epsilon_0}\int_0^\infty ds \int_0^\infty dk_\tau \, \frac{k_\tau}{\kappa}\left[\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}\left(\kappa^2 r_\parallel(\theta,is)+\frac{s^2}{c^2}r_\perp(\theta,is)\right)+\alpha_{zz}(is)k_\tau^2 r_\parallel(\theta,is)\right]e^{-2\kappa z}, \tag{14.215}$$

let us consider this for asymmetric atoms and molecules, instead of spherically symmetric atoms as in Section 14.3.5.3. Putting Eq. (14.215) into symmetric and asymmetric parts, using $ax+by = (a+b)(x+y)/2 + (a-b)(x-y)/2$, gives

$$\begin{aligned} V_{\text{CP}} = {} & \frac{\hbar}{8\pi^2\epsilon_0}\int_0^\infty ds \int_0^\infty dk_\tau \, \frac{k_\tau}{\kappa}\left[\frac{1}{2}\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}+\alpha_{zz}(is)\right)\left(\kappa^2 r_\parallel+\frac{s^2}{c^2}r_\perp+k_\tau^2 r_\parallel\right)\right]e^{-2\kappa z} \\ & + \frac{\hbar}{8\pi^2\epsilon_0}\int_0^\infty ds \int_0^\infty dk_\tau \, \frac{k_\tau}{\kappa}\left[\frac{1}{2}\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}-\alpha_{zz}(is)\right)\left(\kappa^2 r_\parallel+\frac{s^2}{c^2}r_\perp-k_\tau^2 r_\parallel\right)\right]e^{-2\kappa z}. \end{aligned} \tag{14.216}$$

Cleaning up this expression a bit gives

$$\begin{aligned} V_{\text{CP}} = {} & \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}+\alpha_{zz}(is)\right)\int_0^\infty dk_\tau \, \frac{k_\tau}{\kappa}\left[\left(r_\perp+\left(1+\frac{2k_\tau^2c^2}{s^2}\right)r_\parallel\right)\right]e^{-2\kappa z} \\ & + \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}-\alpha_{zz}(is)\right)\int_0^\infty dk_\tau \, \frac{k_\tau}{\kappa}(r_\perp+r_\parallel)e^{-2\kappa z}. \end{aligned} \tag{14.217}$$

(Casimir–Polder potential, asymmetric atom)

The first term is the symmetric potential that we derived earlier, and the second term is due to the asymmetric contribution of the atom.

Now consider a perfectly reflecting surface ($r = -1$), but consider only the asymmetric term:

$$V_{\rm CP} = -\frac{\hbar}{8\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2 \left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2} - \alpha_{zz}(is)\right)\int_0^\infty dk_\tau\, \frac{k_\tau}{\kappa} e^{-2\kappa z}. \tag{14.218}$$

The polarizability is (with no sum over γ)

$$\alpha_{\gamma\gamma}(is) = \sum_j \frac{2k_{j0}|\langle {\rm g}|d_\gamma|{\rm e}_j\rangle|^2}{\hbar c(s^2/c^2 + k_{j0}^2)}, \tag{14.219}$$

and so the potential becomes[15]

$$\begin{aligned}
V_{\rm CP} &= -\frac{1}{4\pi^2\epsilon_0 c^3}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\int_0^\infty ds\, s^2 \frac{k_{j0}}{s^2/c^2 + k_{j0}^2}\int_{s/c}^\infty d\kappa\; e^{-2\kappa z} \\
&= -\frac{1}{8\pi^2\epsilon_0 c^3 z}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\int_0^\infty ds\, s^2 \frac{k_{j0}}{s^2/c^2 + k_{j0}^2}\, e^{-2sz/c} \\
&= -\frac{1}{8\pi^2\epsilon_0 z}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\int_0^\infty ds'\, s'^2 \frac{k_{j0}}{s'^2 + k_{j0}^2}\, e^{-2s'z} \\
&= -\frac{1}{32\pi^2\epsilon_0 z}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\partial_z^2\int_0^\infty ds'\, \frac{k_{j0}}{s'^2 + k_{j0}^2}\, e^{-2s'z} \\
&= -\frac{1}{32\pi^2\epsilon_0 z}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\partial_z^2\, f(2k_{j0}z).
\end{aligned} \tag{14.220}$$

This is equivalent to what we had before in Eq. (13.33), but this sum includes the spurious $1/z^2$ contribution that vanishes under the Thomas–Reiche–Kuhn sum rule.

Thus, consider

$$V_{\rm CP}^0 = \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2 \int_0^\infty dk_\tau\, \frac{k_\tau}{\kappa}\left(\frac{\left[\alpha_{xx}^>(is)+\alpha_{yy}^>(is)\right]}{2} - \alpha_{zz}^>(is)\right)\big(r_\perp(\infty) + r_\parallel(\infty)\big)\, e^{-2\kappa z} \tag{14.221}$$

where $\alpha^>$ is the same as α, but in the asymptotic limit of large s, and the reflection coefficients are evaluated with the indices of refraction at infinite frequency. This becomes

$$\begin{aligned}
V_{\rm CP}^0 &= \frac{1}{8\pi^2\epsilon_0 c^3}\sum_j \left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\big(r_\perp(\infty) + r_\parallel(\infty)\big)\int_0^\infty ds\, s^2 \frac{k_{j0}}{s^2/c^2}\int_{s/c}^\infty d\kappa\; e^{-2\kappa z} \\
&= \frac{1}{8\pi^2\epsilon_0 c}\sum_j k_{j0}\left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\big(r_\perp(\infty) + r_\parallel(\infty)\big)\int_0^\infty ds\int_{s/c}^\infty d\kappa\; e^{-2\kappa z} \\
&= \frac{1}{16\pi^2\epsilon_0 cz}\sum_j k_{j0}\left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\big(r_\perp(\infty) + r_\parallel(\infty)\big)\int_0^\infty ds\; e^{-2sz/c} \\
&= \frac{1}{32\pi^2\epsilon_0 z^2}\sum_j k_{j0}\left(\frac{d_{\parallel,j}^2}{2} - d_{\perp,j}^2\right)\big(r_\perp(\infty) + r_\parallel(\infty)\big)
\end{aligned} \tag{14.222}$$

This expression vanishes due to the TRK sum rule. Thus the first expression (14.221) can be subtracted from the initial expression (14.217). We see here that the last expression is an artifact of a double-limit at infinity, because the reflection coefficients should go to zero at $\omega \longrightarrow \infty$.

[15] After using the integral formula from I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 7th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2007), Formula 3.354.1, or Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), Eq. (5.2.12).

14.3.5.8 Dielectric Thin Films

Suppose we consider the ground-state interaction of an atom with a thin dielectric film, surrounded on either side by vacuum.[16] The film is described by the reflection coefficient[17]

$$r_{\text{film}}^{(\parallel,\perp)}(\theta, is) = \frac{r_{\parallel,\perp}(\theta, is)\,(1 - e^{i\phi})}{1 - r_{\parallel,\perp}^2(\theta, is)\,e^{i\phi}} \tag{14.223}$$

for the two polarizations in terms of the appropriate Fresnel coefficients, where the round-trip phase ϕ in the film is given by

$$\phi = 2kd\sqrt{n^2(is) - \sin^2\theta}. \tag{14.224}$$

The other parameters here are the refractive index $n(\theta, is)$ of the film and the film thickness d. For a very thin film, we may expand the above expression to first order in the film thickness, with the result

$$r_{\text{film}}^{(\parallel,\perp)}(\theta, is) = -i\frac{r_{\parallel,\perp}\,\phi}{1 - r_{\parallel,\perp}^2}, \tag{14.225}$$

assuming $|r_{\parallel,\perp}| \neq 1$. The Casimir–Polder potential is then given by Eq. (14.186), with the reflection coefficients given by Eq. (14.225) for thin films, or by by Eq. (14.223) in the more general case. Here, we will only consider the limit of very thin films. Obviously, this situation is much more complicated than for a simple dielectric interface, and so we will again concentrate on evaluating the near- and far-field limits.

In the near-field regime,

$$\phi = 2kd\sqrt{n^2(is) - \sin^2\theta} = i\frac{2sd}{c}\sqrt{n^2(is) + \frac{c^2k_\perp^2}{s^2}} \approx i2k_\perp d \approx i2\kappa d. \tag{14.226}$$

Recalling that $\kappa \lesssim 1/2z$, the thin-film approximation will be good if $d \ll z$. Thus, this treatment will be valid for atom–surface distances much smaller than any resonance wavelength, but still far enough away that the film appears to be thin. Physically, because the evanescent modes are most important in the near field, evanescent modes whose skin depths are of the order of the film thickness will be modified appropriately. The film reflection coefficients scale to lowest order as $r_{\parallel,\perp}$, so just as in the case of the simple dielectric interface, we will ignore the small contribution of $r_\perp$ compared to $r_\parallel$. Then we arrive at the same result (14.202), but with the replacement

$$r_\parallel \approx \frac{\epsilon_0 - \epsilon(is)}{\epsilon_0 + \epsilon(is)} \longrightarrow \frac{r_\parallel}{1 - r_\parallel^2}2\kappa d. \tag{14.227}$$

Explicitly, then, we have

$$V_{\text{CP}} \approx \frac{\hbar d}{2\pi^2\epsilon_0}\int_0^\infty ds\,\alpha(is)\,\frac{r_\parallel(is)}{1 - r_\parallel^2(is)}\int_0^\infty d\kappa\,\kappa^3\,e^{-2\kappa z}. \tag{14.228}$$

After evaluating the second integral, we have the result

$$V_{\text{CP}} \approx \frac{3\hbar d}{16\pi^2\epsilon_0 z^4}\int_0^\infty ds\,\alpha(is)\,\frac{r_\parallel(is)}{1 - r_\parallel^2(is)}. \tag{14.229}$$

Writing out the Fresnel reflection coefficient explicitly, e.g., as in Eq. (14.227), we finally have

$$V_{\text{CP}} \approx -\frac{3\hbar d}{64\pi^2\epsilon_0 z^4}\int_0^\infty ds\,\alpha(is)\left[\frac{\epsilon(is)}{\epsilon_0} - \frac{\epsilon_0}{\epsilon(is)}\right]. \tag{14.230}$$

(dielectric thin film, near field)

[16] The results here were derived by Yu. S. Barash, "Van der Waals interaction of thin conducting layers," *Soviet Physics—Solid State* **30**, 1578 (1988), Eqs. (25)-(26); and Fei Zhou and Larry Spruch, "van der Waals and retardation (Casimir) interactions of an electron or an atom with multilayered walls," *Physical Review A* **52**, 297 (1995), Eqs. (4.57)-(4.58) (doi: 10.1103/PhysRevA.52.297).

[17] Daniel A. Steck, *Classical and Modern Optics*, available online at `http://steck.us/teaching`.

Thus, we see that the thin film has a different *scaling* of z^{-4} in the near field, as compared to the z^{-3} scaling for the bulk material. The dependence on the dielectric material has been modified as well.

In the far field of a thin dielectric film, we again consider the limit of $\omega = is \longrightarrow 0$ in the response functions. However, in this limit, the thin-film phase ϕ defined in Eq. (14.224) also vanishes, leading to vanishing thin-film reflection coefficients, as in Eq. (14.223). Thus, for the thin-film reflection coefficients, we should keep them to lowest nonvanishing order in s:

$$\phi = 2kd\sqrt{n^2(is) - \sin^2\theta} \approx i\frac{2sd}{c}\sqrt{n^2(0) - 1 + \xi^2} = i\frac{2sd}{c}\sqrt{\chi + \xi^2}. \tag{14.231}$$

Again, we are using the notation $\xi = c\kappa/s = \cos\theta$, and χ is the dc susceptibility of the film medium. Then our analysis from Section 14.3.5.6 carries through with the replacements

$$r_{\parallel,\perp} \longrightarrow -i\frac{r_{\parallel,\perp}}{1 - r_{\parallel,\perp}^2}\phi = \frac{r_{\parallel,\perp}}{1 - r_{\parallel,\perp}^2}\frac{2sd}{c}\sqrt{\chi + \xi^2}. \tag{14.232}$$

Making this replacement in Eq. (14.205), we find

$$V_{\text{CP}} = \frac{\hbar\alpha_0 d}{4\pi^2\epsilon_0 c^4}\int_1^\infty d\xi\,\sqrt{\chi + \xi^2}\left[\frac{r_\perp(\theta,0)}{1 - r_\perp^2(\theta,0)} + \left(2\xi^2 - 1\right)\frac{r_\parallel(\theta,0)}{1 - r_\parallel^2(\theta,0)}\right]\int_0^\infty ds\,s^4\,e^{-2s\xi z/c}. \tag{14.233}$$

The final integral is easy to evaluate, with the result

$$V_{\text{CP}} = \frac{3\hbar c\alpha_0 d}{16\pi^2\epsilon_0 z^5}\int_1^\infty d\xi\,\frac{\sqrt{\chi + \xi^2}}{\xi^5}\left[\frac{r_\perp(\theta,0)}{1 - r_\perp^2(\theta,0)} + \left(2\xi^2 - 1\right)\frac{r_\parallel(\theta,0)}{1 - r_\parallel^2(\theta,0)}\right]. \tag{14.234}$$

(dielectric thin film, far field)

Of course, we could write the integrand out explicitly in terms of ξ, eliminating θ, using the far-field expressions (14.209). However, we won't be quite so masochistic right now. The important thing to notice from this relation is again, the different *scaling* of the far-field potential with distance due to a thin, dielectric film is z^{-5}, compared to the z^{-4} scaling that we derived for a bulk material, whether a perfect conductor or a dielectric.

14.3.5.9 Metallic Thin Films

For a thin metallic film,[18] we have to modify the above calculation, because now the permittivity may become quite large in relevant frequency ranges due to the existence of the dc pole. Furthermore, we can take as an explicit model of the metal's response the Drude–Lorentz model (Problem 1.4)

$$\frac{\epsilon(is)}{\epsilon_0} = 1 + \frac{\omega_{\text{p}}^2}{s(s + \gamma)}, \tag{14.235}$$

where ω_{p} is the plasma frequency, and γ is a material damping rate, given by $\gamma = \epsilon_0\omega_{\text{p}}^2/\sigma_0$, where σ_0 is the dc conductivity.

In the near field, we may take the limit $\gamma \longrightarrow 0$, as the relevant frequencies that contribute to the result are of the order of the atomic resonances, which are generally much larger than γ. Thus, we will in fact use the plasma model for the metal,

$$\frac{\epsilon(is)}{\epsilon_0} = 1 + \frac{\omega_{\text{p}}^2}{s^2}. \tag{14.236}$$

As we argued in the dielectric case, the near field corresponds to $k_{\text{T}} \gg s/c$. Furthermore, we will note that $n^2 s^2/c^2 = (s^2 + \omega_{\text{p}}^2)/c^2$. But we will consider the case where the plasma frequency is of comparable

[18] These results were derived by Yu. S. Barash, *op. cit.*, Eqs. (28) and (30); and M. Boström and Bo E. Sernelius, "van der Waals energy of an atom in the proximity of thin metal films," *Physical Review A* **61**, 052703 (2000), Eqs. (21) and (23) (doi: 10.1103/PhysRevA.61.052703).

magnitude to the atomic resonance frequencies, which is reasonable for real atoms and metals, so that $\omega \sim \omega_p$. Thus, $k_\tau^2 \gg n^2 s^2/c^2$, just as in the dielectric case. In particular, this implies that as before, we can write the Fresnel reflection coefficients as $r_\perp \approx 0$ and

$$r_\parallel \approx \frac{\epsilon_0 - \epsilon(is)}{\epsilon_0 + \epsilon(is)} = -\frac{\omega_p^2}{2s^2 + \omega_p^2}. \tag{14.237}$$

Then we can use Eq. (14.223) for the thin-film reflection coefficient to write $r_{\text{film}}^{(\perp)} = 0$ and

$$r_{\text{film}}^{(\parallel)}(\theta, is) = \frac{r_\parallel \left(1 - e^{-2k_\tau d}\right)}{1 - r_\parallel^2\, e^{-2k_\tau d}} = -\frac{\omega_p^2(2s^2 + \omega_p^2)\left(1 - e^{-2k_\tau d}\right)}{(2s^2 + \omega_p^2)^2 - \omega_p^4 e^{-2k_\tau d}}, \tag{14.238}$$

where we used the film round-trip phase $\phi = i2d\sqrt{n^2 s^2/c^2 + k_\tau^2} \approx i2k_\tau d$, and again d is the film thickness. Note that we are not yet expanding this expression in d: unlike the dielectric case, the Fresnel reflection coefficient can be close to unity, leading to an unphysical divergence in the thin-film reflection coefficient unless we keep factors of $e^{-2k_\tau d}$ around for now. We can put these bits into Eq. (14.186), replacing the polarizability by the dc value $\alpha(is) \longrightarrow \alpha_0$, since for a very thin film, the reflection coefficient will be small except near zero frequency due to the dc pole in $\epsilon(is)$. We thus obtain

$$V_{\text{CP}} = -\frac{\hbar\omega_p^2\alpha_0}{4\pi^2\epsilon_0}\int_0^\infty d\kappa\,\kappa^2 e^{-2\kappa z}\left(1 - e^{-2\kappa d}\right)\int_0^\infty ds\,\frac{2s^2 + \omega_p^2}{(2s^2 + \omega_p^2)^2 - \omega_p^4 e^{-2\kappa d}}, \tag{14.239}$$

after changing the integration variable k_τ to κ (with $k_\tau \approx \kappa$ as before). Recall [Eq. (14.184)] that κ and s are related by $\kappa^2 = s^2/c^2 + k_\tau^2$, but they are independent as far as integration variables are concerned, since s represents the (imaginary) optical "frequency," whereas κ (or k_τ) parameterizes the (independent) wave vector, whose magnitude is unconstrained when we consider the entire family of propagating and evanescent modes. Now we can evaluate the second integral as

$$\begin{aligned}\int_0^\infty ds\,\frac{2s^2 + \omega_p^2}{(2s^2 + \omega_p^2)^2 - \omega_p^4 e^{-2\kappa d}} &= \frac{\pi}{\sqrt{8}\,\omega_p}\,\frac{1 + \sqrt{1 - e^{-2\kappa d}}}{\sqrt{1 - e^{-2\kappa d}}\left(\sqrt{1 - e^{-\kappa d}} + \sqrt{1 + e^{-\kappa d}}\right)} \\ &\approx \frac{\pi}{4\sqrt{2\kappa d}\,\omega_p},\end{aligned} \tag{14.240}$$

where we have kept only the lowest-order term in d in the last expression. Thus, taking the $d \longrightarrow 0$ limit in the rest of Eq. (14.239), we find

$$V_{\text{CP}} = -\frac{\hbar\omega_p\alpha_0\sqrt{d}}{8\sqrt{2}\,\pi\epsilon_0}\int_0^\infty d\kappa\,\kappa^{5/2}\,e^{-2\kappa z}. \tag{14.241}$$

After evaluating the final integral, we find the result

$$V_{\text{CP}} = -\frac{15\hbar\omega_p\alpha_0\sqrt{d}}{1024\sqrt{\pi}\epsilon_0 z^{7/2}}. \tag{14.242}$$

(metallic thin film, near field)

Thus, for a metallic thin film, the main difference with the case of a bulk metal or dielectric is the *fractional-power* scaling of $z^{-7/2}$, which is faster than z^{-3} for the bulk case and slower than z^{-4} for the dielectric-thin-film case. The other obvious difference from the case of the thin dielectric film is the $\sqrt{d}$ scaling with the metallic-film thickness, as opposed to the linear scaling in the dielectric case.

The far-field case of a metallic film is fairly easy, by comparison. Again, the conductivity leads to a dc pole in $\epsilon(\omega)$, which dominates the far-field response. Thus, in the far-field limit of a metallic film, the Casimir–Polder potential is given by the usual perfect-conductor expression (14.208):

$$V_{\text{CP}} = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}. \tag{14.243}$$

(metallic thin film, far field)

Physically, this is because as the relevant frequencies decrease to zero, so does the depth of penetration of the field modes into the film, and so the thinness of the film becomes irrelevant.

14.3.6 Perfectly Conducting, Spherical Cavity

A geometrically more challenging example of the Casimir–Polder potential is the potential felt by an atom within a metallic, spherical cavity. For simplicity we consider only the limit of perfect conductivity. Inside the spherical cavity, we can use the mode-expansion formula (14.162) for the Green tensor and the appropriate mode functions from Section (8.4.4) to obtain the Green tensor

$$\mathrm{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] = \frac{1}{\epsilon_0}\sum_{nlm}\frac{\omega_{nl}^2}{\omega_{nl}^2-\omega^2} f^{(\mathrm{TE})}_{nlm,\alpha}(\mathbf{r}) f^{(\mathrm{TE})*}_{nlm,\beta}(\mathbf{r}') + \frac{1}{\epsilon_0}\sum_{nlm}\frac{\omega_{nl}^2}{\omega_{nl}^2-\omega^2} f^{(\mathrm{TM})}_{nlm,\alpha}(\mathbf{r}) f^{(\mathrm{TM})*}_{nlm,\beta}(\mathbf{r}'),$$

(spherical-cavity Green tensor) (14.244)

where the TE and TM modes are given from Eqs. (8.111) and (8.118) in terms of spherical Bessel functions and vector spherical harmonics by

$$\begin{aligned}
\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r}) &= \mathcal{N}^{(\mathrm{TE})}_{nl} j_l(k_{nl}r)\,\mathbf{X}^m_l(\theta,\phi) && \text{where } j_l(k_{nl}R)=0,\\
\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r}) &= \frac{\mathcal{N}^{(\mathrm{TM})}_{nl}}{k_{nl}}\nabla\times\left[j_l(k_{nl}r)\,\mathbf{X}^m_l(\theta,\phi)\right], && \text{where } \partial_r\left[r\,j_l(k_{nl}r)\right]\big|_{r=R}=0,
\end{aligned}$$

(spherical-cavity modes) (14.245)

with normalization factors

$$\begin{aligned}
\mathcal{N}^{(\mathrm{TE})}_{nl} &= \left[\frac{R^3}{2} j^2_{l+1}(k_{nl}R)\right]^{-1/2}\\
\mathcal{N}^{(\mathrm{TM})}_{nl} &= \left[\frac{R^3}{2}\left(1-\frac{l(l+1)}{k_{nl}^2R^2}\right) j_l^2(k_{nl}R)\right]^{-1/2}.
\end{aligned}$$

(normalization factors) (14.246)

However, this calculation in the spherical-cavity case turns out to miss a singular term $-\delta_{\alpha\beta}\delta^3(\mathbf{r}-\mathbf{r}')/\epsilon_0$.[19] However, this doesn't matter for the Casimir–Polder-type calculations below, since it corresponds to a term independent of the size of the cavity, which thus disappears after renormalization, as discussed below.

In fact, for Casimir–Polder-type calculations, what we will need is the Green tensor evalutated at $\mathbf{r}'=\mathbf{r}$ and $\omega = is$. Also, for spherically symmetric atoms we only need the trace of the Green tensor. For example, for the spherical cavity we thus need

$$G_{\alpha\alpha}(\mathbf{r},\mathbf{r},is) = \frac{1}{\epsilon_0}\sum_{nlm}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}\left|\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r})\right|^2 + \frac{1}{\epsilon_0}\sum_{nlm}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}\left|\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r})\right|^2. \tag{14.247}$$

We then write out the TE and TM squared modes using Eqs. (8.124), (8.124), and (8.108) as

$$\begin{aligned}
\left|\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r})\right|^2 &= [\mathcal{N}^{(\mathrm{TE})}_{nl}]^2 j_l^2(k_{nl}r)\left|\mathbf{X}^m_l(\theta,\phi)\right|^2\\
\left|\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r})\right|^2 &= [\mathcal{N}^{(\mathrm{TM})}_{nl}]^2\left[l(l+1)\left(\frac{j_l(k_{nl}r)}{kr}\right)^2\left|Y^m_l(\theta,\phi)\right|^2 + \left(\frac{\partial_r[rj_l(k_{nl}r)]}{kr}\right)^2\left|\mathbf{X}^m_l(\theta,\phi)\right|^2\right].
\end{aligned} \tag{14.248}$$

Now employing the sum rules (8.88) and (8.110) for the scalar and vector spherical harmonics, we can compute the sum over m as

$$\begin{aligned}
\sum_{m=-l}^{l}\left|\mathbf{f}^{(\mathrm{TE})}_{nlm}(\mathbf{r})\right|^2 &= \frac{(2l+1)}{4\pi}[\mathcal{N}^{(\mathrm{TE})}_{nl}]^2 j_l^2(k_{nl}r)\\
\sum_{m=-l}^{l}\left|\mathbf{f}^{(\mathrm{TM})}_{nlm}(\mathbf{r})\right|^2 &= \frac{(2l+1)}{4\pi}[\mathcal{N}^{(\mathrm{TM})}_{nl}]^2\left[l(l+1)\left(\frac{j_l(k_{nl}r)}{kr}\right)^2 + \left(\frac{\partial_r[rj_l(k_{nl}r)]}{kr}\right)^2\right],
\end{aligned} \tag{14.249}$$

[19] Chen-To Tai, *Dyadic Green Functions in Electromagnetic Theory*, 2nd ed. (IEEE Press, 1994), Section 10-4, pp. 218-220.

and so the Green tensor finally becomes

$$G_{\alpha\alpha}(\mathbf{r},\mathbf{r},is) = \frac{1}{4\pi\epsilon_0}\left\{\sum_{nl}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TE})}]^2 j_l^{\,2}(k_{nl}r) + \sum_{nl}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TM})}]^2\left[l(l+1)\left(\frac{j_l(k_{nl}r)}{kr}\right)^2+\left(\frac{\partial_r[rj_l(k_{nl}r)]}{kr}\right)^2\right]\right\}. \tag{14.250}$$

Recall the two summations are different the first referring to TE-mode boundary conditions, while the second refers to TM modes. The remaining sums are more cumbersome, and can be carried out numerically or by using simpler but approximate asymptotic forms.

To compute the potential, we will assume a spherically symmetric atom and use the Kramers–Heisenberg formula (14.261) and the scalar form of the polarizability $\alpha_{\mu\nu}(\omega) = \delta_{\mu\nu}\alpha(\omega)$, where

$$\alpha(\omega) = \sum_j \frac{2\omega_{j0}d_{j,z}^2}{\hbar(\omega_{j0}^2-\omega^2)}, \tag{14.251}$$

where we use the shorthand $d_{j,z} := \langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle$ for the dipole matrix elements. Now using Eq. (14.163), we can write

$$\begin{aligned} V_{\mathrm{CP}} &= -\frac{\hbar}{2\pi}\int_0^\infty ds\,\alpha(is)\,G_{\mu\mu}(\mathbf{r},\mathbf{r},is) \\ &= -\frac{1}{\pi(4\pi\epsilon_0)}\sum_j d_{j,z}^2\int_0^\infty ds\,\frac{\omega_{j0}}{\omega_{j0}^2+s^2}\left\{\sum_{nl}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TE})}]^2 j_l^{\,2}(k_{nl}r)\right. \\ &\qquad \left. + \sum_{nl}\frac{\omega_{nl}^2}{\omega_{nl}^2+s^2}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TM})}]^2\left[l(l+1)\left(\frac{j_l(k_{nl}r)}{kr}\right)^2+\left(\frac{\partial_r[rj_l(k_{nl}r)]}{kr}\right)^2\right]\right\}. \end{aligned} \tag{14.252}$$

Again using the formula (14.165) to carry out the imaginary-frequency integral,

$$V_{\mathrm{CP}} = -\frac{1}{2(4\pi\epsilon_0)}\sum_j d_{j,z}^2\left\{\sum_{nl}\frac{\omega_{nl}}{\omega_{nl}+\omega_{j0}}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TE})}]^2 j_l^{\,2}(k_{nl}r) + \sum_{nl}\frac{\omega_{nl}}{\omega_{nl}+\omega_{j0}}(2l+1)[\mathcal{N}_{nl}^{(\mathrm{TM})}]^2\left[l(l+1)\left(\frac{j_l(k_{nl}r)}{kr}\right)^2+\left(\frac{\partial_r[rj_l(k_{nl}r)]}{kr}\right)^2\right]\right\}. \tag{14.253}$$

(Casimir–Polder potential, spherical cavity)

In general, this sum must be performed numerically to obtain an answer. Furthermore, this expression is divergent; it must still be renormalized by subtracting off the same expression, but in the limit $R \longrightarrow \infty$. However, the potential turns out again to be negative and divergent as the atom nears the cavity surface, while the potential becomes weakest at the center of the sphere.[20]

14.3.7 Ground-State Atom–Atom Potentials

This formalism can handle not only the interaction of atoms with macroscopic bodies, but also with other atoms. To see this, we will consider the vacuum atom–atom interaction potential in otherwise free space. We suppose the atoms to have polarizability tensors $\alpha_{\mu\nu}^{(1)}$ and $\alpha_{\mu\nu}^{(2)}$, and without loss of generality we may assume them to be separated along the z axis at a distance r. From Eq. (14.167), we may write the interaction potential from the point of view of the first atom as

$$V_{12}(\mathbf{r}) = -\frac{\hbar}{2\pi}\int_0^\infty ds\,\alpha_{\mu\nu}^{(1)}(is)\,G_{\nu\mu}^{(\mathrm{s})}(0,0,is), \tag{14.254}$$

[20] W. Jhe and K. Jang, "Cavity quantum electrodynamics inside a hollow spherical cavity," *Physical Review A* **53**, 1126 (1996) (doi: 10.1103/PhysRevA.53.1126).

in terms of the scattering Green tensor describing the influence of the second atom, assuming atom 1 to be located at the origin.

To compute the Green tensor, we begin with the free-space Green tensor in the form of Eq. (14.50):

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\Big(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\Big)\frac{e^{ikr}}{r} - \frac{1}{\epsilon_0}\delta_{\alpha\beta}\delta^3(\mathbf{r}). \tag{14.255}$$

For our purposes, this will describe the electric field at atom 2 due to a unit dipole at the origin (i.e., atom 1). Since we will assume a nonzero separation between the two atoms, we will henceforth drop the last (contact) term in this Green tensor.

Atom 2 then responds to this field according to its polarizability, producing a dipole moment given by $\alpha^{(2)}_{\mu\gamma}(\omega)\, G^{(0)}_{\gamma\nu}(\mathbf{r},0,\omega)$. Another multiplication with the free-space Green tensor gives the field back at atom 1, and thus we may write the scattering Green tensor as

$$G^{(s)}_{\nu\mu}(0,0,\omega) = G^{(0)}_{\nu\gamma}(0,\mathbf{r},\omega)\ \alpha^{(2)}_{\gamma\beta}(\omega)\ G^{(0)}_{\beta\mu}(\mathbf{r},0,\omega). \tag{14.256}$$

Thus, the interaction potential is

$$V_{12}(\mathbf{r}) = -\frac{\hbar}{2\pi}\int_0^\infty ds\, \alpha^{(1)}_{\mu\nu}(is)\ G^{(0)}_{\nu\gamma}(0,\mathbf{r},is)\ \alpha^{(2)}_{\gamma\beta}(is)\ G^{(0)}_{\beta\mu}(\mathbf{r},0,is). \tag{14.257}$$

This result is correct, but there is a subtlety with the overall coefficient. Because the second dipole moment is induced by the first, there is an overall factor of 1/2 because this is an induced-dipole energy. However, this cancels the factor of 2 from summing over the (identical) energy shifts of both atoms. This induced-dipole factor is justified more formally later [see the second term in Eq. (14.416)]. Now putting in the form (14.255) for the free-space Green tensor, we find the resulting expression

$$V_{12}(\mathbf{r}) = -\frac{\hbar}{2\pi(4\pi\epsilon_0)^2}\int_0^\infty ds\, \alpha^{(1)}_{\mu\nu}(is)\, \alpha^{(2)}_{\gamma\beta}(is)\left[\Big(\partial_\nu\partial_\gamma - \delta_{\nu\gamma}\nabla^2\Big)\frac{e^{-sr/c}}{r}\right]\left[\Big(\partial_\beta\partial_\mu - \delta_{\beta\mu}\nabla^2\Big)\frac{e^{-sr/c}}{r}\right] \tag{14.258}$$

(atom–atom potential)

for the atom–atom potential.

14.3.7.1 Near-Field van der Waals–London Potential

If we consider this potential in the near field, we can set $e^{-sr/c} \approx 1$, so that

$$\Big(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\Big)\frac{e^{-sr/c}}{r} \approx \Big(\partial_\alpha\partial_\beta - \delta_{\alpha\beta}\nabla^2\Big)\frac{1}{r} = -\partial_\alpha\frac{r_\beta}{r^3} + 4\pi\delta^3(\mathbf{r})\,\delta_{\alpha\beta} = \frac{3r_\alpha r_\beta}{r^5} - \frac{\delta_{\alpha\beta}}{r^3} + 4\pi\delta^3(\mathbf{r})\,\delta_{\alpha\beta}. \tag{14.259}$$

Dropping the contact term in this result, the near-field potential becomes

$$V_{12}(\mathbf{r}) = -\frac{\hbar}{2\pi(4\pi\epsilon_0)^2}\left(\frac{\delta_{\nu\gamma}}{r^3} - \frac{3r_\nu r_\gamma}{r^5}\right)\left(\frac{\delta_{\beta\mu}}{r^3} - \frac{3r_\beta r_\mu}{r^5}\right)\int_0^\infty ds\, \alpha^{(1)}_{\mu\nu}(is)\, \alpha^{(2)}_{\gamma\beta}(is), \tag{14.260}$$

and then using the Kramers–Heisenberg formula (14.147),

$$\alpha_{\mu\nu}(\omega) = \sum_j \frac{2\omega_{j0}\langle \mathrm{g}|d_\mu|\mathrm{e}_j\rangle\langle \mathrm{e}_j|d_\nu|\mathrm{g}\rangle}{\hbar(\omega_{j0}^2 - \omega^2)}, \tag{14.261}$$

we may write the near-field potential as

$$V_{12}(\mathbf{r}) = -\frac{2}{\pi(4\pi\epsilon_0)^2\hbar}\sum_{j,j'}\left|\frac{\mathbf{d}^{(1)}_j\cdot\mathbf{d}^{(2)*}_{j'}}{r^3} - \frac{3\left(\mathbf{d}^{(1)}_j\cdot\mathbf{r}\right)\left(\mathbf{d}^{(2)*}_{j'}\cdot\mathbf{r}\right)}{r^5}\right|^2\int_0^\infty ds\, \frac{\omega_{j0}\,\omega_{j'0}}{(\omega_{j0}^2+s^2)(\omega_{j'0}^2+s^2)}, \tag{14.262}$$

where $\mathbf{d}_j := \langle \mathrm{g}|\mathbf{d}|\mathrm{e}_j\rangle$. Now using the integral formula (14.165), we can evaluate the integral, with the result

$$V_{12}(\mathbf{r}) = -\frac{1}{(4\pi\epsilon_0)^2 r^6} \sum_{j,j'} \frac{\left|\mathbf{d}_j^{(1)} \cdot \mathbf{d}_{j'}^{(2)*} - 3\left(\mathbf{d}_j^{(1)} \cdot \hat{r}\right)\left(\mathbf{d}_{j'}^{(2)*} \cdot \hat{r}\right)\right|^2}{\hbar(\omega_{j0} + \omega_{j'0})}$$

(near-field van der Waals–London potential) (14.263)

that the near-field potential scales at r^{-6}. Comparing this to the classical interaction energy between two dipoles,

$$V_{\mathrm{dip}} = \frac{\mathbf{d}_1 \cdot \mathbf{d}_2 - 3(\hat{r} \cdot \mathbf{d}_1)(\hat{r} \cdot \mathbf{d}_2)}{(4\pi\epsilon_0) r^3}, \tag{14.264}$$

we see a similar dependence on the dipole, but the present interaction behaves more like V_{dip}^2. This is because the ground-state interaction is an interaction of atom 1 with the dipole of atom 2 *induced* by atom 1's ground-state fluctuations. Hence, the much weaker interaction at long ranges. In the simpler case of identical, isotropic atoms, where the dipole matrix elements are independent of direction, the result (14.263) becomes

$$V_{12}(\mathbf{r}) = -\frac{1}{(4\pi\epsilon_0)^2 r^6} \sum_{j,j'} \frac{|d_{z,j}|^2 |d_{z,j'}|^2}{\hbar(\omega_{j0} + \omega_{j'0})} \left(\delta_{\mu\nu} - 3\hat{r}_\mu \hat{r}_\nu\right)\left(\delta_{\nu\mu} - 3\hat{r}_\nu \hat{r}_\mu\right) = -\frac{6}{(4\pi\epsilon_0)^2 r^6} \sum_{j,j'} \frac{|d_{z,j}|^2 |d_{z,j'}|^2}{\hbar(\omega_{j0} + \omega_{j'0})} \tag{14.265}$$

where as usual $d_{z,j} \equiv \hat{z} \cdot \mathbf{d}_j$. If we assume that only one transition makes the dominant contribution to the force, then we can make a two-level-atom approximation and write the potential as[21]

$$V_{12}(\mathbf{r}) = -\frac{3|d_z|^4}{(4\pi\epsilon_0)^2 \hbar\omega_0 r^6} = -\frac{3\hbar\omega_0 \alpha_0^2}{(4\pi\epsilon_0)^2 4 r^6},$$

(near-field van der Waals–London potential, identical two-level atoms) (14.266)

where the two-level static polarizability is $\alpha_0 = 2d_z^2/\hbar\omega_0$.

Note that at *very* close separations, the dipole approximation breaks down, and the atoms should repel due to the overlap of their electron clouds. This is commonly modeled by the **Lennard–Jones potential**,[22] which is literally a kludge of adding a repulsive r^{-12} term (or some other high-order power, but r^{-12} is most common) to the r^{-6} van der Waals–London potential to model the repulsion.

14.3.7.2 Far-Field Potential

The atom–atom potential is also simple in the far-field regime where retardation effects are important. As for the atom–mirror interaction, the exponential factors in (14.258) indicate that only modes with small frequencies contribute to the potential at large separations. Thus, we may replace the atomic polarizabilities with their dc values and pull them out of the integral:

$$V_{12}(\mathbf{r}) = -\frac{\hbar\, \alpha_{\mu\nu}^{(1)}(0)\, \alpha_{\gamma\beta}^{(2)}(0)}{2\pi(4\pi\epsilon_0)^2} \int_0^\infty ds \left[\left(\partial_\nu\partial_\gamma - \delta_{\nu\gamma}\nabla^2\right)\frac{e^{-sr/c}}{r}\right]\left[\left(\partial_\beta\partial_\mu - \delta_{\beta\mu}\nabla^2\right)\frac{e^{-sr/c}}{r}\right] \tag{14.267}$$

Then we can evaluate the integral, with the result

$$V_{12}(\mathbf{r}) = -\frac{\hbar c\, \alpha_{\mu\nu}^{(1)}(0)\, \alpha_{\gamma\beta}^{(2)}(0)}{2\pi(4\pi\epsilon_0)^2}\left(\partial_\nu\partial_\gamma - \delta_{\nu\gamma}\nabla^2\right)\left(\partial'_\beta\partial'_\mu - \delta_{\beta\mu}\nabla'^2\right)\left.\frac{1}{rr'(r+r')}\right|_{r'=r}, \tag{14.268}$$

[21] This result originally derived by F. London, "Über einige Eigenshaften und Anwendungen der Molekularkräfte," *Zeitschrift für Physikalische Chemie* **11**, 222 (1930), Eq. (6). For the more general short-range expression (14.263), see H. B. G. Casimir and D. Polder, "The Influence of Retardation on the London-van der Waals Forces," *Physical Review* **73**, 360 (1948) (doi: 10.1103/PhysRev.73.360), Eq. (51); and A. D. McLachlan, "Retarded dispersion forces between molecules," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **271**, 387 (1963), Eq. (6.6).

[22] J. E. Lennard-Jones, "Cohesion," *Proceedings of the Physical Society* **43**, 461 (1931) (doi: 10.1088/0959-5309/43/5/301).

where the normal derivatives act on r only, the primed derivatives act on r' only, and r' is set to r after the derivatives are evaluated. To continue, we specialize to the case of scalar polarizabilities, $\alpha_{\mu\nu} = \alpha\delta_{\mu\nu}$, so that

$$\begin{aligned} V_{12}(\mathbf{r}) &= -\frac{\hbar c\,\alpha_0^{(1)}\alpha_0^{(2)}}{2\pi(4\pi\epsilon_0)^2}\left(\partial_\mu\partial_\nu - \delta_{\mu\nu}\nabla^2\right)\left(\partial'_\nu\partial'_\mu - \delta_{\nu\mu}\nabla'^2\right)\frac{1}{rr'(r+r')}\bigg|_{r'=r} \\ &= -\frac{\hbar c\,\alpha_0^{(1)}\alpha_0^{(2)}}{2\pi(4\pi\epsilon_0)^2}\left(\partial_\mu\partial'_\mu\partial_\nu\partial'_\nu + \nabla^2\nabla'^2\right)\frac{1}{rr'(r+r')}\bigg|_{r'=r}. \end{aligned} \tag{14.269}$$

The derivatives are cumbersome but easy to carry out with the help of a computer, with the result[23]

$$V_{12}(\mathbf{r}) = -\frac{23\hbar c\,\alpha_0^{(1)}\alpha_0^{(2)}}{(4\pi\epsilon_0)^2 4\pi r^7}.$$

(retarded atom–atom potential, spherically symmetric atoms) (14.270)

Thus, in the far field, where retardation is important, the atom–atom potential scales as r^{-7}, compared to the near-field r^{-6} dependence. This is similar to the case of an atom near a planar mirror, where retardation caused the near-field power-law dependence to gain an extra power (there, from r^{-3} to r^{-4}).

14.3.7.3 General Form for Scalar Polarizabilities

To evaluate the atom–atom interaction potential more generally, we can write out Eq. (14.258) in the case of scalar polarizabilities,

$$\begin{aligned} V_{12}(\mathbf{r}) &= -\frac{\hbar}{2\pi(4\pi\epsilon_0)^2}\left[\left(\partial_\mu\partial'_\mu\partial_\nu\partial'_\nu + \nabla^2\nabla'^2\right)\frac{1}{rr'}\int_0^\infty ds\,\alpha^{(1)}(is)\,\alpha^{(2)}(is)\,e^{-s(r+r')/c}\right]_{r'=r} \\ &= -\frac{2}{9\pi\hbar(4\pi\epsilon_0)^2}\sum_{jj'}|d_j^{(1)}d_{j'}^{(2)}|^2\omega_{j0}\omega_{j'0}\left[\left(\partial_\mu\partial'_\mu\partial_\nu\partial'_\nu + \nabla^2\nabla'^2\right)\frac{1}{rr'}\int_0^\infty ds\,\frac{e^{-s(r+r')/c}}{(s^2+\omega_{j0}^2)(s^2+\omega_{j'0}^2)}\right]_{r'=r}, \end{aligned} \tag{14.271}$$

where $d_j = \langle \mathrm{g}|d|\mathrm{e}_j\rangle = \sqrt{3}\langle \mathrm{g}|d_z|\mathrm{e}_j\rangle$. Thus, we need to evaluate an integral of the form

$$\begin{aligned} I(\omega,\omega') &= \int_0^\infty ds\,\frac{e^{-s(r+r')/c}}{(s^2+\omega^2)(s^2+\omega'^2)} \\ &= \frac{1}{\omega^2-\omega'^2}\int_0^\infty ds\,\left[\frac{1}{s^2+\omega'^2} - \frac{1}{s^2+\omega^2}\right]e^{-s(r+r')/c} \\ &= \frac{\omega f[k'(r+r')] - \omega' f[k(r+r')]}{\omega\omega'(\omega^2-\omega'^2)}, \end{aligned} \tag{14.272}$$

where $k = \omega/c$, $k' = \omega'/c$, and we have used the integral formula (Problem 13.1)[24]

$$\int_0^\infty dx\,\frac{e^{-\mu x}}{\beta^2+x^2} = \frac{f(\beta\mu)}{\beta} \qquad (\mathrm{Re}[\beta] > 0, \mathrm{Re}[\mu] > 0), \tag{14.273}$$

and $f(z)$ is an auxiliary function to the sine and cosine integrals [see Eq. (13.26)]. Note that $I(\omega,\omega')$ has a removable singularity at $\omega = \omega'$, so that we may write

$$I(\omega,\omega) = \frac{f[k(r+r')] + k(r+r')\,g[k(r+r')]}{2\omega^3}. \tag{14.274}$$

Thus, the atom–atom potential finally becomes

$$V_{12}(\mathbf{r}) = -\frac{2}{9\pi\hbar(4\pi\epsilon_0)^2}\sum_{jj'}|d_j^{(1)}d_{j'}^{(2)}|^2\omega_{j0}\omega_{j'0}\left[\left(\partial_\mu\partial'_\mu\partial_\nu\partial'_\nu + \nabla^2\nabla'^2\right)\frac{1}{rr'}I(\omega_{j0},\omega_{j'0})\right]_{r'=r}.$$

(scalar, ground-state atom–atom potential) (14.275)

[23] H. B. G. Casimir and D. Polder, *op. cit.*, Eq. (56); A. D. McLachlan, *op. cit.*, Eq. (6.10).

[24] See I. S. Gradstein and I. M. Ryzhik, *Table of Integrals, Series, and Products*, English translation 6th ed., A. Jeffrey and D. Zwillinger, Eds. (Academic Press, 2000), Formula 3.354.1.

In principle we have obtained the full potential, in terms of analytic functions and derivatives. The derivatives are unfortunately cumbersome. However, we can see that we recover our former results. For example, in the far field, we can use the large-z forms $f(z) \sim 1/z$ and $g(z) \sim 1/z^2$ so that

$$I(\omega, \omega) = \frac{c}{\omega^4 (r + r')}. \tag{14.276}$$

Using this result and restricting to a single, dominant resonance while using $\alpha_0 = 2d^2/3\hbar\omega_0$, we recover the far-field result in the form (14.269) from the general form (14.275). For small separations, we can set $r = r' = 0$ in Eq. (14.272) and use $f(0) = \pi/2$ to obtain

$$I(\omega, \omega') = \frac{\pi}{2\omega\omega'(\omega + \omega')}. \tag{14.277}$$

Putting this into Eq. (14.275), we recover the near-field result (14.265) by evaluating the derivatives in the same way as before in the near-field case.

14.3.7.4 Three-Atom Potential

Another interesting example is the Casimir–Polder potential due to three atoms. By the same reasoning leading to the two-atom potential (14.257), we obtain

$$V_{123}(\mathbf{r}_2, \mathbf{r}_3) = -\frac{3!}{3}\frac{\hbar}{2\pi}\int_0^\infty ds\, \alpha^{(1)}_{\mu\nu}(is)\, G^{(0)}_{\nu\gamma}(0, \mathbf{r}_2, is)\, \alpha^{(2)}_{\gamma\beta}(is)\, G^{(0)}_{\beta\delta}(\mathbf{r}_3, \mathbf{r}_2, is)\, \alpha^{(3)}_{\delta\sigma}(is)\, G^{(0)}_{\sigma\mu}(\mathbf{r}_3, 0, is), \tag{14.278}$$

for scattering from the first atom at $\mathbf{r} = 0$ to atom 3 at $\mathbf{r}_3$, to atom 2 at $\mathbf{r}_2$, and back to the first atom. The factor of $1/3$ is because this is an induced-dipole energy. This induced-dipole factor is justified more formally later [see the thirdterm in Eq. (14.416)]. The factor of 3! counts the permutations of the three atoms (all of which are equivalent as far as the energy is concerned). Using Eq. (14.255) again for the free-space Green tensor, we find the resulting expression

$$\begin{aligned} V_{123}(\mathbf{r}_2, \mathbf{r}_3) = &-\frac{\hbar\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{\pi(4\pi\epsilon_0)^3} \\ &\times \int_0^\infty ds\, \left[\left(\partial_\mu\partial_\nu - \delta_{\mu\nu}\nabla^2\right)\frac{e^{-sr_2/c}}{r_2}\right]\left[\left(\partial_\nu\partial_\gamma - \delta_{\nu\gamma}\nabla^2\right)\frac{e^{-sr_{32}/c}}{r_{32}}\right]\left[\left(\partial_\gamma\partial_\mu - \delta_{\gamma\mu}\nabla^2\right)\frac{e^{-sr_3/c}}{r_3}\right] \end{aligned} \tag{14.279}$$

for the three-atom potential, where we are simplifying the calculation by assuming scalar polarizibilities for the atoms, and we are considering the far-field limit, where only the dc limit of the atomic polarizibilities matter. Note also that $\mathbf{r}_{32} := \mathbf{r}_3 - \mathbf{r}_2$. Carrying out the s integral, we obtain

$$V_{123}(\mathbf{r}_2, \mathbf{r}_3) = -\frac{\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{\pi(4\pi\epsilon_0)^2}\left(\partial_\mu\partial_\nu - \delta_{\mu\nu}\nabla^2\right)\left(\partial'_\nu\partial'_\gamma - \delta_{\nu\gamma}\nabla'^2\right)\left(\partial''_\gamma\partial''_\mu - \delta_{\gamma\mu}\nabla''^2\right)\frac{1}{rr'r''(r + r' + r'')}\bigg|_{\substack{\mathbf{r}=\mathbf{r}_2 \\ \mathbf{r}'=\mathbf{r}_{32} \\ \mathbf{r}''=\mathbf{r}_3}}. \tag{14.280}$$

Multiplying out the derivatives leads to

$$\begin{aligned} V_{123}(\mathbf{r}_2, \mathbf{r}_3) = &-\frac{\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{\pi(4\pi\epsilon_0)^3}\left(\partial''_\mu\partial_\mu\partial_\nu\partial'_\nu\partial'_\gamma\partial''_\gamma - \partial_\mu\partial'_\mu\partial_\nu\partial'_\nu\nabla''^2 - \partial'_\mu\partial''_\mu\partial'_\nu\partial''_\nu\nabla^2 - \partial''_\mu\partial_\mu\partial''_\nu\partial_\nu\nabla'^2\right) \\ &\times \frac{1}{rr'r''(r + r' + r'')}\bigg|_{\substack{\mathbf{r}=\mathbf{r}_2 \\ \mathbf{r}'=\mathbf{r}_{32} \\ \mathbf{r}''=\mathbf{r}_3}}, \end{aligned} \tag{14.281}$$

or in index-free notation,

$$V_{123}(\mathbf{r}_2,\mathbf{r}_3) = -\frac{\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{\pi(4\pi\epsilon_0)^3}\Big((\nabla''\cdot\nabla)(\nabla\cdot\nabla')(\nabla'\cdot\nabla'') - (\nabla\cdot\nabla')^2\nabla''^2 - (\nabla'\cdot\nabla'')^2\nabla^2 - (\nabla''\cdot\nabla)^2\nabla'^2\Big) \times \frac{1}{rr'r''(r+r'+r'')}\bigg|_{\substack{\mathbf{r}=\mathbf{r}_2\\ \mathbf{r}'=\mathbf{r}_{32}\\ \mathbf{r}''=\mathbf{r}_3}}. \tag{14.282}$$

At this point it is best to consider specific configurations. For example, for three atoms in a straight line with distance r between adjacent atoms, we have $\mathbf{r}_2 = r\hat{z}$, $\mathbf{r}_3 = 2r\hat{z}$, and $\mathbf{r}_{32} = r\hat{z}$, in which case evaluating the derivatives above leads to

$$V_{123}(r) = -\frac{93\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{512\pi(4\pi\epsilon_0)^2 r^{10}} = -\frac{93\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{2^9\pi(4\pi\epsilon_0)^2 r^{10}}. \tag{14.283}$$

equilateral triangle: $\mathbf{r}_2 = r\hat{z}$, $\mathbf{r}_3 = \hat{z}r/2 + \hat{y}\sqrt{3}r/2$, and $\mathbf{r}_{32} = -\hat{z}r/2 + \hat{y}\sqrt{3}r/2$

$$V_{123}(r) = -\frac{1264\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{243\pi(4\pi\epsilon_0)^2 r^{10}} = -\frac{2^4\times 79\hbar c\alpha_0^{(1)}\alpha_0^{(2)}\alpha_0^{(3)}}{3^5\pi(4\pi\epsilon_0)^2 r^{10}}. \tag{14.284}$$

Note that at the same order of accuracy there are other lower-body interactions, for example proportional to $(\alpha_0^{(1)})^3$ or $(\alpha_0^{(1)})^2\alpha_0^{(2)}$ that we are ignoring; we are only focusing on the true three-body contribution to the potential.

14.3.8 Temperature Dependence

In our treatment above, we have computed energy expectation values always respect to the ground/vacuum state of the combined atom/field system. However, we can also extend this formalism to cover other states, in particular thermal states

$$\rho = \sum_n P(n)|n\rangle\langle n|, \tag{14.285}$$

where the occupation probability of each energy eigenstate $|n\rangle$ is

$$P(n) = \frac{1}{Z}e^{-E_n/k_\mathrm{B}T}, \tag{14.286}$$

where

$$Z := \sum_n e^{-E_n/k_\mathrm{B}T} \tag{14.287}$$

is the partition function. Of course, we can compute the energy shifts for other states, but the thermal state is a reasonable equilibrium state that accounts to some extent for excited states, and reduces to the vacuum-state results that we have derived above in the limit $T \longrightarrow 0$.

14.3.8.1 Fluctuation–Dissipation Relation

To look at the fluctuations of a physical quantity, we can recall as motivation the optical Wiener–Khinchin theorem (Sections 2.2 and 5.7), and then write down the power spectral density as a Fourier transform of a correlation function. Specifically, we will write the *two-sided* power spectral density for the dipole fluctuations for the system in state $|n\rangle$ in terms of a *symmetrized* correlation tensor as

$$\tilde{S}^{(n)}_{\mu\nu}(\omega) := \frac{1}{4\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\langle n|[d_\mu(0), d_\nu(\tau)]_+|n\rangle, \tag{14.288}$$

and then write the *one-sided* tensor spectral density as

$$S^{(n)}_{\mu\nu}(\omega) := \tilde{S}^{(n)}_{\mu\nu}(\omega) + \tilde{S}^{(n)}_{\mu\nu}(-\omega) = \tilde{S}^{(n)}_{\mu\nu}(\omega) + \tilde{S}^{(n)}_{\nu\mu}(\omega). \tag{14.289}$$

We are as usual assuming steady state, so we suppress any explicit dependence on the absolute time t. This is somewhat different from the spectrum we have written down before, but it clearly represents *some* fluctuation at frequency ω, and the sum over all frequencies properly represents the total dipole fluctuation in state $|n\rangle$,

$$\begin{aligned}\int_0^\infty d\omega\, S_{\mu\nu}^{(n)}(\omega) &= \int_{-\infty}^\infty d\omega\, \tilde{S}_{\mu\nu}^{(n)}(\omega) \\ &= \frac{1}{4\pi}\int_{-\infty}^\infty d\tau \int_{-\infty}^\infty d\omega\, e^{i\omega\tau}\langle n|[d_\mu(0), d_\nu(\tau)]_+|n\rangle \\ &= \frac{1}{2}\int_{-\infty}^\infty d\tau\, \delta(\tau)\langle n|[d_\mu(0), d_\nu(\tau)]_+|n\rangle \\ &= \frac{1}{2}\langle n|[d_\mu, d_\nu]_+|n\rangle = \langle n|d_\mu d_\nu|n\rangle\end{aligned} \tag{14.290}$$

where in the second step we used the fact that the correlation function is an even function of τ, if we assume the power spectral density $S_{\mu\nu}^{(n)}(\omega)$ to be symmetric. In particular, the trace of this relation is mostly what we would associate with the total fluctuations,

$$\int_0^\infty d\omega\, S_{\mu\mu}^{(n)}(\omega) = \langle n|d^2|n\rangle \tag{14.291}$$

(with the usual implied summation). Technically speaking, the diagonal elements of the spectral tensor represent the fluctuations, while the off-diagonal elements represent correlations (covariances) between fluctuations of different dipole-operator components.

Now our goal will be to relate these fluctuations to the dissipation (absorption) in the system. We will start by rewriting the spectral (fluctuation) tensor as

$$\begin{aligned}2\tilde{S}_{\mu\nu}^{(n)}(\omega) &= \frac{1}{2\pi}\int_{-\infty}^\infty d\tau\, e^{i\omega\tau}\langle n|\Big[d_\mu(0)\, d_\nu(\tau) + d_\nu(\tau)\, d_\mu(0)\Big]|n\rangle \\ &= \frac{1}{2\pi}\sum_j \int_{-\infty}^\infty d\tau\, e^{i\omega\tau}\Big[\langle n|d_\mu(0)|j\rangle\langle j|d_\nu(\tau)|n\rangle + \langle n|d_\nu(\tau)|j\rangle\langle j|d_\mu(0)|n\rangle\Big] \\ &= \frac{1}{2\pi}\sum_j \int_{-\infty}^\infty d\tau\, e^{i\omega\tau}\Big[\langle n|d_\mu(0)|j\rangle\langle j|d_\nu(0)|n\rangle e^{i\omega_{jn}\tau} + \langle n|d_\nu(0)|j\rangle\langle j|d_\mu(0)|n\rangle e^{i\omega_{nj}\tau}\Big] \\ &= \sum_j \Big[\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\delta(\omega+\omega_{jn}) + \langle n|d_\nu|j\rangle\langle j|d_\mu|n\rangle\delta(\omega+\omega_{nj})\Big] \\ &= \sum_j \langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\Big[\delta(\omega+\omega_{jn}) + \delta(\omega+\omega_{nj})\Big].\end{aligned} \tag{14.292}$$

Here, the frequency interval ω_{nj} is as usual

$$\omega_{nj} = \frac{E_n - E_j}{\hbar} \tag{14.293}$$

in terms of the eigenstate energies.

Now we can compute the spectral tensor in the case of a thermal state by summing over all states $|n\rangle$,

with each term weighted by the occupation probability (14.286).

$$\begin{aligned} 2\tilde{S}_{\mu\nu}(\omega) &= \sum_{nj} P(n)\,\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\Big[\delta(\omega+\omega_{jn})+\delta(\omega+\omega_{nj})\Big] \\ &= \sum_{nj}\big[P(n)+P(j)\big]\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\,\delta(\omega+\omega_{nj}) \\ &= \sum_{nj} P(n)\left[1+e^{\hbar\omega_{nj}/k_\mathrm{B}T}\right]\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\,\delta(\omega+\omega_{nj}) \\ &= \sum_{nj} P(n)\left[1+e^{-\hbar\omega/k_\mathrm{B}T}\right]\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle]\,\delta(\omega+\omega_{nj}). \end{aligned} \tag{14.294}$$

In the second step here we interchanged (relabeled) the summation indices, and in the last step we used the projection property of the delta function.

Now that we have the spectral tensor in this form, we will turn our attention to the dissipation. Recall from Section 14.1.4.1 that the imaginary part of a response function is responsible for the loss or dissipation from energy. This result certainly applies to the polarizability, and we showed explicitly this to be the case in its classical treatment, as in Eq. (1.85). From Eq. (14.153), the imaginary part of the atomic polarizability for an atom in state $|n\rangle$ is

$$\mathrm{Im}[\alpha^{(n)}_{\mu\nu}(\omega)] = \frac{\pi}{\hbar}\sum_j \langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\Big[\delta(\omega+\omega_{nj})-\delta(\omega+\omega_{jn})\Big]. \tag{14.295}$$

Similarly averaging this expression over the thermal state (14.285), we find

$$\begin{aligned} \mathrm{Im}[\alpha_{\mu\nu}(\omega)] &= \frac{\pi}{\hbar}\sum_{nj} P(n)\,\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\Big[\delta(\omega+\omega_{nj})-\delta(\omega+\omega_{jn})\Big] \\ &= \frac{\pi}{\hbar}\sum_{nj}\big[P(n)-P(j)\big]\,\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\,\delta(\omega+\omega_{nj}) \\ &= \frac{\pi}{\hbar}\sum_{nj} P(n)\big[1-e^{-\hbar\omega/k_\mathrm{B}T}\big]\,\langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle\,\delta(\omega+\omega_{nj}). \end{aligned} \tag{14.296}$$

Comparing this result to Eq. (14.294), we can see the similarity and identify

$$\tilde{S}_{\mu\nu}(\omega) = \frac{\hbar}{2\pi}\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\,\frac{1+e^{-\hbar\omega/k_\mathrm{B}T}}{1-e^{-\hbar\omega/k_\mathrm{B}T}} = \frac{\hbar}{2\pi}\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\coth\left[\frac{\hbar\omega}{2k_\mathrm{B}T}\right]. \tag{14.297}$$

(fluctuation–dissipation relation)

This result is known as the **fluctuation–dissipation relation**,[25] relating the spectral density $S_{\mu\nu}(\omega)$ of fluctuations at frequency ω to the dissipation part of the response function $\mathrm{Im}[\alpha_{\mu\nu}(\omega)]$. Again, summing Eq. (14.288) over all levels,

$$\tilde{S}_{\mu\nu}(\omega) = \frac{1}{4\pi}\int_{-\infty}^{\infty} d\tau\, e^{i\omega\tau}\langle[d_\mu(0),d_\nu(\tau)]_+\rangle\,, \tag{14.298}$$

where the expectation value here is again taken with respect to the thermal state at temperature T. Note that

$$\coth\left[\frac{\hbar\omega}{2k_\mathrm{B}T}\right] = \frac{e^{\hbar\omega/k_\mathrm{B}T}+1}{e^{\hbar\omega/k_\mathrm{B}T}-1} = 2\left[\frac{1}{2}+\frac{1}{e^{\hbar\omega/k_\mathrm{B}T}-1}\right], \tag{14.299}$$

where the last quantity in brackets represents the mean thermal energy of a quantum harmonic oscillator at frequency ω, in units of $\hbar\omega$—the first term is the zero-point energy, while the second term represents the thermal contribution.

[25] Herbert B. Callen and Theodore A. Welton, "Irreversibility and Generalized Noise," *Physical Review* **83**, 34 (1951) (doi: 10.1103/PhysRev.83.34); L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Pergamon, 1980), §124.

We can invert the Fourier transform in the fluctuation–dissipation relation (14.297) and use Eq. (14.298) to write the fluctuations directly in terms of the dipole autocorrelation function:

$$\frac{1}{2}\langle[d_\mu(0), d_\nu(\tau)]_+\rangle = \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\coth\left[\frac{\hbar\omega}{2k_\mathrm{B}T}\right].$$

(fluctuation–dissipation relation) (14.300)

Then the covariance matrix for the dipole fluctuations is given by taking $\tau = 0$ in this expression, with the result

$$\langle d_\mu d_\nu\rangle = \frac{\hbar}{\pi}\int_0^{\infty} d\omega\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\coth\left[\frac{\hbar\omega}{2k_\mathrm{B}T}\right].$$

(fluctuation–dissipation relation) (14.301)

As written here, the covariance matrix is obviously symmetric, so we have dropped the anticommutator. Thus, *any absorptive character of the atomic dipole necessarily leads to dipole fluctuations.* Of course, any *dispersion* implies some absorption by the Kramers–Kronig relations, so dispersion also implies fluctuations. Note that $\coth x \longrightarrow 1$ as $x \longrightarrow \infty$, so fluctuations persist even as $T \longrightarrow 0$. These zero-temperature fluctuations are obviously quantum-mechanical in nature. However, for high temperatures, we can use $\coth x \longrightarrow 1/x$ for small x to write

$$\langle d_\mu d_\nu\rangle = \frac{2k_\mathrm{B}T}{\pi}\int_0^{\infty}\frac{d\omega}{\omega}\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)] = k_\mathrm{B}T\,\mathrm{Re}[\alpha_{\mu\nu}(0)], \qquad \begin{matrix}(14.302)\\(\text{large } T)\end{matrix}$$

which no longer involves $\hbar$. (We used the Kramers–Kronig relations (14.90) to evaluate the integral here.) In this case, the quantum fluctuations are negligible compared to the (classical) thermal fluctuations.

Of course, all of these results apply as well to *any* observable and its linear response function, assuming a linear interaction Hamiltonian. In particular, for the electric field,

$$\frac{1}{2}\langle[E_\mu(\mathbf{r}, 0), E_\nu(\mathbf{r}', \tau)]_+\rangle = \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\,\mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)]\coth\left[\frac{\hbar\omega}{2k_\mathrm{B}T}\right].$$

(fluctuation–dissipation relation) (14.303)

The imaginary part of the Green tensor represents dissipation (material absorption) of the electromagnetic field, which again implies field fluctuations.

14.3.8.2 Fluctuation–Dissipation Example: Johnson Noise

As a simple example and application of the fluctuation–dissipation relation, we can consider **Johnson noise**,[26] the intrinsic noise in a any resistor, independent of the details of its material composition, geometry, and so on. The resistance obviously represents the dissipation, and we will show that it implies fluctuations in the form of voltage noise. We start with an interaction Hamiltonian in the form of (14.102), which will represent the energy of electrons in a resistor:

$$H_\mathrm{int} = V\sum_j \frac{q_j x_j}{L}. \tag{14.304}$$

Here, V is the "voltage operator," L is the length of the conduction path of the resistor, and q_j and x_j are respectively the charge and position of particle j. In linear-response theory, the voltage operator will respond to the classical "force"

$$F(t) = -\sum_j \frac{q_j x_j}{L}. \tag{14.305}$$

[26] Johnson noise is named after the first person to measure it: J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Physical Review Letters* **32**, 97 (1928) (doi: 10.1103/PhysRev.32.97). It is also called Johnson–Nyquist noise, named additionally after the first to describe it theoretically: H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Physical Review Letters* **32**, 110 (1928) (doi: 10.1103/PhysRev.32.110).

The current represents the flow of charge as a rate of charge per unit time, and thus

$$I(t) = -\dot{F}(t) = \sum_j \frac{q_j \dot{x}_j}{L}. \tag{14.306}$$

That is, $\dot{H}_{\text{int}} = VI(t)$ is the power dissipated due to motion of the charges in the resistor, assuming a constant voltage. Expressed as a Fourier transform,

$$I(\omega) = i\omega F(\omega). \tag{14.307}$$

The Fourier transform of Ohm's law reads

$$V(\omega) = Z(\omega) I(\omega), \tag{14.308}$$

where $Z(\omega)$ is the frequency-dependent **impedance** of the resistor. In terms of the ''force'' function,

$$V(\omega) = i\omega Z(\omega) F(\omega), \tag{14.309}$$

and thus $i\omega Z(\omega)$ is the generalized susceptibility for the resistor. Now using the fluctuation–dissipation relation in the high-temperature limit (14.302), we can write the variance of the voltage fluctuations as

$$\left\langle V^2 \right\rangle = \frac{2k_{\text{B}}T}{\pi} \int_0^\infty d\omega \, \frac{\text{Im}[i\omega Z(\omega)]}{\omega} = \frac{2k_{\text{B}}T}{\pi} \int_0^\infty d\omega \, \text{Re}[Z(\omega)]. \tag{14.310}$$

The real part of the impedance is the **resistance**, and thus[27]

$$\left\langle V^2 \right\rangle = \frac{2k_{\text{B}}T}{\pi} \int_0^\infty d\omega \, R(\omega). \tag{14.311}$$
(Johnson noise)

These voltage fluctuations are what are referred to as Johnson noise. The voltage noise is typically measured only over some bandwidth $\Delta\nu$. Changing to a ''regular'' frequency from the angular frequency,

$$\left\langle V^2 \right\rangle = 4k_{\text{B}}T \int_0^{\Delta\nu} d\nu \, R(\nu), \tag{14.312}$$

and if the resistance is roughly constant over the measurement bandwidth, we arrive at the well-known expression

$$\left\langle V^2 \right\rangle = 4Rk_{\text{B}}T \, \Delta\nu. \tag{14.313}$$
(Johnson noise, $\Delta\nu$ bandwidth limit)

The noise is proportional to temperature (in the classical limit of high temperature), and has the character of white noise, so long as $R(\omega)$ is constant over the range of the bandwidth limit. For example, at $T = 293$ K, a 10 kΩ resistor measured over a 10 kHz bandwidth has an intrinsic, thermal rms voltage noise of 1.2 μV. At the same temperature, a 1 MΩ resistor measured over a 1 MHz bandwidth has an rms voltage noise of 0.12 mV, which is starting to become significant on the scale of laboratory voltages.

If we use the fluctuation–dissipation relation in the more general form (14.301), we similarly find the general result

$$\left\langle V^2 \right\rangle = \frac{\hbar}{\pi} \int_0^\infty d\omega \, \omega \, R(\omega) \coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right]. \tag{14.314}$$
(Johnson noise, arbitrary T)

If we take the limit of small temperature, we can replace the coth by 1:

$$\left\langle V^2 \right\rangle = \frac{\hbar}{\pi} \int_0^\infty d\omega \, \omega \, R(\omega). \tag{14.315}$$
(Johnson noise, small T)

[27] Herbert B. Callen and Theodore A. Welton, *op. cit.*, Eq. (4.11).

The same bandwidth limit leads in this case to zero-temperature noise of

$$\langle V^2 \rangle = 2\pi\hbar R\,\Delta\nu^2, \qquad \text{(Johnson noise, small } T\text{, } \Delta\nu \text{ bandwidth limit)} \quad (14.316)$$

assuming a constant resistance over the bandwidth. Clearly, quantum fluctuations persist even at zero temperature, producing "quantum Johnson noise."[28]

14.3.8.3 Temperature-Dependent Shifts

Now on to the level shift.[29] In the zero-temperature case, we used the second-order perturbation expression

$$\delta E_n = \sum_j \frac{|\langle n|H_{\rm int}|j\rangle|^2}{E_n - E_j} \quad (14.317)$$

for the shift of level $|n\rangle$ due to the interaction energy $H_{\rm int}$, where the indices label combined states of the atom and field. Now, for a thermal state, we must perform the average

$$\delta E = \sum_{nj} P(n)\,\frac{|\langle n|H_{\rm int}|j\rangle|^2}{E_n - E_j} \quad (14.318)$$

over the thermal occupation probabilities (14.286) to treat the shift at nonzero temperature. We will now claim that this second-order shift may be written in terms of the correlation function

$$\delta E = \frac{i}{2\hbar}\int_{-\infty}^{0} d\tau\,\langle [H_{\rm int}(\tau), H_{\rm int}(0)]\rangle\,. \qquad \text{(perturbative energy shift)} \quad (14.319)$$

This result is, of course, valid for *any* stationary state, not just the thermal ones, and gives a nice, representation-independent expression for the shift. Strictly speaking, this result assumes that the perturbation is turned on adiabatically in the distant past, so that we may insert a convergence factor to guarantee a sensible result:

$$\delta E = \lim_{\sigma\to 0^+}\frac{i}{2\hbar}\int_{-\infty}^{0} d\tau\,\langle [H_{\rm int}(\tau), H_{\rm int}(0)]\rangle\; e^{\sigma\tau}. \quad (14.320)$$

To see this, we perform algebraic steps that are similar to what we used for the fluctuation–dissipation relation:

$$\begin{aligned}
\delta E &= \frac{i}{2\hbar}\sum_n P(n)\int_{-\infty}^{0} d\tau\,\langle n|[H_{\rm int}(\tau), H_{\rm int}(0)]|n\rangle \\
&= \frac{i}{2\hbar}\sum_{nj} P(n)\int_{-\infty}^{0} d\tau\,\langle n|H_{\rm int}(\tau)|j\rangle\langle j|H_{\rm int}(0)|n\rangle - \langle n|H_{\rm int}(0)|j\rangle\langle j|H_{\rm int}(\tau)\Big]|n\rangle \\
&= \frac{i}{2\hbar}\sum_{nj} P(n)\int_{-\infty}^{0} d\tau\,|\langle n|H_{\rm int}|j\rangle|^2\left(e^{i\omega_{nj}\tau} - e^{i\omega_{jn}\tau}\right) \\
&= \frac{i}{2\hbar}\sum_{nj}\big[P(n) - P(j)\big]\int_{-\infty}^{0} d\tau\,|\langle n|H_{\rm int}|j\rangle|^2\,e^{i\omega_{nj}\tau}.
\end{aligned} \quad (14.321)$$

[28] For a measurement of the coth dependence of Johnson noise at low temperature, see R. Movshovich, B. Yurke, P. G. Kaminsky, A. D. Smith, A. H. Silver, R. W. Simon, and M. V. Schneider, "Observation of zero-point noise squeezing via a Josephson-parametric amplifier," *Physical Review Letters* **65**, 1419 (1990) (doi: 10.1103/PhysRevLett.65.1419).

[29] The derivation here closely follows A. D. McLachlan, "Retarded Dispersion Forces in Dielectrics at Finite Temperatures," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **274**, 80 (1963).

As we mentioned, we should really have a convergence factor here, so

$$\begin{aligned}\delta E &= \frac{i}{2\hbar}\sum_{nj}\big[P(n)-P(j)\big]\lim_{\sigma\to 0^+}\int_{-\infty}^{0} d\tau\,|\langle n|H_{\rm int}|j\rangle|^2\,e^{i\omega_{nj}\tau}\,e^{\sigma\tau}\\ &= \frac{1}{2\hbar}\sum_{nj}\big[P(n)-P(j)\big]\frac{|\langle n|H_{\rm int}|j\rangle|^2}{\omega_{nj}}\\ &= \sum_{nj}P(n)\frac{|\langle n|H_{\rm int}|j\rangle|^2}{\hbar\omega_{nj}}.\end{aligned} \tag{14.322}$$

This last expression is equivalent to the second-order expression Eq. (14.318).

Now to evaluate the commutator in Eq. (14.319) for the dipole interaction Hamiltonian $H_{\rm int} = -\mathbf{d}\cdot\mathbf{E} = -d_\mu E_\mu$. We can then write

$$\begin{aligned}[H_{\rm int}(\tau), H_{\rm int}(0)] &= [d_\mu(\tau)E_\mu(\tau), d_\nu(0)E_\nu(0)]\\ &= d_\mu E_\mu d_\nu E_\nu - d_\nu E_\nu d_\mu E_\mu\\ &= d_\mu d_\nu E_\mu E_\nu - d_\nu d_\mu E_\nu E_\mu\\ &= \frac{1}{2}\Big[\big(d_\mu d_\nu E_\mu E_\nu + d_\mu d_\nu E_\nu E_\mu - d_\nu d_\mu E_\mu E_\nu - d_\nu d_\mu E_\nu E_\mu\big)\\ &\qquad + \big(d_\mu d_\nu E_\mu E_\nu - d_\mu d_\nu E_\nu E_\mu + d_\nu d_\mu E_\mu E_\nu - d_\nu d_\mu E_\nu E_\mu\big)\Big]\\ &= \frac{1}{2}\Big([d_\mu(\tau), d_\nu(0)]\,[E_\mu(\tau), E_\nu(0)]_+ + [d_\mu(\tau), d_\nu(0)]_+\,[E_\mu(\tau), E_\nu(0)]\Big),\end{aligned} \tag{14.323}$$

where we used the shorthands $d_\mu \equiv d_\mu(\tau)$, $d_\nu \equiv d_\nu(0)$, $E_\mu \equiv E_\mu(\tau)$, and $E_\nu \equiv E_\nu(0)$, and we have used the fact that under unperturbed evolution, $\mathbf{d}(\tau)$ and $\mathbf{E}(\tau')$ commute even at different times. We can then take the expectation value of the commutator $[H_{\rm int}(\tau), H_{\rm int}(0)]$ and then use the commutator-correlation-function expressions in the forms of Eqs. (14.125) and (14.130), as well as the fluctuation–dissipation relations in the forms (14.300) and (14.303), to write

$$\begin{aligned}\langle[H_{\rm int}(\tau), H_{\rm int}(0)]\rangle &= \frac{1}{2}\Big(\langle[d_\mu(\tau), d_\nu(0)]\rangle\,\langle[E_\mu(\tau), E_\nu(0)]_+\rangle + \langle[d_\mu(\tau), d_\nu(0)]_+\rangle\,\langle[E_\mu(\tau), E_\nu(0)]\rangle\Big)\\ &= \frac{\hbar^2}{2\pi^2}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\,\mathrm{Im}[G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega')]\left\{\coth\left[\frac{\hbar\omega}{2k_{\rm B}T}\right] + \coth\left[\frac{\hbar\omega'}{2k_{\rm B}T}\right]\right\}e^{-i(\omega+\omega')t}.\end{aligned} \tag{14.324}$$

The energy shift (14.319) represents the finite-temperature version of the Casimir–Polder potential, and thus we can now write

$$\begin{aligned}V_{\rm CP} &= \frac{i\hbar}{4\pi^2}\int_{-\infty}^{0}d\tau\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\,\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\,\mathrm{Im}[G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega')]\left\{\coth\left[\frac{\hbar\omega}{2k_{\rm B}T}\right] + \coth\left[\frac{\hbar\omega'}{2k_{\rm B}T}\right]\right\}e^{-i(\omega+\omega')\tau}\\ &= -\frac{\hbar}{4\pi^2}\int_{-\infty}^{\infty}d\omega\int_{-\infty}^{\infty}d\omega'\,\frac{\mathrm{Im}[\alpha_{\mu\nu}(\omega)]\,\mathrm{Im}[G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega')]}{\omega+\omega'}\left\{\coth\left[\frac{\hbar\omega}{2k_{\rm B}T}\right] + \coth\left[\frac{\hbar\omega'}{2k_{\rm B}T}\right]\right\},\end{aligned} \tag{14.325}$$

where we implicitly used the usual convergence factor in the time integral. We can then use the Kramers–Kronig relations (14.90) adapted to the dipole and field response functions,

$$\begin{aligned}\mathrm{Re}[\alpha_{\mu\nu}(\omega)] &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{Im}[\alpha_{\mu\nu}(\omega')]}{\omega'-\omega}\,d\omega'\\ \mathrm{Re}[G_{\mu\nu}(\mathbf{r},\mathbf{r}',\omega)] &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\mathrm{Im}[G_{\mu\nu}(\mathbf{r},\mathbf{r}',\omega')]}{\omega'-\omega}\,d\omega',\end{aligned} \tag{14.326}$$

to carry out the ω' integral (changing variables beforehand in the second term), with the result

$$\begin{aligned} V_{\text{CP}} &= -\frac{\hbar}{4\pi}\int_{-\infty}^{\infty} d\omega \left[\text{Re}[\alpha_{\mu\nu}(-\omega)]\,\text{Im}[G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega)] + \text{Im}[\alpha_{\mu\nu}(\omega)]\,\text{Re}[G_{\nu\mu}(\mathbf{r},\mathbf{r},-\omega)]\right] \coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right] \\ &= -\frac{\hbar}{4\pi}\int_{-\infty}^{\infty} d\omega\, \text{Im}[\alpha_{\mu\nu}(\omega)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega)] \coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right], \end{aligned} \tag{14.327}$$

where we have used the fact that the real parts of the response functions are even functions of the real frequency ω. We may rewrite this last expression as

$$V_{\text{CP}} = -\frac{\hbar}{4\pi i}\int_{-\infty}^{\infty} d\omega\, \alpha_{\mu\nu}(\omega)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega) \coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right], \tag{14.328}$$

if we recall that $\coth x$ has a simple pole at $x = 0$ and that $\text{Re}[\alpha_{\mu\nu}(\omega)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},\omega)]$ is an even function of ω, so that it leads to a vanishing contribution in the principal-value integral.

Now we will reduce this integral to a summation as follows. Since $\coth ix = -i\cot x$, $\coth x$ has simple poles at $x = i\pi n$ for every integer n. Thus, $\coth(\hbar\omega/2k_{\text{B}}T)$ has poles at $\omega = is_n$, where the imaginary frequencies are given by

$$s_n = n\frac{2\pi k_{\text{B}}T}{\hbar}. \tag{14.329}$$

(Matsubara frequencies)

These discrete frequencies are called the **Matsubara frequencies**.[30] Furthermore, the residues of the thermal function $\coth(\hbar\omega/2k_{\text{B}}T)$ at each $\omega = is_n$ is simply $2k_{\text{B}}T/\hbar$. We can then change the integral in Eq. (14.328) to a contour integral over the great semicircle in the upper half-plane, as we did to derive the Kramers–Kronig relations in Eq. (14.1.4.2). The semicircular part of the contour vanishes because $\alpha_{\mu\nu}(\omega)$ decays at least as fast as $1/|\omega|^2$ for large $|\omega|$, as we saw in our derivation of the Kramers–Kronig relations, Eq. (14.83). Thus, by Cauchy's integral formula, the integral in Eq. (14.328) changes to $2\pi i$ times the sum over residues at frequencies is_n, with the result

$$V_{\text{CP}} = -\frac{k_{\text{B}}T}{2}\alpha_{\mu\nu}(is_0)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},is_0) - k_{\text{B}}T\sum_{n=1}^{\infty}\alpha_{\mu\nu}(is_n)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},is_n) \tag{14.330}$$

Notice that the pole at $\omega = 0$ only contributes half its residue because of the principal value that we take in Eq. (14.328) knocks out half the contribution of any real-axis pole. Using the original notation of Lifshitz,[31] we may write this sum as

$$V_{\text{CP}} = -k_{\text{B}}T{\sum_{n=0}^{\infty}}'\alpha_{\mu\nu}(is_n)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},is_n).$$

(temperature-dependent Casimir–Polder shift) (14.331)

where the primed summation symbol $\sum'$ denotes that the $n = 0$ term is accompanied by an extra factor of $1/2$. Again, to avoid a divergence and to focus only on the interaction of the atom with a macrocopic body, we should remove the free-field contribution and use only the scattering part of the Green tensor:

$$V_{\text{CP}} = -k_{\text{B}}T{\sum_{n=0}^{\infty}}'\alpha_{\mu\nu}(is_n)\, G^{(\text{s})}_{\nu\mu}(\mathbf{r},\mathbf{r},is_n). \tag{14.332}$$

(renormalized form)

Thus, compared to the zero-temperature expression (14.163), which involved an integral over imaginary frequency of the product of the dipole and field susceptibilities, the finite-temperature case involves a discrete sum over the Matsubara frequencies.

[30] After Takeo Matsubara, "A New Approach to Quantum Statistical Mechanics," *Progress in Theoretical Physics* **14**, 351 (1955) (doi: 10.1143/PTP.14.351).

[31] E. M. Lifshitz, "The Theory of Molecular Attractive Forces between Solids," *Soviet Physics JETP* **2**, 73 (1956).

14.3.8.4 Imaginary Time and the Low-Temperature Limit

Recalling from Eq. (14.129) that the Green tensor is given as a correlation function as

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = \frac{i}{\hbar}\int_0^\infty d\tau\,\langle[E_\alpha(\mathbf{r},\tau),E_\beta(\mathbf{r}',0)]\rangle\; e^{i\omega\tau}, \tag{14.333}$$

we can see that for imaginary frequencies $\omega = is$, the Green tensor amounts to a Laplace transform:

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is) = \frac{i}{\hbar}\int_0^\infty d\tau\,\langle[E_\alpha(\mathbf{r},\tau),E_\beta(\mathbf{r}',0)]\rangle\; e^{-s\tau}. \tag{14.334}$$

Shifting to an imaginary time $\tau \longrightarrow -i\tau$, we return to the Fourier-type integral expression

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is) = \frac{1}{\hbar}\int_0^\infty d\tau\,\langle[E_\alpha(\mathbf{r},-i\tau),E_\beta(\mathbf{r}',0)]\rangle\; e^{is\tau}. \tag{14.335}$$

However, the point of the discussion above is that the Green tensor effectively vanishes everywhere except at the Matsubara frequencies is_n,

$$V_{\text{CP}} = -\frac{\hbar}{2\pi}\int_{0^-}^\infty ds\,\alpha_{\mu\nu}(is)\,\tilde{G}_{\nu\mu}(\mathbf{r},\mathbf{r},is), \tag{14.336}$$

where

$$\tilde{G}_{\alpha\beta}(\mathbf{r},\mathbf{r},is) := \frac{2\pi k_{\text{B}}T}{\hbar}{\sum_{n=0}^\infty}' G_{\alpha\beta}(\mathbf{r},\mathbf{r},is_n)\,\delta(s-s_n). \tag{14.337}$$

Thus, the finite-temperature potential takes on the same form as the zero-temperature potential (14.163), under the replacement $G_{\alpha\beta}(\mathbf{r},\mathbf{r},is) \longrightarrow \tilde{G}_{\alpha\beta}(\mathbf{r},\mathbf{r},is)$. Since the spectrum $G_{\alpha\beta}(\mathbf{r},\mathbf{r},is)$ is then effectively discrete and periodic, we may refer to our discussion of sampling and the sampling theorem (Section 25.1) to note that we may think of its Fourier transform as a periodic function in the imaginary time, with the ''samples'' given by

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is_n) = \frac{1}{\hbar}\int_0^{\hbar/k_{\text{B}}T} d\tau\,\langle[E_\alpha(\mathbf{r},-i\tau),E_\beta(\mathbf{r}',0)]\rangle\; e^{is_n\tau}. \tag{14.338}$$

That is, the values $(k_{\text{B}}T/\hbar)G_{\alpha\beta}(\mathbf{r},\mathbf{r}',is_n)$ are the Fourier components of the **imaginary-time Green tensor**

$$\mathscr{G}_{\alpha\beta}(\mathbf{r},\mathbf{r}',\tau) := \frac{1}{\hbar}\langle[E_\alpha(\mathbf{r},-i\tau),E_\beta(\mathbf{r}',0)]\rangle\,\Theta(\tau), \tag{14.339}$$

which we may regard as time-periodic with period $\hbar/k_{\text{B}}T$. In the low-temperature limit as $T \longrightarrow 0$, the period diverges, and under an integral sign we see from the definition (14.337) that $\tilde{G}_{\alpha\beta}(\mathbf{r},\mathbf{r},is) \longrightarrow G_{\alpha\beta}(\mathbf{r},\mathbf{r},is)$, so that we obtain the previous expression for the zero-temperature Casimir–Polder shift.

14.3.8.5 High-Temperature Limit

The high-temperature limit of the potential (14.332) comes by noting that the spacing between the Matsubara frequencies becomes very large for large T. Since the summand of Eq. (14.332) decreases monotonically with frequency, for sufficiently high temperature only the dc term will make a substantial contribution. Thus, to leading order,

$$V_{\text{CP}} = -\frac{1}{2}k_{\text{B}}T\,\alpha_{\mu\nu}(0)\,G^{(\text{s})}_{\nu\mu}(\mathbf{r},\mathbf{r},0). \tag{14.340}$$
(large T)

This expression is evidently the classical Stark shift of the atom due to the presence of thermal photons, where the quantum zero-point contribution is negligible, since $\hbar$ is absent in this expression. (In the general time-dependent expression, $\hbar$ appears in the definition of the frequencies s_n.) In the case of a nondegenerate

ground state, the imaginary parts of the susceptibilities vanish at $\omega = 0$, since there is no mechanism for dissipation. Then[32]

$$V_{\rm CP} = -\frac{1}{2}k_{\rm B}T\,{\rm Re}[\alpha_{\mu\nu}(0)]\,{\rm Re}[G^{(\rm s)}_{\nu\mu}(\mathbf{r},\mathbf{r},0)] = -\frac{\langle d_\mu d_\nu\rangle\langle E_\mu E_\nu\rangle}{2k_{\rm B}T} = -\frac{\langle H^2_{\rm AF}\rangle}{2k_{\rm B}T}, \tag{14.341}$$

where we used the high-temperature fluctuation–dissipation relation (14.302) and its analog for the field. Again the expectation values here are relative to their values in free space, since we have discarded the divergent free-space contribution.

14.3.8.6 Planar Boundaries at Nonzero Temperature

As an example of temperature-dependent effects, we will consider an atom near a planar surface. For simplicity, we will consider the high-temperature limit in the far field, and also a spherically symmetric atom. Comparing the general expression (14.167) at zero temperature to the high-temperature expression (14.340) we see that we can obtain the high-temperature result from the zero-temperature result by omitting the integral over s, setting $s = 0$ in the remaining integrand, and multiplying by $\pi k_{\rm B}T/\hbar$. Doing this in Eq. (14.205), we obtain

$$V_{\rm CP} = \frac{k_{\rm B}T\alpha_0}{4\pi\epsilon_0}\int_0^\infty d\kappa\,\kappa^2 r_\parallel(\theta,0)\,e^{-2\kappa z}, \tag{14.342}$$

where with $s = 0$, we can use $k_\tau = \kappa$. Furthermore, for $s = 0$, the reflection coefficient is replaced as in the far-field limit by $[\epsilon_0 - \epsilon(0)]/[\epsilon_0 + \epsilon(0)]$. Then evaluating the remaining integral, we find

$$V_{\rm CP} = -\frac{k_{\rm B}T\alpha_0}{16\pi\epsilon_0 z^3}\left(\frac{\epsilon(0)-\epsilon_0}{\epsilon(0)+\epsilon_0}\right).$$

(planar dielectic, high-temperature/large-distance limit) (14.343)

For a conductor, $\epsilon(0) \longrightarrow \infty$, and thus

$$V_{\rm CP} = -\frac{k_{\rm B}T\alpha_0}{16\pi\epsilon_0 z^3}.$$

(conductor, high-temperature/large-distance limit) (14.344)

When do these high-temperature expressions hold? Essentially, the second term in the Matsubara sum (14.331), at frequency $s_1 = 2\pi k_{\rm B}T/\hbar$, must be negligible compared to the dc term. Since the Green tensor is damped as $e^{-2\kappa z} \le e^{-2sz/c}$, this will occur for

$$z \gg \frac{c}{2s_1} = \frac{\hbar c}{4\pi k_{\rm B}T}. \tag{14.345}$$

(high-temperature condition)

At room temperature, this condition amounts to $z \gg 0.6$ μm. The low-temperature limit corresponds to the opposite regime $z \ll \hbar c/4\pi k_{\rm B}T$, where the terms in the Matsubara sum are closely spaced in frequency and can be well approximated by the zero-temperature integral. Thus, even for normal temperatures corresponding to $k_{\rm B}T \ll \hbar\omega_0$ (i.e., $T \ll 10$ kK for optical transitions), where the atom is essentially in the ground state, it is still possible to be in a regime of ''high temperature'' if the distance is sufficiently large.[33] Thus, for a ground-state atom near the planar surface of a bulk dielectric, the interaction potential scales as z^{-3} in the near-field regime, then as z^{-4} in the retarded regime of $(2k_{0j})^{-1} \ll z \ll \hbar c/4\pi k_{\rm B}T$ (for all integer j). Then, in the very long-distance regime of $z \gg \hbar c/4\pi k_{\rm B}T$, the potential scales again as z^{-3}.

Comparing the general expression (14.167) at zero temperature to the temperature-dependent expression (14.332) we can write the temperature-dependent result by adapting the zero-temperature result

[32] A. D. McLachlan, *op. cit.*

[33] For the experimental observation of the large-distance temperature-dependent corrections, but in a nonequilibrium thermal state, see J. M. Obrecht, R. J. Wild, M. Antezza, L. P. Pitaevskii, S. Stringari, and E. A. Cornell, ''Measurement of the Temperature Dependence of the Casimir-Polder Force,'' *Physical Review Letters* **98**, 063201 (2007) (doi: 10.1103/PhysRevLett.98.063201).

as follows: replace the integral over s by an appropriate sum over Matsubara frequencies, and multiplying by $2\pi k_B T/\hbar$. Doing this in Eq. (14.186), we find a general expression for the temperature-dependent Casimir–Polder potential near a planar surface in terms of the reflection coefficients:

$$V_{CP} = \frac{k_B T}{4\pi\epsilon_0 c^2} {\sum_{n=0}^{\infty}}' s_n^2\, \alpha(is_n) \int_0^\infty dk_\tau\, \frac{k_\tau}{\kappa_n} \left[r_\perp(\theta, is_n) + \left(1 + \frac{2k_\tau^2 c^2}{s_n^2}\right) r_\parallel(\theta, is_n) \right] e^{-2\kappa_n z}.$$

(Lifshitz formula: temperature-dependent potential, planar surface) (14.346)

Here $\kappa_n = \sqrt{s_n^2/c^2 + k_\tau^2}$, and the angle θ still depends on s_n and k_τ. This is the **Lifshitz expression** for the atom–surface potential, after the original treatment of Lifshitz for the temperature-dependent potential between two surfaces,[34] from which this expression for the atom–surface force may be deduced.[35]

Again, in the limit of small temperature, the sum goes over to an integral, and we recover the zero-temperature expression (14.186). Then at small temperatures, by how much does the sum differ from the integral? To obtain a perturbative correction due to nonzero temperature, we can use the Euler–Maclaurin summation formula in the form[36]

$$\begin{aligned} {\sum_{j=0}^{\infty}}' f(j\Delta t) &= \frac{1}{\Delta t}\int_0^\infty dt\, f(t) - \sum_{j=1}^{\infty} \frac{(\Delta t)^{2j-1}}{(2j)!} B_{2j} f^{(2j-1)}(0) \\ &= \frac{1}{\Delta t}\int_0^\infty dt\, f(t) - \frac{\Delta t}{12} f'(0) + \frac{(\Delta t)^3}{720} f'''(0) + \cdots \end{aligned} \tag{14.347}$$

to look at precisely this difference, where B_n are the Bernoulli numbers. Then in the general case, we can approximately evaluate the Matsubara sum in Eq. (14.332) for small T by keeping only the correction terms shown explicitly in Eq. (14.347):

$$\begin{aligned} V_{CP} \approx &-\frac{\hbar}{2\pi}\int_0^\infty ds\, \alpha_{\mu\nu}(is)\, G^{(s)}_{\nu\mu}(\mathbf{r}, \mathbf{r}, is) \\ &+ \frac{2\pi(k_B T)^2}{12\hbar} \Big[\partial_s \Big(\alpha_{\mu\nu}(is)\, G^{(s)}_{\nu\mu}(\mathbf{r}, \mathbf{r}, is)\Big)\Big]_{s=0} - \frac{(2\pi)^3 (k_B T)^4}{720\hbar^3} \Big[\partial_s^3 \Big(\alpha_{\mu\nu}(is)\, G^{(s)}_{\nu\mu}(\mathbf{r}, \mathbf{r}, is)\Big)\Big]_{s=0} + O(T^6). \end{aligned}$$

(small T expansion) (14.348)

The first term is the usual zero-temperature expression, while the rest are effectively a power series in the temperature. Let's evaluate these perturbative corrections for a perfect conductor, in which case Eq. (14.346) becomes

$$V_{CP} = \frac{k_B T}{2\pi\epsilon_0 c^2} {\sum_{n=0}^{\infty}}' s_n^2\, \alpha(is_n) \int_0^\infty dk_\tau\, \frac{k_\tau}{\kappa_n} \left(1 + \frac{k_\tau^2 c^2}{s_n^2}\right) e^{-2\kappa_n z}. \tag{14.349}$$

To expand this expression as in Eq. (14.348), we first consider the zero-temperature formula (14.192), which we may write in the form

$$V_{CP} = -\frac{\hbar}{16\pi^2\epsilon_0 z^3} \int_0^\infty ds\, \alpha(is) \left(1 + \frac{2sz}{c} + \frac{2s^2 z^2}{c^2}\right) e^{-2sz/c}. \tag{14.350}$$

To write the temperature expansion in Eq. (14.348), the zeroth-order term is simply given by this expression. The order T^2 correction is given by this same expression, if we remove the integral sign, hit the integrand with ∂_s, set $s \longrightarrow 0$, and multiply by $-(2\pi k_B T/\hbar)^2/12$. The order T^4 correction is given by this same expression, if we remove the integral sign, hit the integrand with ∂_s^3, set $s \longrightarrow 0$, and multiply by $(2\pi k_B T/\hbar)^4/720$. Noting that $\partial_s \alpha(is) = 0$ at $s = 0$, we simply need to expand the remaining integrand

$$\left(1 + \frac{2sz}{c} + \frac{2s^2 z^2}{c^2}\right) e^{-2sz/c} = 1 - \frac{4z^3 s^3}{3c^3} + O(s^4), \tag{14.351}$$

[34] E. M. Lifshitz, *op. cit.*

[35] See J. F. Babb, G. L. Klimchitskaya, and V. M. Mostepanenko, "Casimir–Polder interaction between an atom and a cavity wall under the influence of real conditions," *Physical Review A* **70**, 042901 (2004) (doi: 10.1103/PhysRevA.70.042901).

[36] E. M. Lifshitz, *op. cit.*; Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 806, Eq. (23.1.30).

to see that the order T^2 correction vanishes. Then with the order T^4 correction the potential becomes

$$V_{\rm CP}(z,T) = V_{\rm CP}(z,0) + \frac{\pi^2 (k_{\rm B}T)^4 \alpha_0}{90\epsilon_0 \hbar^3 c^3} + O(T^6), \tag{14.352}$$

where $V_{\rm CP}(z,0)$ is given by Eq. (14.194). The lowest-order nonvanishing temperature correction thus just amounts to a z-independent offset. If we compare the correction to the far-field expression (14.208), the temperature-dependent corrections should be negligible so long as

$$z \ll \sqrt[4]{\frac{135}{16}} \left(\frac{\hbar c}{\pi k_{\rm B} T} \right). \tag{14.353}$$

(zero-temperature validity condition)

At room temperature, this condition amounts to $z \ll 4$ μm.

14.3.9 Excited-Level Shifts

Thus far, we have mainly treated level shifts only of the ground state, but what happens to excited levels due to the presence of some extra body? The shift of level $|n\rangle$ is given in second-order perturbation theory by the same expression (14.133) as for the ground state,

$$V_n = -\sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle j|\mathbf{d}|n\rangle \cdot \langle 0|\mathbf{E}|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar(\omega_{jn} + \omega_{\mathbf{k}})}, \tag{14.354}$$

but now the difference is that ω_{jn} may be negative for states $|j\rangle$ lower in energy than $|n\rangle$, whereas for the ground state this frequency was always positive. To see how the level shift changes due to this sign change, we can try out the ground-state expression (14.163)

$$V_n^{(1)} = -\frac{\hbar}{2\pi} \int_0^\infty ds\, \alpha_{\mu\nu}^{(n)}(is)\, G_{\nu\mu}(\mathbf{r},\mathbf{r},is),$$

("ground-state part" of level shift) (14.355)

where now $\alpha_{\mu\nu}^{(n)}$ is the polarizability tensor for $|n\rangle$, given by rewriting the Kramers–Heisenberg formula (14.147) as

$$\alpha_{\mu\nu}^{(n)}(\omega) = \sum_j \frac{2\omega_{jn} \langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle}{\hbar(\omega_{jn}^2 - \omega^2)}. \tag{14.356}$$

Substituting this expression and also Eq. (14.159) for the Green tensor into this relation as before, we find

$$\begin{aligned} V_n^{(1)} &= -\frac{\hbar}{2\pi} \int_0^\infty ds \sum_j \frac{2\omega_{jn} \langle n|d_\mu|j\rangle\langle j|d_\nu|n\rangle}{\hbar(\omega_{jn}^2 + s^2)} \sum_{\mathbf{k},\zeta} \frac{2\omega_{\mathbf{k}} \langle 0|E_\nu(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle\langle 1_{\mathbf{k},\zeta}|E_\mu(\mathbf{r}',\omega_{\mathbf{k}})|0\rangle}{\hbar\,(\omega_{\mathbf{k}}^2 + s^2)} \\ &= -\frac{2}{\pi\hbar} \sum_j \sum_{\mathbf{k},\zeta} |\langle n|\mathbf{d}|j\rangle \cdot \langle 0|\mathbf{E}(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle|^2 \int_0^\infty ds\, \frac{\omega_{jn}\,\omega_{\mathbf{k}}}{(\omega_{jn}^2 + s^2)(\omega_{\mathbf{k}}^2 + s^2)} \\ &= -\sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle n|\mathbf{d}|\mathrm{e}_j\rangle \cdot \langle 0|\mathbf{E}(\mathbf{r},\omega_{\mathbf{k}})|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar(\omega_{jn} + \omega_{\mathbf{k}} \operatorname{sgn}\omega_{jn})}, \end{aligned} \tag{14.357}$$

where we have used the generalization of the integral formula (14.165)

$$\int_0^\infty dx\, \frac{ab}{(a^2+x^2)(b^2+x^2)} = \frac{\pi}{2(a^2-b^2)}(a \operatorname{sgn} b - b \operatorname{sgn} a) \qquad (a,b \in \mathbb{R},\ a,b \neq 0). \tag{14.358}$$

(See Problem 14.5.) Thus, we don't *quite* recover the perturbation expression (14.355), due to the presence of the extra $\operatorname{sgn}\omega_{jn}$, whenever $\omega_{jn} < 0$. However, note that we may write the total shift as

$$V_n = V_n^{(1)} + V_n^{(2)}, \tag{14.359}$$

(total level shift)

where $V_n^{(1)}$ is the expression (14.357) that describes the energy shift of $|n\rangle$, and $V_n^{(2)}$ is the difference between the full shift (14.354) and $V_n^{(1)}$ in the form (14.357):

$$\begin{aligned} V_n^{(2)} &= -\sum_j \sum_{\mathbf{k},\zeta} \frac{|\langle j|\mathbf{d}|n\rangle \cdot \langle 0|\mathbf{E}|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar} \left(\frac{1}{\omega_{jn} + \omega_{\mathbf{k}}} - \frac{1}{\omega_{jn} + \omega_{\mathbf{k}} \operatorname{sgn} \omega_{jn}} \right) \\ &= -\sum_j \Theta(\omega_{nj}) \sum_{\mathbf{k},\zeta} \frac{|\langle j|\mathbf{d}|n\rangle \cdot \langle 0|\mathbf{E}|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar} \left(\frac{1}{\omega_{jn} + \omega_{\mathbf{k}}} - \frac{1}{\omega_{jn} - \omega_{\mathbf{k}}} \right) \\ &= 2 \sum_j \Theta(\omega_{nj}) \sum_{\mathbf{k},\zeta} \frac{\omega_{\mathbf{k}} |\langle j|\mathbf{d}|n\rangle \cdot \langle 0|\mathbf{E}|1_{\mathbf{k},\zeta}\rangle|^2}{\hbar \left(\omega_{jn}^2 - \omega_{\mathbf{k}}^2\right)}. \end{aligned} \tag{14.360}$$

Notice that the Heaviside function $\Theta(\omega_{nj})$ "activates" whenever $E_n > E_j$, that is, whenever $|n\rangle$ acts as an excited state with respect to $|j\rangle$. Then we may use the Kramers–Heisenberg formula (14.159) for the Green tensor to eliminate the electric-field matrix elements and the sum over modes to obtain

$$V_n^{(2)} = -\sum_j \Theta(\omega_{nj}) \langle n|d_\alpha|j\rangle \langle j|d_\beta|n\rangle \operatorname{Re}[G_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega_{jn})].$$

(extra level shift for excited states) (14.361)

The extra shift here is due to the interaction of the atomic dipole with fields at the atomic resonance frequencies for every transition where $|n\rangle$ is the excited state. In fact, we may regard this shift as the Stark shift of the atom due to coupling to its own field. We have already treated this in the case of a perfectly conducting plane in terms of a mode sum in Section 13.7 along with the comparison to the same predictions of the Lorentz model in Section 1.5.1. From our previous discussion, we may conclude that this shift is a *classical* shift, which is consistent with its form: it is simply the coupling of the dipole covariance matrix to the Green tensor, which represents the light backscattered to the atom by the external body. We can see this directly from our construction of the Green tensor in Section (14.1.3). Given a classical dipole d_β at location $\mathbf{r}$ and oscillating at frequency ω, $d_\beta\, G_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega)$ gives the electric field at $\mathbf{r}$ due to the dipole and any other bodies that may reflect or otherwise influence the dipole's radiated field. The interaction energy will then be the product of the original dipole with the field, or $d_\alpha\, d_\beta\, G_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega)$. The real part is then taken in Eq. (14.361) since the Green tensor represents a complex field amplitude, and the Heaviside function limits this mechanism to radiative (excited) states. The Green tensor is then evaluated only at the transition frequencies, since those are the frequencies of the dipole radiation; the radiation rates of each transition are given by the magnitudes of the corresponding dipole matrix elements.

Again, we must renormalize this shift to remove the divergent free-field contribution, with the result[37]

$$\begin{aligned} V_n &= V_n^{(1)} + V_n^{(2)} \\ V_n^{(1)} &= -\frac{\hbar}{2\pi} \int_0^\infty ds\, \alpha_{\mu\nu}^{(n)}(is)\, G_{\nu\mu}^{(\mathrm{s})}(\mathbf{r},\mathbf{r},is) \\ V_n^{(2)} &= -\sum_j \Theta(\omega_{nj}) \langle n|d_\alpha|j\rangle \langle j|d_\beta|n\rangle \operatorname{Re}[G_{\alpha\beta}^{(\mathrm{s})}(\mathbf{r},\mathbf{r},\omega_{jn})] \end{aligned} \tag{14.362}$$

(renormalized level shift)

for the shift of an arbitrary atomic level $|n\rangle$.

14.3.9.1 Example: Spherically Symmetric Atom, Perfectly Conducting Plane

As an example of applying this formalism, let's consider a spherically symmetric atom near a perfectly conducting plate. The spherically symmetric atom has a scalar polarizability, and hence the shift reduces

[37] cf. J. M. Wylie and J. E. Sipe, "Quantum electrodynamics near an interface. II," *Physical Review A* **32**, 2030 (1985) (doi: 10.1103/PhysRevA.32.2030), Eqs. (4.3)-(4.4); also Werner Vogel and Dirk-Gunnar Welsch, *Quantum Optics*, 3rd ed. (Wiley-VCH, 2006), Eqs. (10.75)-(10.77), noting the different normalization convention there for the Green tensor.

to

$$\begin{aligned} V_n &= V_n^{(1)} + V_n^{(2)} \\ V_n^{(1)} &= -\frac{\hbar}{2\pi}\int_0^\infty ds\,\alpha^{(n)}(is)\,G^{(\mathrm{s})}_{\mu\mu}(\mathbf{r},\mathbf{r},is) \\ V_n^{(2)} &= -\sum_j \Theta(\omega_{nj})|\langle n|d_z|j\rangle|^2\,\mathrm{Re}[G^{(\mathrm{s})}_{\alpha\alpha}(\mathbf{r},\mathbf{r},\omega_{jn})] \end{aligned}$$

(spherically symmetric atom) (14.363)

The $V_n^{(1)}$ part of the shift is already given in terms of the reflection coefficients by (14.186),

$$V_n^{(1)} = \frac{\hbar}{8\pi^2\epsilon_0 c^2}\int_0^\infty ds\,s^2\,\alpha^{(n)}(is)\int_0^\infty dk_\tau\,\frac{k_\tau}{\kappa}\left[r_\perp(\theta,is) + \left(1+\frac{2k_\tau^2c^2}{s^2}\right)r_\parallel(\theta,is)\right]e^{-2\kappa z}, \tag{14.364}$$

while we now must compute $V_n^{(2)}$. The diagonal Green-tensor components are given by Eqs. (14.178) and (14.182) and the trace of the Green tensor is thus given by

$$G^{(s)}_{\alpha\alpha}(z,z,\omega) = \frac{i}{4\pi\epsilon_0}\frac{\omega^2}{c^2}\int_0^\infty dk_\tau\,\frac{k_\tau}{k_z}\left[r_\perp(\theta,\omega) + \left(1-\frac{2k_\tau^2c^2}{\omega^2}\right)r_\parallel(\theta,\omega)\right]e^{i2k_zz}, \tag{14.365}$$

where $k_z = \sqrt{\omega^2/c^2 - k_\tau^2}$. Then $V_n^{(2)}$ can be written directly in terms of this form.

For a perfect conductor, we replace the reflection coefficients by unity, and so

$$\begin{aligned} G^{(s)}_{\alpha\alpha}(z,z,\omega) &= \frac{i}{2\pi\epsilon_0}\frac{\omega^2}{c^2}\int_0^\infty dk_\tau\,\frac{k_\tau}{k_z}\left(1-\frac{k_\tau^2c^2}{\omega^2}\right)e^{i2k_zz} \\ &= \frac{1}{8\pi\epsilon_0}\partial_z\int_0^\infty d(k_\tau^2)\,e^{i2z\sqrt{\omega^2/c^2-k_\tau^2}} \\ &= \frac{1}{16\pi\epsilon_0}\partial_z\left(\frac{i2\omega z}{c}-1\right)\frac{e^{i2\omega z/c}}{z^2} \\ &= \frac{1}{16\pi\epsilon_0}\partial_z^2\frac{e^{i2\omega z/c}}{z}. \end{aligned} \tag{14.366}$$

Putting this into (14.363), we find

$$V_n^{(2)}(z) = -\frac{1}{16\pi\epsilon_0}\sum_j \Theta(\omega_{nj})|\langle n|d_z|j\rangle|^2\,\partial_z^2\frac{1}{z}\cos(2k_{jn}z), \tag{14.367}$$

where $k_{jn} = \omega_{jn}/c$. We already computed the other part of the shift in Eq. (14.194), with the result

$$V_n^{(1)}(z) = -\frac{\operatorname{sgn}\omega_{jn}}{16\pi^2\epsilon_0}\sum_j |\langle n|d_z|j\rangle|^2\,\partial_z^2\frac{1}{z}f(2|k_{jn}|z), \tag{14.368}$$

though we have had to modify it here by introducing the factor $\operatorname{sgn}\omega_{jn}$ and introducing the absolute value of k_{jn} in the argument of $f(z)$. This is because due to the form (14.356) of the polarizability, the integral for $V_n^{(1)}$ in Eqs. (14.363) is of the form

$$\begin{aligned} \int_0^\infty ds\,\frac{\omega_{jn}\,h(s)}{\omega_{jn}^2+s^2} &= \frac{1}{2}\int_{-\infty}^\infty ds\frac{\omega_{jn}\,h(s)}{(s+i|\omega_{jn}|)(s-i|\omega_{jn}|)} \\ &= \frac{1}{2}\frac{2\pi i\omega_{jn}\,h(i|\omega_{jn}|)}{2i|\omega_{jn}|} \\ &= \frac{\pi}{2}h(i|\omega_{jn}|)\operatorname{sgn}\omega_{jn} \end{aligned} \tag{14.369}$$

for the appropriate (even) function $h(s)$, where we have used Cauchy's integral formula applied to the contour around the upper half-plane. Of course, for the ground state $\omega_{jn} > 0$, so the absolute value and sgn function were unnecessary, but they are needed now.

Putting these parts together, the total level shift is

$$V_n(z) = V_n^{(1)}(z) + V_n^{(2)}(z) = -\frac{\operatorname{sgn}\omega_{jn}}{16\pi^2\epsilon_0}\sum_j |\langle n|d_z|j\rangle|^2\,\partial_z^2\frac{1}{z}\Big[f(|2k_{j0}|z) - \Theta(\omega_{nj})\pi\cos(2|k_{jn}|z)\Big],$$

(level shift near perfectly conducting plane, spherically symmetric atom) (14.370)

This agrees with the result of our previous mode-summation calculation, Eq. (13.68), if we restrict that result to a spherically symmetric atom. Of course, the formalism here covers the case of an anisotropic atom, with only a bit more work.

14.3.10 Lifetime Shifts

Now we will consider the complementary problem to the body-induced level shifts: the shifts in the lifetimes or decay rates of atomic excited levels. In our classical treatment (Section 1.5) of this problem in the special cases of an atom near a planar mirror or another atom, we saw that the shifts of the decay rate and the transition frequency were different aspects of the same effect. Here we treat the decay-rate shifts separately from the level shifts, due to the additional complexity of the formalism here.

14.3.10.1 Decay Rate Near a Macroscopic Body

To treat the general decay problem in the presence of a macroscopic body, we start with Fermi's golden rule in the form (11.58)

$$\Gamma_{\mathrm{i}\to\mathrm{f}} = \frac{2\pi}{\hbar}|\langle \mathrm{i}|H_{\mathrm{int}}|\mathrm{f}\rangle|^2\delta(E_\mathrm{i} - E_\mathrm{f}). \tag{14.371}$$

We are interested in the decay from atomic state $|\mathrm{i}\rangle$ to state $|\mathrm{f}\rangle$ due to the interaction with the field; we should thus also include initial and final field states:

$$\Gamma_{\mathrm{i}\to\mathrm{f}} = \frac{2\pi}{\hbar}\sum_{I\,F} P(I)\,|\langle \mathrm{i}\;I|H_{\mathrm{int}}|\mathrm{f}\;F\rangle|^2\delta(E_\mathrm{i} + E_I - E_\mathrm{f} - E_F). \tag{14.372}$$

Here, I and F are parameters labeling initial and final field states, respectively, and $P(I)$ is the initial ($t = 0$) occupation probability for the field state $|I\rangle$. If we use the integral representation of the delta function, this becomes

$$\begin{aligned}\Gamma_{\mathrm{i}\to\mathrm{f}} &= \frac{1}{\hbar^2}\sum_{I\,F} P(I)\int_{-\infty}^{\infty} d\tau\,|\langle \mathrm{i}\;I|H_{\mathrm{int}}|\mathrm{f}\;F\rangle|^2 e^{i(E_\mathrm{i}+E_I-E_\mathrm{f}-E_F)\tau/\hbar}\\ &= \frac{1}{\hbar^2}\sum_{I\,F} P(I)\int_{-\infty}^{\infty} d\tau\,|\langle \mathrm{i}\;I|H_{\mathrm{int}}|\mathrm{f}\;F\rangle|^2 e^{i\omega_{\mathrm{if}}\tau}e^{i(E_I-E_F)\tau/\hbar},\end{aligned} \tag{14.373}$$

where $\omega_{\mathrm{if}} = (E_\mathrm{i} - E_\mathrm{f})/\hbar > 0$ as usual. For the dipole interaction Hamiltonian,

$$\begin{aligned}\Gamma_{\mathrm{i}\to\mathrm{f}} &= \frac{1}{\hbar^2}\sum_{I\,F} P(I)\int_{-\infty}^{\infty} d\tau\,\big|\langle \mathrm{i}|\mathbf{d}|\mathrm{f}\rangle\cdot\langle I|\mathbf{E}(\mathbf{r})|F\rangle\big|^2 e^{i\omega_{\mathrm{if}}\tau}e^{i(E_I-E_F)\tau/\hbar}\\ &= \frac{1}{\hbar^2}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\sum_{I\,F} P(I)\int_{-\infty}^{\infty} d\tau\,\langle I|E_\alpha(\mathbf{r})|F\rangle\langle F|E_\beta(\mathbf{r})|F\rangle\,e^{i\omega_{\mathrm{if}}\tau}e^{i(E_I-E_F)\tau/\hbar}\\ &= \frac{1}{\hbar^2}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\sum_{I\,F} P(I)\int_{-\infty}^{\infty} d\tau\,\langle I|E_\alpha(\mathbf{r},\tau)|F\rangle\langle F|E_\beta(\mathbf{r},0)|I\rangle\,e^{i\omega_{\mathrm{if}}\tau}\\ &= \frac{1}{\hbar^2}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\sum_{I} P(I)\int_{-\infty}^{\infty} d\tau\,\langle I|E_\alpha(\mathbf{r},\tau)E_\beta(\mathbf{r},0)|I\rangle\,e^{i\omega_{\mathrm{if}}\tau}\\ &= \frac{1}{\hbar^2}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\int_{-\infty}^{\infty} d\tau\,\langle E_\alpha(\mathbf{r},\tau)E_\beta(\mathbf{r},0)\rangle\,e^{i\omega_{\mathrm{if}}\tau},\end{aligned} \tag{14.374}$$

where the final expectation value is an ensemble average over initial states, and we have transformed the field operators to the Heisenberg picture (technically, the interaction picture, since they evolve as if they were unperturbed).

To evaluate the field correlation function, we will need Eq. (14.130) for the commutator correlation function,

$$\langle [E_\alpha(\mathbf{r},\tau), E_\beta(\mathbf{r}',0)]\rangle = \frac{\hbar}{\pi}\int_{-\infty}^{\infty} d\omega\, \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)]\, e^{-i\omega\tau}, \tag{14.375}$$

along with the fluctuation–dissipation theorem (14.303),

$$\langle [E_\alpha(\mathbf{r},0), E_\beta(\mathbf{r}',\tau)]_+\rangle = \frac{\hbar}{\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\, \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] \coth\left[\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right]. \tag{14.376}$$

It follows, for example, from the Kramers–Heisenberg formulae (14.159) and (14.160) that the Green tensor is symmetric, $G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = G_{\beta\alpha}(\mathbf{r}',\mathbf{r},\omega)$, and thus we may switch the order of the anticommutator and rewrite the fluctuation–dissipation relation as

$$\langle [E_\alpha(\mathbf{r},\tau), E_\beta(\mathbf{r}',0)]_+\rangle = \frac{\hbar}{\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\, \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] \coth\left[\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right]. \tag{14.377}$$

We can alternately see this since the left-hand side is real, the time dependence on the right-hand side is of the form $\cos(\omega\tau)$; thus the correlation function is an even function of the time difference τ. Now adding Eqs. (14.375) and (14.377), we obtain

$$\begin{aligned}\langle E_\alpha(\mathbf{r},\tau) E_\beta(\mathbf{r}',0)\rangle &= \frac{\hbar}{2\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\, \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)] \left\{1 + \coth\left[\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right]\right\} \\ &= \frac{\hbar}{\pi}\int_{-\infty}^{\infty} d\omega\, e^{-i\omega\tau}\, \frac{\mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)]}{1-\exp\left[-\dfrac{\hbar\omega}{k_{\mathrm{B}}T}\right]}.\end{aligned} \tag{14.378}$$

Inverting the Fourier transform leads to

$$\int_{-\infty}^{\infty} d\tau\, \langle E_\alpha(\mathbf{r},\tau) E_\beta(\mathbf{r}',0)\rangle\, e^{i\omega\tau} = 2\hbar\, \frac{\mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)]}{1-\exp\left[-\dfrac{\hbar\omega}{k_{\mathrm{B}}T}\right]}, \tag{14.379}$$

and putting this into Eq. (14.374), we find[38]

$$\Gamma_{\mathrm{i}\to\mathrm{f}} = \frac{2}{\hbar}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\, \frac{\mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega_{\mathrm{if}})]}{1-\exp\left[-\dfrac{\hbar\omega_{\mathrm{if}}}{k_{\mathrm{B}}T}\right]}.$$

(atomic spontaneous decay rate, finite T) (14.380)

Notice that this rate diverges linearly with T when $k_{\mathrm{B}}T \gg \hbar\omega_{\mathrm{if}}$. Taking the $T \longrightarrow 0$ limit, we find

$$\Gamma_{\mathrm{i}\to\mathrm{f}} = \frac{2}{\hbar}\langle \mathrm{i}|d_\alpha|\mathrm{f}\rangle\langle \mathrm{f}|d_\beta|\mathrm{i}\rangle\, \mathrm{Im}[G_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega_{\mathrm{if}})]$$

(atomic spontaneous decay rate, $T = 0$) (14.381)

for the rate of spontaneous decay for the atomic $|\mathrm{i}\rangle \longrightarrow |\mathrm{f}\rangle$ transition. Comparing this result to the excited level shift $V^{(2)}$ from Eqs. (14.362) we see that the two expressions have the nearly same form. The obvious

[38] cf. J. M. Wylie and J. E. Sipe, "Quantum electrodynamics near an interface," *Physical Review A* **30**, 1185 (1984) (doi: 10.1103/PhysRevA.30.1185), Eq. (2.5). The derivation here is essentially identical to theirs. For the zero-temperature limit, see also J. M. Wylie and J. E. Sipe, "Quantum electrodynamics near an interface. II," *Physical Review A* **32**, 2030 (1985) (doi: 10.1103/PhysRevA.32.2030), Eq. (4.5); and Werner Vogel and Dirk-Gunnar Welsch, *Quantum Optics*, 3rd ed. (Wiley, 2006), Eq. (10.27), again noting the different normalization convention there for the Green tensor.

differences here are: the extra factor of $2/\hbar$ here because we are considering a decay rate of population rather than a level shift; the absence here of a level sum and a Heaviside function, although we are implicitly considering only transitions of unstable excited states to lower-energy states, and the *total* decay rate from any level follows from summing over all posssible decay paths; and finally, the important difference is the presence of the imaginary part of the Green tensor, as opposed to the real part from the level shift. We saw precisely this dependence on the field quadratures before in the classical treatment of these problems [cf. Eq. (1.125)]. This is also what we expected from what we know about generalized susceptibilities from Section 14.1.4.1; the imaginary part alone leads to dissipation.

14.3.10.2 Free Space: Green-Tensor Example

Now we can work out the general decay rate (14.381) for the case of free space. Recall that the free-space Green tensor is, from Eq. (14.46),

$$G^{(0)}_{\alpha\beta}(\mathbf{r},0,\omega) = \frac{1}{4\pi\epsilon_0}\left\{[3\hat{r}_\alpha\hat{r}_\beta - \delta_{\alpha\beta}]\left[\frac{1}{r^3} - i\frac{k}{r^2}\right] - [\hat{r}_\alpha\hat{r}_\beta - \delta_{\alpha\beta}]\,\frac{k^2}{r}\right\} e^{ikr}. \tag{14.382}$$

In computing the decay rate, we need a product of the form $d_\alpha G^{(0)}_{\alpha\beta} d_\beta$. The first (near-field) term of the free-space Green tensor then gives a contribution of the form $d_\alpha[3\hat{r}_\alpha\hat{r}_\beta - \delta_{\alpha\beta}]d_\beta = 3(\mathbf{d}\cdot\hat{r})^2 - d^2$, which vanishes for a spherically symmetric atom. The tensor second term is of the form $-d_\alpha[\hat{r}_\alpha\hat{r}_\beta - \delta_{\alpha\beta}]d_\beta = d^2 - (\mathbf{d}\cdot\hat{r})^2 = 2d^2/3$; taking the imaginary part and then letting $\mathbf{r} \longrightarrow 0$ gives a factor

$$\lim_{r\to 0}\frac{k^2\mathrm{Im}[e^{ikr}]}{r} = k^3. \tag{14.383}$$

Putting all the pieces together, we find the free-space decay rate for a two-level atom:

$$\Gamma_0 = \frac{2}{\hbar}\langle \mathrm{e}|d_\alpha|\mathrm{g}\rangle\langle \mathrm{g}|d_\beta|\mathrm{e}\rangle\,\mathrm{Im}[G^{(0)}_{\alpha\beta}(0,0,\omega_{\mathrm{eg}})] = \frac{\omega^3_{\mathrm{eg}}|\langle \mathrm{g}|\mathbf{d}|\mathrm{e}\rangle|^2}{3\pi\epsilon_0\hbar c^3}.$$

(free-space atomic spontaneous decay rate) (14.384)

This is the same result that we found directly from Fermi's golden rule, Eq. (11.68), and of course our result, Eq. (11.29), from our previous Weisskopf–Wigner calculation.

In general, then, we can write the decay rate in the presence of a macroscopic body as

$$\Gamma(\mathbf{r}) = \Gamma_0 + \frac{2}{\hbar}\langle \mathrm{e}|d_\alpha|\mathrm{g}\rangle\langle \mathrm{g}|d_\beta|\mathrm{e}\rangle\,\mathrm{Im}[G^{(\mathrm{s})}_{\alpha\beta}(\mathbf{r},\mathbf{r},\omega_{\mathrm{eg}})],$$

(atomic spontaneous decay rate) (14.385)

so that the deviation from the free-space decay rate is given in terms of the scattering part of the Green tensor.

14.3.10.3 Planar Reflector

As another example, we take the case of an atom near a planar interface. We recall from Eqs. (14.182) and (14.178) the nonvanishing Green-tensor components

$$\begin{aligned} G^{(s)}_{xx}(z,z,\omega) = G^{(s)}_{yy}(z,z,\omega) &= \frac{i}{8\pi\epsilon_0}\int_0^\infty dk_\intercal\,\frac{k_\intercal}{k_z}\left(k_z^2 r_\parallel(\theta,\omega) + k^2 r_\perp(\theta,\omega)\right) e^{2ik_z z} \\ G^{(s)}_{zz}(z,z,\omega) &= -\frac{i}{4\pi\epsilon_0}\int_0^\infty dk_\intercal\,\frac{k_\intercal^3}{k_z}\,r_\parallel(\theta,\omega)\,e^{2ik_z z}. \end{aligned} \tag{14.386}$$

Taking the matrix elements to be the same in the x and y directions, the shift in the decay rate due to the interface is

$$\begin{aligned}\delta\Gamma(z) = \Gamma(z) - \Gamma_0 &= \frac{2}{\hbar}\langle \mathrm{e}|d_\alpha|\mathrm{g}\rangle\langle \mathrm{g}|d_\beta|\mathrm{e}\rangle\,\mathrm{Im}[G^{(\mathrm{s})}_{\alpha\beta}(z,z,\omega_{\mathrm{eg}})] \\ &= \frac{1}{2\pi\epsilon_0\hbar}\mathrm{Re}\left\{\int_0^\infty dk_\tau\,\frac{k_\tau}{k_z}\left[(d^2_{\mathrm{ge},\parallel}/2)\Big(k_z^2 r_\parallel(\theta,\omega_{\mathrm{eg}}) + k_{\mathrm{eg}}^2 r_\perp(\theta,\omega_{\mathrm{eg}})\Big) - d^2_{\mathrm{ge},z}\left(k_{\mathrm{eg}}^2 - k_z^2\right) r_\parallel(\theta,\omega_{\mathrm{eg}})\right] e^{2ik_z z}\right\}\end{aligned}$$

(decay-rate shift near planar interface) (14.387)

where $k_\tau^2 + k_z^2 = k_{\mathrm{eg}}^2 = (\omega_{\mathrm{eg}}/c)^2$. Unfortunately, we can't go much further here without a specific assumption regarding the angle dependence of the reflection coefficients.

Near a perfectly conducting interface, we may set both reflection coefficients to unity. Then we change integration variables to $\kappa = -ik_z$, so that $k_\tau^2 - \kappa^2 = k_{\mathrm{eg}}^2$, with the result

$$\begin{aligned}\delta\Gamma(z) &= \frac{k_{\mathrm{eg}}^2}{2\pi\epsilon_0\hbar}\left[\left(d^2_{\mathrm{ge},\parallel}/2 - d^2_{\mathrm{ge},z}\right) - \frac{1}{4k_{\mathrm{eg}}^2}\left(d^2_{\mathrm{ge},\parallel}/2 + d^2_{\mathrm{ge},z}\right)\partial_z^2\right]\mathrm{Re}\left\{i\int_{-ik_{\mathrm{eg}}}^\infty d\kappa\, e^{-2\kappa z}\right\} \\ &= \frac{k_{\mathrm{eg}}^2}{4\pi\epsilon_0\hbar}\left[\left(d^2_{\mathrm{ge},\parallel}/2 - d^2_{\mathrm{ge},z}\right) - \frac{1}{4k_{\mathrm{eg}}^2}\left(d^2_{\mathrm{ge},\parallel}/2 + d^2_{\mathrm{ge},z}\right)\partial_z^2\right]\frac{1}{z}\,\mathrm{Re}\left[ie^{2ik_{\mathrm{eg}}z}\right] \\ &= -\frac{k_{\mathrm{eg}}^2}{4\pi\epsilon_0\hbar}\left[\left(d^2_{\mathrm{ge},\parallel}/2 - d^2_{\mathrm{ge},z}\right) - \frac{1}{4k_{\mathrm{eg}}^2}\left(d^2_{\mathrm{ge},\parallel}/2 + d^2_{\mathrm{ge},z}\right)\partial_z^2\right]\frac{\sin(2k_{\mathrm{eg}}z)}{z} \\ &= -\frac{k_{\mathrm{eg}}^3}{2\pi\epsilon_0\hbar}\left[\left(d^2_{\mathrm{ge},\parallel}/2 - d^2_{\mathrm{ge},z}\right) - \left(d^2_{\mathrm{ge},\parallel}/2 + d^2_{\mathrm{ge},z}\right)\partial_{z'}^2\right]\frac{\sin z'}{z'},\end{aligned} \tag{14.388}$$

where we have switched to the scaled coordinate $z' := 2k_{\mathrm{eg}}z$. Now using the free-space decay rate (14.384), we may write the shift as

$$\delta\Gamma(z) = -\frac{3}{2}\Gamma_0\left[\left(\hat{\varepsilon}_\parallel^2/2 - \hat{\varepsilon}_z^2\right) - \left(\hat{\varepsilon}_\parallel^2/2 + \hat{\varepsilon}_z^2\right)\partial_{z'}^2\right]\frac{\sin z'}{z'},$$

(decay-rate shift near perfect mirror) (14.389)

where $\hat{\varepsilon}_\parallel^2 := d^2_{\mathrm{ge},\parallel}/d^2_{\mathrm{ge}}$ and $\hat{\varepsilon}_z^2 := d^2_{\mathrm{ge},z}/d^2_{\mathrm{ge}}$. This result is identical to what we derived using the classical model, Eq. (1.132). Recall that the spatial dependence of the quantum modes is purely classical, and that's what determines the spatial dependence of the shift. Evidently, all the "quantumness" in this relation is buried in the details of the free-space decay rate Γ_0.

14.4 Casimir Energy

Thus far, we have concentrated on the evaluation of Casimir–Polder energies—energies of an atom (or molecule or other point particle) interacting with a dielectric body or bodies. Here we will (relatively briefly) consider the analogous problem of **Casimir energies**, or interactions involving only macroscopic bodies. Of course, the Casimir–Polder effect is just a special case of this more general Casimir effect, and we will take advantage of this fact.

In the case where we neglect dispersion, it is relatively straightforward to come up with an expression for the Casimir energy within linear-response theory. For example, for a dispersionless medium with dielectric and magnetic response, we can write the field Hamiltonian as [see Eq. (8.309) for the dielectric case, and we are also appropriately generalizing the magnetic response here]

$$H = \int d^3r\,(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}) = \int d^3r\left[\epsilon(\mathbf{r})E^2 + \frac{1}{\mu(\mathbf{r})}B^2\right].$$

(electromagnetic Hamiltonian) (14.390)

The Casimir energy V_{C} is then just the expectation value in the vacuum state:

$$V_{\mathrm{C}} = \int d^3r\left\langle \epsilon(\mathbf{r})E^2 + \frac{1}{\mu(\mathbf{r})}B^2\right\rangle. \tag{14.391}$$

Formally this is a divergent quantity, as in the Casimir–Polder case, where we subtracted the vacuum contribution to the energy in Eq. (14.167). Here, we will simply note that it is important to compute the *difference* in energy between two configurations of the dielectric bodies in order to obtain a sensible result. To write this energy in terms of the Green tensor, we can adapt the correlation function (14.303), with $T = 0$ and at equal times, so that

$$\langle [E_\mu(\mathbf{r}), E_\nu(\mathbf{r}')]_+ \rangle = \frac{\hbar}{\pi} \int_{-\infty}^{\infty} d\omega \, \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)] \, \mathrm{sgn}(\omega). \tag{14.392}$$

Adding this to the $\tau = 0$ form of the anticommutator version (14.130),

$$\langle [E_\mu(\mathbf{r}), E_\nu(\mathbf{r}')] \rangle = \frac{\hbar}{\pi} \int_{-\infty}^{\infty} d\omega \, \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)], \tag{14.393}$$

we obtain

$$\langle E_\mu(\mathbf{r}) \, E_\nu(\mathbf{r}') \rangle = \frac{\hbar}{\pi} \int_{0}^{\infty} d\omega \, \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)]. \tag{14.394}$$

For the magnetic contribution, we can similarly write

$$\langle \partial_t B_\mu(\mathbf{r}) \, \partial_t B_\nu(\mathbf{r}') \rangle = \nabla \times \nabla' \times \langle E_\mu(\mathbf{r}) \, E_\nu(\mathbf{r}') \rangle = \frac{\hbar}{\pi} \int_{0}^{\infty} d\omega \, \nabla \times \nabla' \times \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)], \tag{14.395}$$

or

$$\langle B_\mu(\mathbf{r}) \, B_\nu(\mathbf{r}') \rangle = -\frac{\hbar}{\pi} \int_{0}^{\infty} d\omega \, \omega^2 \, \nabla \times \nabla' \times \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)] = \frac{\hbar}{\pi} \int_{0}^{\infty} \frac{d\omega}{\omega^2} \, \nabla \times \mathrm{Im}[G_{\mu\nu}(\mathbf{r}, \mathbf{r}', \omega)] \times \overleftarrow{\nabla}'. \tag{14.396}$$

In these expressions, ∇ operates only on $\mathbf{r}$, ∇' operates only on $\mathbf{r}'$, and the arrow on the last ∇' indicates that it operates to the left on the Green tensor. Putting these expressions into Eq. (14.391), we have[39]

$$V_{\mathrm{C}} = \frac{\hbar}{\pi} \int d^3r \int_0^\infty d\omega \left[\epsilon(\mathbf{r}) \, \mathrm{Im}[G_{\mu\mu}(\mathbf{r}, \mathbf{r}', \omega)] + \frac{1}{\mu(\mathbf{r})\omega^2} \nabla \times \mathrm{Im}[G_{\mu\mu}(\mathbf{r}, \mathbf{r}', \omega)] \times \overleftarrow{\nabla}' \Big|_{\mathbf{r}=\mathbf{r}'} \right], \tag{14.397}$$

(Casimir energy, no dispersion)

where the repeated μ's are summed.

Although it is reasonably simple to obtain this expression in the absence of dispersion, the case where we include dispersion is more subtle, and must be handled with some care. For example, simply replacing $\epsilon(\mathbf{r})$ by $\epsilon(\mathbf{r}, \omega)$ does not work in either Eq. (14.391) or Eq. (14.390).

14.4.1 Casimir Energy of Dispersive Dielectric Bodies

To incorporate dispersion into the Casimir energy, we will take advantage of the fact that we have *already* incorporated dispersion into the *Casimir–Polder* energy, which in terms of an imaginary-frequency integral is

$$V_{\mathrm{CP}} = -\frac{\hbar}{2\pi} \int_0^\infty ds \, \mathrm{Tr} \Big[\boldsymbol{\alpha}(is) \cdot \mathbf{G}^{(\mathrm{s})}(\mathbf{r}, \mathbf{r}, is) \Big], \tag{14.398}$$

from Eq. (14.167), or as a real-frequency integral from Eq. (14.327),

$$V_{\mathrm{CP}} = -\frac{\hbar}{2\pi} \mathrm{Im} \int_0^\infty d\omega \, \mathrm{Tr}[\boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(\mathrm{s})}(\mathbf{r}, \mathbf{r}, \omega)] \tag{14.399}$$

after taking the zero-temperature limit, exploiting the even nature of the integrand, and renormalizing by subtracting the free-space Green tensor.

[39] A similar expression can be found in K. A. Milton, *The Casimir Effect: Physical Manifestations of Zero-Point Energy* (World Scientific, 2001), p. 84, Eq. (5.27a) (ISBN: 9810243979).

Although these corrections correctly model a dispersive interaction of an atom with a macroscopic body (or collection of bodies), it implicitly assumes that the effect of introducing the atom has a negligible effect on the vacuum field (e.g., via the Green tensor). In introducing a *macroscopic* body in the context of the Casimir effect, this is no longer true, and we have to be more careful with the notion of the interaction energy. This procedure will allow us to generalize these expressions into the correct expressions to model Casimir energies for dispersive objects.[40] To simplify this treatment somewhat, we will only consider dielectric bodies, ignoring any magnetic effects.

14.4.1.1 Induced Interaction Energies

First, we will have a short digression to treat the general notion of an ''induced'' interaction energy.[41] That is, consider the Hamiltonian

$$H_\lambda = H_0 + \lambda V, \tag{14.400}$$

which is parameterized by λ, which satisfies $0 \leq \lambda \leq 1$. Then λ smoothly interpolates between the ''unperturbed'' Hamiltonian H_0 and the ''perturbed'' Hamiltonian $H_0 + V$. Unlike perturbation theory, however, we will not make any assumptions about having a weak perturbation: the interaction V may be as strong as we like.

We will assume we know the energy E_λ corresponding to an eigenstate $|E_\lambda\rangle$, for every possible value of λ:

$$H_\lambda |E_\lambda\rangle = E_\lambda |E_\lambda\rangle. \tag{14.401}$$

Often this technique will be useful for analyzing ground-state shifts (such as Casimir energies), but we need not assume this explicitly. Now differentiating the energy E_λ with respect to λ, we obtain

$$\begin{aligned}
\frac{\partial E_\lambda}{\partial \lambda} &= \frac{\partial}{\partial \lambda}\langle E_\lambda | H_\lambda | E_\lambda \rangle \\
&= \left[\frac{\partial}{\partial \lambda}\langle E_\lambda|\right] H_\lambda |E_\lambda\rangle + \langle E_\lambda | H_\lambda \left[\frac{\partial}{\partial \lambda}|E_\lambda\rangle\right] + \langle E_\lambda | \frac{\partial H_\lambda}{\partial \lambda} |E_\lambda\rangle \\
&= E_\lambda \left[\frac{\partial}{\partial \lambda}\langle E_\lambda|\right] |E_\lambda\rangle + E_\lambda \langle E_\lambda| \left[\frac{\partial}{\partial \lambda}|E_\lambda\rangle\right] + \langle E_\lambda | V | E_\lambda\rangle \\
&= E_\lambda \frac{\partial}{\partial \lambda}\langle E_\lambda | E_\lambda\rangle + \langle E_\lambda | V | E_\lambda\rangle \\
&= \langle E_\lambda | V | E_\lambda\rangle \\
&= \frac{1}{\lambda}\langle E_\lambda | \lambda V | E_\lambda\rangle.
\end{aligned} \tag{14.402}$$

Note that in the next-to-last equality, we used $\langle E_\lambda | E_\lambda\rangle = 1$. The result above is often called the **Hellmann–Feynman theorem**;[42] we will write this in integral form here as

$$E_1 - E_0 = \int_0^1 \frac{d\lambda}{\lambda} \langle E_\lambda | \lambda V | E_\lambda\rangle. \tag{14.403}$$

(Hellmann–Feynman theorem)

This gives the energy difference $E_1 - E_0$ between the system in the presence and absence of the interaction V.

[40] From here through Section 14.4.1.7, we are closely following the derivation of F. S. S. Rosa, D. A. R. Dalvit, and P. W. Milonni, ''Electromagnetic energy, absorption, and Casimir forces. II. Inhomogeneous dielectric media,'' *Physical Review A* **84**, 053813 (2011) (doi: 10.1103/PhysRevA.84.053813).

[41] Alexander L. Fetter and John Dirk Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, 1971), pp. 69-70 (ISBN: 0070206538).

[42] For a history of this theorem, see David Wallace, *An Introduction To Hellmann–Feynman Theory*, Master's thesis (University of Central Florida, 2005), available at `http://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1412&context=etd`.

As a simple example, consider the energy due to an induced dipole in a static electric field. Setting $\omega = 0$ in the Kramers–Heisenberg expression (14.148) for the polarizability gives the static polarizability

$$\alpha_0 = \sum_j \frac{2|\langle \mathrm{g}|d_z|\mathrm{e}_j\rangle|^2}{\hbar\omega_{j0}}. \tag{14.404}$$

The interaction Hamiltonian for the dipole–field interaction is $V = -\mathbf{d}\cdot\mathbf{E}$. The perturbation-theory analysis of the dipole response in Section 14.3.1 gives $\mathbf{d} = \alpha_0\mathbf{E}$ in the static limit. Then $\langle V\rangle = \alpha_0(\lambda)E^2$, where the λ-dependent polarizability is $\alpha_0(\lambda) = \lambda^2\alpha_0$, because if we associate λ with the dipole moment, we can note that $V \sim d$ but $\alpha_0 \sim d^2$. Thus the energy shift (14.403) becomes

$$\Delta E = \int_0^1 \frac{d\lambda}{\lambda}\,\alpha_0(\lambda)\,E^2, \tag{14.405}$$

and so putting in the explicit factor of λ^2, we find

$$\Delta E = \int_0^1 d\lambda\,\lambda\,\alpha_0 E^2 = \frac{1}{2}\alpha_0 E^2. \tag{14.406}$$

This is exactly what we expect for the energy of an induced dipole, specifically the factor of $1/2$. For example, we can obtain the same result by bringing the electric field up from zero to the final value, integrating the work in doing this along the way, we obtain

$$\Delta E = -\int_0^E dE'\,d(E') = -\int_0^E dE'\,\alpha_0 E' = \frac{1}{2}\alpha_0 E^2, \tag{14.407}$$

which gives the same factor of $1/2$. However, Eq. (14.405) is a more general expression than $(1/2)\alpha_0E^2$ of the induced dipole energy, because different functional forms for $\alpha_0(\lambda)$, due to more complicated interactions, can lead to different final expressions for the energy. This is precisely the path we will follow in adapting the Casimir–Polder potential to the Casimir energy.

14.4.1.2 Assembling a Dielectric Medium, Atom by Atom

Now we will model the dielectric media as a collection of N atoms. Since a dipole responds to an applied electric field via the tensor polarizability $\boldsymbol{\alpha}$ as

$$\mathbf{d} = \boldsymbol{\alpha}\cdot\mathbf{E}, \tag{14.408}$$

we can write the dipole response $\mathbf{d}_n(\omega)$ at frequency ω for each atom at $\mathbf{r}_n$ as

$$\mathbf{d}_n(\omega) = \boldsymbol{\alpha}(\omega)\cdot\mathbf{E}(\mathbf{r}_n,\omega). \tag{14.409}$$

We assume the electric field to comprise an externally applied field $\mathbf{E}_0$ as well as the induced field from each dipole. The latter contribution can be written in terms of the free-space Green tensor $\mathbf{G}_0$, so that the total dipole response for the nth atom is

$$\mathbf{d}_n(\omega) = \boldsymbol{\alpha}(\omega)\cdot\mathbf{E}_0(\mathbf{r}_n,\omega) + \boldsymbol{\alpha}(\omega)\cdot\sum_{j=1}^{N}\mathbf{G}^{(0)}(\mathbf{r}_n,\mathbf{r}_j,\omega)\cdot\mathbf{d}_j(\omega). \tag{14.410}$$

Grouping the dipole moments together, we may formally write this system of equations in matrix form as

$$\Big[1 - \boldsymbol{\alpha}(\omega)\cdot\mathbf{G}^{(0)}(\omega)\Big]\cdot\mathbf{d}(\omega) = \boldsymbol{\alpha}(\omega)\cdot\mathbf{E}_0(\omega), \tag{14.411}$$

where the matrices encode both polarization and ‘‘which-atom’’ degrees of freedom. That is, the ‘‘1’’ in the left-hand matrix should be interpreted as the $3N\times 3N$ identity matrix $\mathcal{I}_{3N}$, and $\boldsymbol{\alpha}(\omega)$ and $\mathbf{G}_0(\omega)$ are

similarly $3N \times 3N$, while $\mathbf{d}(\omega)$ and $\mathbf{E}_0(\omega)$ are column vectors of dimension $3N$. In this form, we may just as well introduce the freedom to assign a different polarizability tensor to each atom, which corresponds to replacing $\boldsymbol{\alpha}(\omega)$ by $\boldsymbol{\alpha}^{(n)}(\omega)$ in Eq. (14.409). Then solving for the collection of dipole moments, we obtain

$$\mathbf{d}(\omega) = \Big[1 - \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega)\Big]^{-1} \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{E}_0(\omega). \tag{14.412}$$

The collective polarizability of the N atomic dipoles is just the entire matrix that acts on the field vector $\mathbf{E}_0(\omega)$ on the right-hand side. Thus, to adapt the unrenormalized form of the single-atom interaction energy (14.399), we can replace the polarizability in that expression with the collective polarizability and the Green tensor with the free-space tensor, with the result

$$V_{\text{C}} = -\frac{\hbar}{\pi} \int_0^1 \frac{d\lambda}{\lambda} \,\text{Im} \int_0^\infty d\omega \,\text{Tr}\bigg\{ \Big[1 - \boldsymbol{\alpha}(\omega,\lambda) \cdot \mathbf{G}^{(0)}(\omega)\Big]^{-1} \cdot \boldsymbol{\alpha}(\omega,\lambda) \cdot \mathbf{G}^{(0)}(\omega) \bigg\}, \tag{14.413}$$

which follows from introducing the correct integral over the interaction parameter λ as in Eq. (14.403). As in the induced-dipole example of Eq. (14.405), the coupled form of the polarizability tensor is $\boldsymbol{\alpha}(\omega,\lambda) = \lambda^2 \boldsymbol{\alpha}(\omega)$, so that

$$V_{\text{C}} = -\frac{\hbar}{\pi} \,\text{Im} \int_0^\infty d\omega \,\text{Tr}\bigg\{ \int_0^1 d\lambda \,\lambda \Big[1 - \lambda^2 \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega)\Big]^{-1} \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \bigg\}. \tag{14.414}$$

Note again that the trace here extends over not only the polarization degrees of freedom, but also the coordinates of all N atoms. Hence, this expression corresponds to a Casimir energy of the entire collection of atoms. Carrying out the λ integral then gives the expression

$$V_{\text{C}} = \frac{\hbar}{2\pi} \,\text{Im} \int_0^\infty d\omega \,\text{Tr} \log \Big[1 - \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}_0(\omega)\Big] \tag{14.415}$$

for the Casimir energy. Note that since the matrices do not commute, technically we should verify this by Taylor expansion, which we can do as follows. The nth term (starting with $n = 0$) in the expansion of $(1 - \lambda^2 A)^{-1}$ is $\lambda^{2n} A^n$, so the nth term in the expansion of $(1 - \lambda^2 A)^{-1} \lambda A$ is $\lambda^{2n+1} A^{n+1}$. Integrating over λ gives $A^{n+1}/2(n+1)$. This matches the series expansion of $(-1/2) \log A$. The Taylor expansion of Eq. (14.415) is itself interesting,

$$\begin{aligned} V_{\text{C}} = -\frac{\hbar}{2\pi} \,\text{Im} \int_0^\infty d\omega \,\text{Tr} \Big[& \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \\ & + \frac{1}{2} \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \\ & + \frac{1}{3} \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) + \cdots \Big], \end{aligned} \tag{14.416}$$

since the terms correspond to one-body, two-body, three-body, etc. interactions. As a simple example, consider a collection of $N = 2$ atoms, in which case the two-body term gives the two-atom interaction

$$\begin{aligned} V_{12} &= -\frac{\hbar}{4\pi} \,\text{Im} \int_0^\infty d\omega \,\text{Tr} \Big[\boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \cdot \boldsymbol{\alpha}(\omega) \cdot \mathbf{G}^{(0)}(\omega) \Big] \\ &= -\frac{\hbar}{4\pi} \,\text{Im} \int_0^\infty d\omega \sum_{n,k=1}^{2} \,\text{Tr} \Big[\boldsymbol{\alpha}^{(n)}(\omega) \cdot \mathbf{G}^{(0)}(\mathbf{r}_n, \mathbf{r}_k, \omega) \cdot \boldsymbol{\alpha}^{(k)}(\omega) \cdot \mathbf{G}^{(0)}(\mathbf{r}_k, \mathbf{r}_n, \omega) \Big] \\ &= -\frac{\hbar}{2\pi} \,\text{Im} \int_0^\infty d\omega \,\text{Tr} \Big[\boldsymbol{\alpha}^{(1)}(\omega) \cdot \mathbf{G}^{(0)}(\mathbf{r}_1, \mathbf{r}_2, \omega) \cdot \boldsymbol{\alpha}^{(2)}(\omega) \cdot \mathbf{G}^{(0)}(\mathbf{r}_2, \mathbf{r}_1, \omega) \Big] \end{aligned} \tag{14.417}$$

In the second step, we wrote out the spatial part of the trace, so that the remaining trace refers only to polarization. In the last step, we dropped terms with $n = k$, corresponding to (divergent) one-body interactions. The result gives the Casimir–Polder potential between two atoms; this result is equivalent via Wick rotation to the result (14.257) that we derived directly.

14.4.1.3 Conversion to Permittivity

In going over from a collection of atoms to the continuum limit, we can, for example, take the two-body interaction term from Eq. (14.416) and modify it via

$$\begin{aligned}
&\mathrm{Tr}\big[\boldsymbol{\alpha}(\omega)\cdot\mathbf{G}^{(0)}(\omega)\cdot\boldsymbol{\alpha}(\omega)\cdot\mathbf{G}^{(0)}(\omega)\big] \\
&\qquad\longrightarrow \mathrm{Tr}\int d^3r\int d^3r'\,N(\mathbf{r})\,N(\mathbf{r}')\,\boldsymbol{\alpha}(\mathbf{r},\omega)\cdot\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega)\cdot\boldsymbol{\alpha}(\mathbf{r}',\omega)\cdot\mathbf{G}^{(0)}(\mathbf{r}',\mathbf{r},\omega) \\
&\qquad= \mathrm{Tr}\int d^3r\int d^3r'\,\epsilon_0^2\,\boldsymbol{\chi}(\mathbf{r},\omega)\cdot\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega)\cdot\boldsymbol{\chi}(\mathbf{r}',\omega)\cdot\mathbf{G}^{(0)}(\mathbf{r}',\mathbf{r},\omega) \\
&\qquad= \mathrm{Tr}\big[\epsilon_0\boldsymbol{\chi}(\omega)\cdot\mathbf{G}^{(0)}(\omega)\big]^2,
\end{aligned} \tag{14.418}$$

where note that in the middle two expressions, the overall spatial integration is written out explicitly, while it is buried in the the trace in the other two expressions. Also, the dielectric susceptibility tensor $\boldsymbol{\chi}$ is related to the polarizability tensor [in a tensor generalization of Eq. (1.17)] via

$$\boldsymbol{\chi}(\mathbf{r},\omega) = \frac{N(\mathbf{r})}{\epsilon_0}\boldsymbol{\alpha}(\mathbf{r},\omega), \tag{14.419}$$

where $N(\mathbf{r})$ is the atomic number density, and recall that the susceptibility is related to the dielectric permittivity tensor $\boldsymbol{\epsilon}$ by $\boldsymbol{\chi}(\mathbf{r},\omega) = \boldsymbol{\epsilon}(\mathbf{r},\omega)/\epsilon_0 - 1$. Now performing these same replacements term-by-term in the series expansion of the logarithm, Eq. (14.415) becomes

$$V_{\mathrm{C}} = \frac{\hbar}{2\pi}\,\mathrm{Im}\int_0^\infty d\omega\,\mathrm{Tr}\log\Big[1-\epsilon_0\boldsymbol{\chi}(\omega)\cdot\mathbf{G}^{(0)}(\omega)\Big], \tag{14.420}$$

(dispersive Casimir energy, with free-space Green tensor)

where again the trace in this expression encompasses position integral.

14.4.1.4 Introduction of the Dielectric Green Tensor

Thus far, the expression (14.420) represents the field in terms of the vacuum Green tensor, but more it would be more sensible to have the Green tensor in the presence of the dielectric to represent the field. Recall the defining relation (14.28) for the Green tensor:

$$\nabla\times[\nabla\times\mathbf{G}(\mathbf{r},\mathbf{r}',\omega)] - \omega^2\mu_0\boldsymbol{\epsilon}(\mathbf{r},\omega)\cdot\mathbf{G}(\mathbf{r},\mathbf{r}',\omega) = \mu_0\omega^2\delta^3(\mathbf{r}-\mathbf{r}'). \tag{14.421}$$

Taking the difference between the dielectric and vacuum cases gives

$$\nabla\times\nabla\times\big[\mathbf{G}-\mathbf{G}^{(0)}\big] - \omega^2\mu_0\big[\boldsymbol{\epsilon}(\omega)\cdot\mathbf{G}-\epsilon_0\mathbf{G}^{(0)}\big] = 0. \tag{14.422}$$

Then rearranging and subtracting $\omega^2\mu_0\epsilon_0\mathbf{G}$ from both sides,

$$\nabla\times\nabla\times\big[\mathbf{G}-\mathbf{G}^{(0)}\big] - \omega^2\mu_0\epsilon_0\big[\mathbf{G}-\mathbf{G}^{(0)}\big] = \omega^2\mu_0\big[\boldsymbol{\epsilon}(\omega)-\epsilon_0\big]\cdot\mathbf{G} = \omega^2\mu_0\epsilon_0\boldsymbol{\chi}(\omega)\cdot\mathbf{G}. \tag{14.423}$$

This has the same form as the defining wave equation (14.421), but in vacuum, and with ''source'' $\omega^2\mu_0\epsilon_0\chi(\omega)\mathbf{G}$. The solution is thus $\mathbf{G}^{(0)}$ ''convolved'' with the ''source'' $\epsilon_0\boldsymbol{\chi}(\omega)\cdot\mathbf{G}$:

$$\mathbf{G}-\mathbf{G}^{(0)} = \mathbf{G}^{(0)}\cdot\epsilon_0\boldsymbol{\chi}(\omega)\cdot\mathbf{G}. \tag{14.424}$$

With explicit position dependence, this solution is

$$\mathbf{G}(\mathbf{r},\mathbf{r}',\omega) - \mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}',\omega) = \int d^3r''\,\mathbf{G}^{(0)}(\mathbf{r},\mathbf{r}'',\omega)\cdot\epsilon_0\boldsymbol{\chi}(\mathbf{r}'',\omega)\cdot\mathbf{G}(\mathbf{r}'',\mathbf{r}',\omega). \tag{14.425}$$

Now a transpose gives

$$\mathbf{G} - \mathbf{G}^{(0)} = \mathbf{G} \cdot \epsilon_0 \boldsymbol{\chi}(\omega) \cdot \mathbf{G}^{(0)}, \tag{14.426}$$

while operating with the left-inverse of $\mathbf{G}$ gives

$$1 - \epsilon_0 \boldsymbol{\chi}(\omega) \cdot \mathbf{G}^{(0)} = \mathbf{G}^{-1} \cdot \mathbf{G}^{(0)}. \tag{14.427}$$

Using this result in Eq. (14.420) gives the energy expression

$$V_{\text{C}} = -\frac{\hbar}{2\pi} \operatorname{Im} \int_0^\infty d\omega \operatorname{Tr} \log \left[\left(\mathbf{G}^{(0)} \right)^{-1} \cdot \mathbf{G} \right]$$

(dispersive Casimir energy, trace-log form) (14.428)

after trading an overall minus sign for the inverse of the matrix inside the log. We can also write this expression more simply as

$$V_{\text{C}} = -\frac{\hbar}{2\pi} \operatorname{Im} \int_0^\infty d\omega \operatorname{Tr} \log \mathbf{G}(\omega),$$

(dispersive Casimir energy, trace-log form) (14.429)

with the understanding that we should renormalize the energy (which in this expression is divergent) by subtracting the energy of the vacuum or some other comparable dielectric configuration. Although we have been working in the space of *real* frequencies ω, any of these expressions may be Wick-rotated to imaginary frequency. For example, Eq. (14.429) becomes

$$V_{\text{C}} = -\frac{\hbar}{2\pi} \int_0^\infty ds \operatorname{Tr} \log \mathbf{G}(is),$$

(dispersive Casimir energy, Wick-rotated trace-log form) (14.430)

upon letting $\omega \longrightarrow is$.

14.4.1.5 Explicit Permittivity Dependence

We can arrive at an alternate expression for the Casimir energy that highlights the direct dependence on the dispersive permittivity and allows a direct comparison to the nondispersive expression (14.397). Continuing with the unrenormalized expression (14.429), and keeping in mind the necessity of renormalizing the final result, we can integrate by parts to obtain

$$V_{\text{C}} = \frac{\hbar}{2\pi} \operatorname{Im} \operatorname{Tr} \int_0^\infty d\omega\, \omega \partial_\omega \log \mathbf{G} = \frac{\hbar}{2\pi} \operatorname{Im} \operatorname{Tr} \int_0^\infty d\omega\, \omega\, \mathbf{G}^{-1} \partial_\omega \mathbf{G}. \tag{14.431}$$

To evaluate $\partial_\omega G(\omega)$, we can multiply through by ω^{-2} and then differentiate the wave equation (14.421) to obtain

$$\nabla \times \nabla \times \partial_\omega [\omega^{-2} \mathbf{G}] - \omega^2 \mu_0 \boldsymbol{\epsilon} \cdot \partial_\omega [\omega^{-2} \mathbf{G}] = 2\omega^{-1} \mu_0 \boldsymbol{\epsilon} \cdot \mathbf{G} + \mu_0 (\partial_\omega \boldsymbol{\epsilon}) \cdot \mathbf{G}. \tag{14.432}$$

In analogy with Eq. (14.424), we can write the solution as

$$\partial_\omega \left[\omega^{-2} \mathbf{G} \right] = \mathbf{G} \cdot \omega^{-2} \left[2\omega^{-1} \boldsymbol{\epsilon} + (\partial_\omega \boldsymbol{\epsilon}) \right] \cdot \mathbf{G}, \tag{14.433}$$

or simplifying this,

$$\partial_\omega \mathbf{G} = 2\omega^{-1} \mathbf{G} + \mathbf{G} \cdot \left[2\omega^{-1} \boldsymbol{\epsilon} + (\partial_\omega \boldsymbol{\epsilon}) \right] \cdot \mathbf{G} = 2\omega^{-1} \mathbf{G} + \mathbf{G} \cdot \frac{\partial_\omega (\omega^2 \boldsymbol{\epsilon})}{\omega^2} \cdot \mathbf{G}. \tag{14.434}$$

Putting this into Eq. (14.432), but discarding the first term (which produces a constant term that vanishes under renormalization), we arrive at

$$V_{\text{C}} = \frac{\hbar}{2\pi} \operatorname{Im} \operatorname{Tr} \int_0^\infty \frac{d\omega}{\omega} \left[\partial_\omega (\omega^2 \boldsymbol{\epsilon}) \right] \cdot \mathbf{G}, \tag{14.435}$$

(dispersive Casimir energy)

or writing out the trace in the spatial degrees of freedom,

$$V_{\text{C}} = \frac{\hbar}{2\pi} \operatorname{Im} \operatorname{Tr} \int d^3r \int_0^\infty \frac{d\omega}{\omega} \left[\partial_\omega \left[\omega^2 \boldsymbol{\epsilon}(\mathbf{r}, \omega) \right] \right] \cdot \mathbf{G}(\mathbf{r}, \mathbf{r}, \omega),$$

(dispersive Casimir energy) (14.436)

In the limit of a nondispersive, isotropic dielectric, $[\omega^2 \epsilon]' = 2\omega\epsilon$, and this relation reduces to the first (electric-field) term of the nondispersive expression (14.397), if we also ignore any imaginary part of the permittivity. However, the nondispersive expression (14.397) does not generalize in a completely obvious way to a dispersive medium, because of the presence of a term of the form $\omega\epsilon'(\omega)$.

14.4.1.6 Free-Space Energy

A simple check on the Casimir-energy expression is to examine the energy in free space, by letting $\boldsymbol{\epsilon} \longrightarrow \epsilon_0$ in Eq. (14.436):

$$V_{\text{C}} = \frac{\hbar\epsilon_0}{\pi} \int d^3r \int_0^\infty d\omega \operatorname{Tr} \operatorname{Im} \mathbf{G}^{(0)}(\mathbf{r}, \mathbf{r}, \omega). \tag{14.437}$$

Using the free-space Green tensor (14.46),

$$G^{(0)}_{\alpha\beta}(\mathbf{r}, 0, \omega) = \frac{1}{4\pi\epsilon_0} \left\{ [3\hat{r}_\alpha \hat{r}_\beta - \delta_{\alpha\beta}] \left[\frac{1}{r^3} - i\frac{k}{r^2} \right] - [\hat{r}_\alpha \hat{r}_\beta - \delta_{\alpha\beta}] \frac{k^2}{r} \right\} e^{ikr}, \tag{14.438}$$

we can discard the first portion in the trace and keep only the imaginary part to find

$$\operatorname{Tr} \operatorname{Im} \mathbf{G}^{(0)}(\mathbf{r}, 0, \omega) = \frac{1}{2\pi\epsilon_0} \left\{ \frac{k^2}{r} \right\} \sin kr. \tag{14.439}$$

Then at the same source and effect points, we have

$$\operatorname{Tr} \operatorname{Im} \mathbf{G}^{(0)}(\mathbf{r}, \mathbf{r}, \omega) = \frac{k^3}{2\pi\epsilon_0} = \frac{\omega^3}{2\pi\epsilon_0 c^3}. \tag{14.440}$$

In Eq. (14.437), this leads to the free-space energy

$$V_{\text{C}} = \int d^3r \int_0^\infty d\omega \, \frac{\hbar\omega^3}{2\pi^2 c^3} = \int d^3r \int_0^\infty d\omega \left(\frac{1}{2}\hbar\omega \right) 2 \left(\frac{\omega^2}{2\pi^2 c^3} \right), \tag{14.441}$$

where we can interpret the last expression as counting zero-point energies $\hbar\omega/2$ for each mode, with 2 polarizations per frequency and a vacuum mode density $\omega^2/2\pi^2c^3$. We can verify this energy by direct mode summation in a large quantization volume V, in analogy with the steps in the calculation of the spontaneous-emission rate in Section 11.3:

$$\begin{aligned} V_0 &:= \sum_{\mathbf{k},\zeta} \frac{\hbar\omega_k}{2} \\ &= \hbar \sum_{\text{k}} \omega_k \\ &\longrightarrow \frac{\hbar c V}{8\pi^3} \int d^3k \, k \\ &= \frac{\hbar c V}{2\pi^2} \int dk \, k^3 \\ &= \frac{\hbar V}{2\pi^2 c^3} \int d\omega \, \omega^3. \end{aligned} \tag{14.442}$$

The vacuum mode density here also agrees with the density of states from the Fermi-Golden-rule calculation of the spontaneous-emission rate, where Eq. (11.66) gives $\rho(E)\,dE = E^2V\,dE/\pi^2\hbar^3c^3$, and letting $E = \hbar\omega$ gives $\rho(\omega)\,d\omega = \omega^2V\,d\omega/\pi^2c^3$, which already counts both polarization.

14.4.1.7 Reduction to the Casimir–Polder Potential

Now to make contact again with the Casimir–Polder potential, consider a small variation $\delta\boldsymbol{\epsilon}$ in the permittivity ϵ. Then the variation in the Casimir energy (14.435) is

$$\delta V_{\text{C}} = \frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int_0^\infty \frac{d\omega}{\omega}\left\{\Big[\partial_\omega(\omega^2\delta\boldsymbol{\epsilon})\Big]\cdot\mathbf{G} + \Big[\partial_\omega(\omega^2\boldsymbol{\epsilon})\Big]\cdot\delta\mathbf{G}\right\}. \tag{14.443}$$

Note that it is critical here to consider the variation $\delta\mathbf{G}$ in the Green tensor, which depends implicitly on the permittivity, as well as the direct variation $\delta\boldsymbol{\epsilon}$. The same variation in the wave equation (14.421) gives

$$\nabla\times\nabla\times\delta\mathbf{G} - \omega^2\mu_0\boldsymbol{\epsilon}\cdot\delta\mathbf{G} = \omega^2\mu_0\delta\boldsymbol{\epsilon}\cdot\mathbf{G}, \tag{14.444}$$

which has solution

$$\delta\mathbf{G} = \mathbf{G}\cdot\delta\boldsymbol{\epsilon}\cdot\mathbf{G}. \tag{14.445}$$

We can use this result to eliminate $\delta\mathbf{G}$ in the integrand of Eq. (14.435):

$$\Big[\partial_\omega(\omega^2\delta\boldsymbol{\epsilon})\Big]\cdot\mathbf{G} + \Big[\partial_\omega(\omega^2\boldsymbol{\epsilon})\Big]\cdot\delta\mathbf{G} = \Big[\partial_\omega(\omega^2\delta\boldsymbol{\epsilon})\Big]\cdot\mathbf{G} + \Big[\partial_\omega(\omega^2\boldsymbol{\epsilon})\Big]\cdot\mathbf{G}\cdot\delta\boldsymbol{\epsilon}\cdot\mathbf{G}. \tag{14.446}$$

Then using Eq. (14.434) in the form

$$\mathbf{G}^{-1}\cdot\omega^2\partial_\omega\mathbf{G} = 2\omega + \partial_\omega(\omega^2\boldsymbol{\epsilon})\cdot\mathbf{G}, \tag{14.447}$$

we can rewrite the last term of Eq. (14.446), with the result

$$\begin{aligned}\Big[\partial_\omega(\omega^2\delta\boldsymbol{\epsilon})\Big]\cdot\mathbf{G} + \Big[\partial_\omega(\omega^2\boldsymbol{\epsilon})\Big]\cdot\delta\mathbf{G} &= \Big[\partial_\omega(\omega^2\delta\boldsymbol{\epsilon})\Big]\cdot\mathbf{G} + \mathbf{G}^{-1}\cdot\omega^2\partial_\omega\mathbf{G}\cdot\delta\boldsymbol{\epsilon}\cdot\mathbf{G} - 2\omega\delta\boldsymbol{\epsilon}\cdot\mathbf{G}\\ &= \Big[\omega^2\partial_\omega\delta\boldsymbol{\epsilon}\Big]\cdot\mathbf{G} + \mathbf{G}^{-1}\cdot\omega^2\partial_\omega\mathbf{G}\cdot\delta\boldsymbol{\epsilon}\cdot\mathbf{G}\end{aligned} \tag{14.448}$$

Then Eq. (14.435) becomes

$$\delta V_{\text{C}} = \frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int_0^\infty \frac{d\omega}{\omega}\left\{\omega^2\partial_\omega\delta\boldsymbol{\epsilon}\cdot\mathbf{G} + \omega^2\delta\boldsymbol{\epsilon}\cdot\partial_\omega\mathbf{G}\right\} \tag{14.449}$$

after cyclic permutation of matrices under the trace. Simplifying further gives

$$\delta V_{\text{C}} = \frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int_0^\infty d\omega\,\omega\,\partial_\omega\big(\delta\boldsymbol{\epsilon}\cdot\mathbf{G}\big), \tag{14.450}$$

and then integration by parts gives

$$\delta V_{\text{C}} = -\frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int_0^\infty d\omega\,\delta\boldsymbol{\epsilon}\cdot\mathbf{G} = -\frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int d^3r\int_0^\infty d\omega\,\delta\boldsymbol{\epsilon}(\mathbf{r},\omega)\cdot\mathbf{G}(\mathbf{r},\mathbf{r},\omega), \tag{14.451}$$

(Casimir-energy variation)

where we have written out the spatial part of the trace in the second expression. Now using Eq. (14.419) to identify the permittivity variation with a single localized atom at $\mathbf{r}_{\text{A}}$,

$$\delta\boldsymbol{\epsilon}(\mathbf{r},\omega) = \frac{N(\mathbf{r})}{\epsilon_0}\boldsymbol{\alpha}(\mathbf{r},\omega) = \frac{\delta(\mathbf{r}-\mathbf{r}_{\text{A}})}{\epsilon_0}\boldsymbol{\alpha}(\omega), \tag{14.452}$$

the delta function removes the spatial integral, and we are left with

$$V_{\text{CP}}(\mathbf{r}_{\text{A}}) \equiv \delta V_{\text{C}} = -\frac{\hbar}{2\pi}\,\text{Im}\,\text{Tr}\int_0^\infty d\omega\,\boldsymbol{\alpha}(\omega)\cdot\mathbf{G}(\mathbf{r},\mathbf{r},\omega), \tag{14.453}$$

in agreement with Eq. (14.399), if we renormalize by subtracting the same expression with the vacuum Green tensor. Note that the variation expression (14.451) is also useful in other situations with a small change in the dielectric, such as computing the *force* between two planar dielectric slabs (where the idea is to compute the difference between two slightly different distances, so $\delta\epsilon$ corresponds to a thin film on one of the surfaces).

14.4.1.8 Temperature Dependence

In generalizing these expressions for the Casimir energy to the case of nonzero temperature, we can appeal to the following expression,

$$V_{\text{CP}} = -\frac{\hbar}{2\pi}\,\text{Im}\int_0^\infty d\omega\,\text{Tr}\big[\boldsymbol{\alpha}(\omega)\cdot\mathbf{G}(\mathbf{r},\mathbf{r},\omega)\big]\,\coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right], \tag{14.454}$$

which follows from the temperature-dependent Casimir–Polder expression(14.327), exploiting the odd nature of the integrand (specifically, the imaginary parts of the response functions). Note that this is precisely the same as the starting expression (14.399) for the Casimir–Polder energy at zero temperature, except for the presence of the coth factor here. Thus, all the results here carry through by judiciously inserting this factor. For example, the trace-log expression (14.428) becomes[43]

$$V_{\text{C}} = -\frac{\hbar}{2\pi}\,\text{Im}\int_0^\infty d\omega\,\text{Tr}\log\left[\left(\mathbf{G}^{(0)}\right)^{-1}\cdot\mathbf{G}\right]\coth\left[\frac{\hbar\omega}{2k_{\text{B}}T}\right].$$

(dispersive Casimir energy, trace-log form) (14.455)

The same contour integration that we used in the Casimir–Polder case applies here, leading to the Matsubara frequencies appearing due to the poles in the coth function.

[43] Compare to, for exmaple, Francesco Intravaia and Ryan Behunin, ''Casimir effect as a sum over modes in dissipative systems,'' *Physical Review A* **86**, 062517 (2012), Eq. (42) (doi: 10.1103/PhysRevA.86.062517).

14.5 Exercises

Problem 14.1

Suppose we denote the permittivity including conduction by

$$\tilde{\epsilon}(\omega) := \epsilon(\omega) + \frac{i\sigma(\omega)}{\omega}, \tag{14.456}$$

where $\epsilon(\omega)$ is analytic in the upper half-plane ($\mathrm{Im}[\omega] \geq 0$) and obeys the usual Kramers–Kronig relations (14.90)

$$\begin{aligned} \mathrm{Re}[\epsilon(\omega) - \epsilon_0] &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Im}[\epsilon(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega' \\ \mathrm{Im}[\epsilon(\omega) - \epsilon_0] &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Re}[\epsilon(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega'. \end{aligned} \tag{14.457}$$

Show that $\tilde{\epsilon}(\omega)$ satisfies the modified Kramers–Kronig relations

$$\begin{aligned} \mathrm{Re}[\tilde{\epsilon}(\omega) - \epsilon_0] &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Im}[\tilde{\epsilon}(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega' \\ \mathrm{Im}[\tilde{\epsilon}(\omega) - \epsilon_0] &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Re}[\tilde{\epsilon}(\omega') - \epsilon_0]}{\omega' - \omega}\, d\omega' + \frac{\sigma_0}{\omega}, \end{aligned} \tag{14.458}$$

where $\sigma_0 = \sigma(0)$ is the dc conductivity, which is a real number. You may also assume $\sigma(\omega) \longrightarrow 0$ as $\omega \longrightarrow \infty$.

Hint: the algebra is simple if you keep the relations as much as possible in terms of the Hilbert-transform operator $\mathscr{H}$.

Problem 14.2

Show that the Kramers–Kronig relations for the permittivity

$$\begin{aligned} \mathrm{Re}[\epsilon(\omega)] &= \epsilon_0 + \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Im}[\epsilon(\omega')]}{\omega' - \omega}\, d\omega' \\ \mathrm{Im}[\epsilon(\omega)] &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\mathrm{Re}[\epsilon(\omega')] - \epsilon_0}{\omega' - \omega}\, d\omega' \end{aligned} \tag{14.459}$$

can be written in the equivalent form

$$\begin{aligned} \mathrm{Re}[\epsilon(\omega)] &= \epsilon_0 + \frac{2}{\pi} \int_{0}^{\infty} \frac{\omega' \mathrm{Im}[\epsilon(\omega')]}{\omega'^2 - \omega^2}\, d\omega' \\ \mathrm{Im}[\epsilon(\omega)] &= -\frac{2\omega}{\pi} \int_{0}^{\infty} \frac{\mathrm{Re}[\epsilon(\omega')] - \epsilon_0}{\omega'^2 - \omega^2}\, d\omega'. \end{aligned} \tag{14.460}$$

Hint: use the property $\epsilon(-\omega^*) = \epsilon^*(\omega)$.

Problem 14.3

Consider the (classical) forced, damped harmonic oscillator

$$\ddot{x} + \gamma \dot{x} + \omega_0^2 x = f(t), \tag{14.461}$$

subject to the forcing function $f(t)$.

(a) Find the Green function $g(t,t')$ for the initially undisturbed system, defined as the solution $x(t)$ where the forcing function $f(t) = \delta(t-t')$ is an impulse at time t', and with $x(t) = 0$ for $t < t'$. That is, $g(t,t')$ satisfies

$$\ddot{g} + \gamma\dot{g} + \omega_0^2 g = \delta(t-t'), \tag{14.462}$$

subject to the boundary condition $g(t',t') = 0$. Assume underdamped oscillation, and don't bother to derive the solution to Eq. (14.461) (look it up from a reputable source).

(b) Write down the general solution to Eq. (14.461) [by integrating Eq. (14.462) over t'] for an arbitrary forcing function $f(t)$ as an integral involving the Green function. Noting that $g(t,t') = g(t-t')$, show that the integral is in fact a convolution of $f(t)$ with $g(t)$.

(c) Derive the frequency-space Green function $\tilde{g}(\omega)$, defined as the amplitude $\tilde{x}(\omega)$, where

$$x(t) = \tilde{x}(\omega)e^{-i\omega t}, \tag{14.463}$$

and $x(t)$ is the solution to Eq. (14.461) due to a unit-amplitude, monochromatic forcing at frequency ω,

$$f(t) = e^{-i\omega t}. \tag{14.464}$$

(d) For an arbitrary forcing function $f(t)$ with Fourier transform $\tilde{f}(\omega)$, use the convolution theorem to show that the solution $x(t)$ may be written

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \tilde{g}(\omega)\tilde{f}(\omega)e^{-i\omega t}. \tag{14.465}$$

Thus, $\tilde{g}(\omega)$ is the **transfer function** for the damped harmonic oscillator, because it gives the "transfer efficiency" for the forcing amplitude at frequency ω through the system. Of course $\tilde{g}(\omega)$ is also the generalized susceptibility for the damped harmonic oscillator, and thus, for example, obeys the Kramers–Kronig relations.

Problem 14.4

In Section 14.1.4, we argued that $g_\chi(t \longrightarrow \infty) \longrightarrow 0$, where $g_\chi(t)$ is the inverse Fourier transform of the susceptibility $\chi(\omega)$,

$$g_\chi(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \chi(\omega)\, e^{-i\omega t}, \tag{14.466}$$

was a reasonable requirement, since a dielectric medium "forgets" any electromagnetic perturbations in the distant past. However, this conclusion is *not* valid for a conductor.

(a) For a conductor with permittivity

$$\tilde{\epsilon}(\mathbf{r},\omega) = \left[\epsilon(\mathbf{r},\omega) + i\frac{\sigma(\mathbf{r},\omega)}{\omega}\right], \tag{14.467}$$

take $\chi(\omega) = \tilde{\epsilon}(\mathbf{r},\omega)/\epsilon_0 - 1$ to be the susceptibility. Then show that $\lim_{t\to\infty} g_\chi(t) = \sigma_0/\epsilon_0$, where $\sigma_0 = \sigma(0)$.

Note: this is not a technically difficult problem, but it is rather tricky for two reasons. First, when you set up the semicircular contour integral, the contour you choose will depend on the sign of t, in order to get the curved part to go away. Second, the $\chi(\omega)$ is actually not *quite* correct at $\omega = 0$; go ahead and compute $g_\chi(t)$ from $\chi(\omega)$, but then adjust the $\omega = 0$ component of $\chi(\omega)$ to obtain a result consistent with a causal response.

(b) Since $g_\chi(t \longrightarrow \infty) \longrightarrow 0$ fails to hold for a conductor, this means that a conductor *doesn't* forget perturbations in the distant past. Give a physical interpretation of this statement by considering a (near-dc) pulsed electric field applied to a medium (assume that the sign of the electric field is always positive). What is the net effect of the pulse long after it finishes for a dielectric vs. a conductor?

Problem 14.5

Prove the integral formula (14.358)

$$\int_0^\infty dx\,\frac{ab}{(a^2+x^2)(b^2+x^2)} = \frac{\pi}{2(a^2-b^2)}(a\,\mathrm{sgn}\,b - b\,\mathrm{sgn}\,a) \qquad (a,b\in\mathbb{R},\ a,b\neq 0) \tag{14.468}$$

by considering the contour around the great upper half-plane.

Problem 14.6

Justify the formula

$$\lim_{\epsilon\to 0^+}\frac{1}{x\pm i\epsilon} = \mathrm{P}\,\frac{1}{x} \mp i\pi\delta(x). \tag{14.469}$$

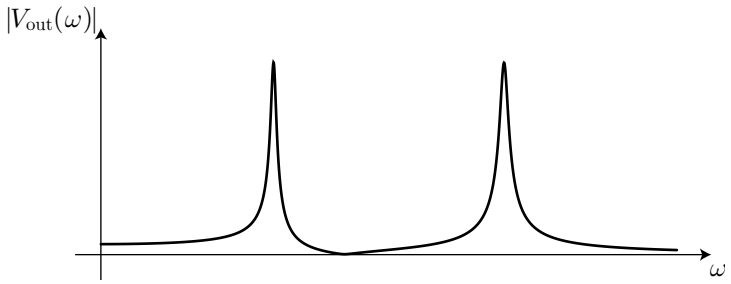
Here, the "P" denotes the Cauchy principal value, i.e.,

$$\int_{-\infty}^{\infty} dx\,\mathrm{P}\,f(x) \equiv \mathrm{P}\int_{-\infty}^{\infty} dx\,f(x) \equiv \fint_{-\infty}^{\infty} dx\,f(x). \tag{14.470}$$

Note: the delta function is defined in terms of its action on a test function $f(x)$ under an integral, i.e. $\int dx\,f(x)\,\delta(x) = f(0)$. Show that the formula (14.469) holds in the same sense. Are there restrictions on the test functions $f(x)$ for which this formula is valid?

Problem 14.7

A linear circuit has the following frequency response for the output voltage amplitude $|V_{\mathrm{out}}(\omega)|$ as a function of the constant input-voltage amplitude $V_{\mathrm{in}}(\omega)$. Sketch the corresponding phase lag $\delta(\omega)$ of the output voltage, given by $V_{\mathrm{out}} = |g(\omega)|V_{\mathrm{in}}(\omega)\,e^{i\delta(\omega)}$ for some (complex) gain function $g(\omega)$. *Explain.*



Problem 14.8

Understand and use the expression (14.210)

$$V_{\mathrm{CP}} = -\frac{3\hbar c\alpha_0}{64\pi^2\epsilon_0 z^4}\int_1^\infty \frac{d\xi}{\xi^4}\left[\frac{\sqrt{\chi+\xi^2}-\xi}{\sqrt{\chi+\xi^2}+\xi} + \left(1-2\xi^2\right)\frac{\sqrt{\chi+\xi^2}-\xi(1+\chi)}{\sqrt{\chi+\xi^2}+\xi(1+\chi)}\right] \tag{14.471}$$

for the Casimir–Polder energy, and adapt it to the magnetic-field case as follows. First, replace the dc electric polarizability with the analogous dc magnetizability $\alpha_{\mathrm{m}0}$, defined such that the induced magnetic-dipole moment is $\mu = \alpha_{\mathrm{m}0}B$, making sure to keep the dimensions correct. Then use the considerations of the dielectric-mode-sum analysis of Section 13.13.4 to modify the integrals for the magnetic field. To keep things simple, consider only the limits of very small and very large χ.

Problem 14.9

Beginning with the expression (14.217)

$$V_{\text{CP}} = \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}+\alpha_{zz}(is)\right)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}\left[\left(r_\perp+\left(1+\frac{2k_\text{T}^2c^2}{s^2}\right)r_\parallel\right)\right]e^{-2\kappa z} + \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}-\alpha_{zz}(is)\right)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}(r_\perp+r_\parallel)e^{-2\kappa z}. \tag{14.472}$$

for the Casimir–Polder force between an atom and a planar dielectric interface, take the large-distance (long-wavelength) limit. Show that this may be written

$$V_{\text{CP}} = -\frac{3\hbar c}{64\pi^2\epsilon_0 z^4}\left[\left(\frac{\alpha_{0,xx}+\alpha_{0,yy}}{2}+\alpha_{0,zz}\right)\eta(\chi)+\left(\frac{\alpha_{0,xx}+\alpha_{0,yy}}{2}-\alpha_{0,zz}\right)\eta^{(a)}(\chi)\right], \tag{14.473}$$

where the $\alpha_{0,\beta\beta}$ are diagonal components of the static polarizability tensor, and the functions $\eta(\chi)$ and $\eta^{(a)}(\chi)$ are χ-dependent functions that are zero in the limit $\chi \longrightarrow 0$, with $\lim_{\chi\to\infty}\eta(\chi) = 1$ and $\lim_{\chi\to\infty}\eta^{(a)}(\chi) = 1/2$.

Problem 14.10

Beginning with the same expression as in Problem 14.9 [Eq. (14.472)],

$$V_{\text{CP}} = \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}+\alpha_{zz}(is)\right)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}\left[\left(r_\perp+\left(1+\frac{2k_\text{T}^2c^2}{s^2}\right)r_\parallel\right)\right]e^{-2\kappa z} + \frac{\hbar}{16\pi^2\epsilon_0 c^2}\int_0^\infty ds\, s^2\left(\frac{[\alpha_{xx}(is)+\alpha_{yy}(is)]}{2}-\alpha_{zz}(is)\right)\int_0^\infty dk_\text{T}\,\frac{k_\text{T}}{\kappa}(r_\perp+r_\parallel)e^{-2\kappa z}, \tag{14.474}$$

compute the Casimir–Polder potential in the near-field approximation.

Chapter 15

Resolvent Operator

We will now develop a method of calculation based on the resolvent operator.[1] This method is quite formal but also quite powerful, especially in scattering problems and problems involving transitions to continua.

15.1 Definition

The **resolvent operator** is defined in the complex plane in terms of the *time-independent* Hamiltonian H by

$$G(z) := \frac{1}{z - H}. \tag{15.1}$$
(resolvent operator)

Expanding into a basis $|\alpha\rangle$ of eigenstates of H (i.e., multiplying by the identity in this basis on either side of the resolvent), we see that

$$G(z) = \sum_{\alpha} \frac{|\alpha\rangle\langle\alpha|}{z - E_\alpha}. \tag{15.2}$$

Thus, we see that $G(z)$ has a simple pole at each eigenvalue E_α of H. In the case of a continuum of states, where the eigenvalues cluster together into a continuous interval along the real line, $G(z)$ has instead a branch cut. For example, for a bound system with ionized states, $G(z)$ has poles at each bound-state energy, but a branch cut beginning at the smallest ionization energy and extending to $+\infty$ along the real axis. There are some subtleties in dealing with branch cuts in $G(z)$ that we will return to later. However, away from the real axis, the resolvent operator is analytic (in the sense that all its matrix elements are analytic functions off the real axis).

15.2 Green Functions for the Schrödinger Equation

15.2.1 Energy-Space Green Functions

Consider the following function, defined by the integral expression

$$G^+(E) := -\lim_{\delta \to 0^+} \frac{i}{\hbar} \int_0^\infty d\tau \, e^{i(E-H)\tau/\hbar} e^{-\tau\delta/\hbar}, \tag{15.3}$$

[1] The oft-cited canonical references are Marvin L. Goldberger and Kenneth M. Watson, *Collision Theory* (Wiley, 1964); and Albert Messiah, *Quantum Mechanics* (Wiley, 1958) (ISBN: 0486409244). However, a detailed description with many examples appears in Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Atom–Photon Interactions: Basic Processes and Applications* (Wiley, 1992), Chapter III (ISBN: 0471625566). Also, in the context of spontaneous emission, see the section on the ''Goldberger–Watson method'' in G. S. Agarwal, ''Quantum Statistical Theories of Spontaneous Emission and their Relation to Other Approaches,'' *Springer Tracts Modern Physics* **70**, 1 (1974) (doi: 10.1007/BFb0042382). Another nice, more general introduction is given by Doron Cohen, ''Lecture Notes in Quantum Mechanics,'' arXiv.org preprint (arXiv: quant-ph/0605180v3) (2008). For the resolvent method applied to the master equation, see P. Lambropoulos, ''Spectral Line Shape in the Presence of Weak Collisions and Intense Fields,'' *Physical Review* **164**, 84 (1967) (doi: 10.1103/PhysRev.164.84).

where the δ factor is inserted to guarantee convergence of the integral. Carrying out this integral,

$$\begin{aligned} G^+(E) &= -\lim_{\delta\to 0^+} \frac{i}{\hbar}\int_0^\infty d\tau\, e^{i(E-H+i\delta)\tau/\hbar} \\ &= -\lim_{\delta\to 0^+} \frac{e^{i(E-H+i\delta)\tau/\hbar}}{E-H+i\delta}\Big|_0^\infty \\ &= \lim_{\delta\to 0^+} \frac{1}{E-H+i\delta} \\ &= \frac{1}{E-H+i0^+} \\ &= G(E+i0^+). \end{aligned} \tag{15.4}$$

Similarly, we can define the function

$$G^-(E) := \lim_{\delta\to 0^+} \frac{i}{\hbar}\int_{-\infty}^0 d\tau\, e^{i(E-H)\tau/\hbar} e^{+\delta\tau/\hbar}, \tag{15.5}$$

which becomes

$$\begin{aligned} G^-(E) &= \lim_{\delta\to 0^+} \frac{i}{\hbar}\int_{-\infty}^0 d\tau\, e^{i(E-H-i\delta)\tau/\hbar} \\ &= \frac{1}{E-H-i0^+} \\ &= G(E-i0^+). \end{aligned} \tag{15.6}$$

For reasons we will see, $G^+(E)$ is called the **retarded Green function**, in energy (frequency) space, while $G^-(E)$ is called the **advanced Green function** in energy space. We have thus shown that both Green functions are related to the resolvent via

$$G^\pm(E) = G(E\pm i0^+) = \frac{1}{E-H\pm i0^+}.$$

(retarded and advanced Green functions) (15.7)

That is, they are essentially the resolvent along the line displaced infinitesimally above and below the real axis. These differ due to the singular nature of the resolvent along the real axis.

15.2.2 Time-Dependent Green Functions and Propagators

Now note that the definition (15.3) can be rewritten

$$G^+(E) = \frac{1}{i\hbar}\int_{-\infty}^\infty d\tau\, e^{i(E+i0^+)\tau/\hbar}\, U(\tau,0)\,\Theta(\tau), \tag{15.8}$$

where $\Theta(\tau)$ is the Heaviside step function and $U(\tau,0) = e^{-iH\tau/\hbar}$ is the unitary time-evolution operator from time 0 to τ for evolution under the time-independent Hamiltonian H. That is, $G^+(E)$ is (within a specific normalization convention) the Fourier transform of "half" of the time-evolution operator, $U(\tau,0)\Theta(\tau)$. Similarly, from the definition (15.5), we see that

$$G^-(E) = \frac{1}{i\hbar}\int_{-\infty}^\infty d\tau\, e^{i(E-i0^+)\tau/\hbar}\,[-U(\tau,0)\,\Theta(-\tau)], \tag{15.9}$$

so that up to a minus sign, $G_-(E)$ is the Fourier transform of the "other half" of the time-evolution operator, $U(\tau,0)\,\Theta(-\tau)$. Thus, defining the time-dependent Green functions

$$G^\pm(t,t_0) := \pm U(t,t_0)\,\Theta[\pm(t-t_0)],$$

(retarded and advanced Green functions) (15.10)

these are related by the above Green functions by a Fourier transform:

$$G^{\pm}(E) = \frac{1}{i\hbar}\int_{-\infty}^{\infty} d\tau\, e^{i[E+(\mathrm{sgn}\tau)i0^+]\tau/\hbar}\, G^{\pm}(\tau,0).$$

(Green-function Fourier transform) (15.11)

Since the energy-space Green function $G^{\pm}(E)$ is the Fourier transform of the time-domain function $G^{\pm}(\tau,0)$, the former is commonly written $\tilde{G}^{\pm}(E)$. However, since we will usually be in the energy domain, here we will just use the argument of the Green function to determine in which space it lives. Note that since we are dealing with time-independent systems, $G^{\pm}(t,t_0)$ only depends on $t-t_0$.

Why the terminology of advanced and retarded Green functions? First, recall [Eq. (4.36)] that the time-evolution operator satisfies the Schrödinger equation,

$$i\hbar\partial_t U(t,t_0) = HU(t,t_0). \tag{15.12}$$

We can use this relation and $\partial_\tau\Theta(\tau)=\delta(\tau)$ to differentiate the Green functions:

$$\begin{aligned} i\hbar\partial_t G^{\pm}(t,t_0) &= \pm i\hbar\partial_t\big\{U(t,t_0)\Theta[\pm(t-t_0)]\big\} \\ &= \pm HU(t,t_0)\Theta[\pm(t-t_0)] \pm i\hbar U(t,t_0)\big[\pm\delta(t-t_0)\big] \\ &= HG^{\pm}(t,t_0) + i\hbar\delta(t-t_0). \end{aligned} \tag{15.13}$$

In the last step, we used $U(t_0,t_0)=1$. Thus, we have shown that

$$(i\hbar\partial_t - H)G^{\pm}(t,t_0) = i\hbar\delta(t-t_0).$$

(Green functions for the Schrödinger equation) (15.14)

Thus, $G^{\pm}(t,t_0)$ is the solution to the Schrödinger equation, "driven" by a delta-function impulse at $t=t_0$. In particular, $G^{+}(t,t_0)$ is the "retarded" Green function, because the "source" is in the past, and the response follows after the impulse, $t>t_0$. Similarly, $G^{-}(t,t_0)$ is the "advanced" Green function, because the source is in the future, and the response comes *before* the impulse, $t<t_0$. Both Green functions obey the same equation, but correspond to different boundary conditions.

Inverting the Fourier-transform relation (15.11) gives

$$G^{\pm}(\tau,0) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^{\pm}(E).$$

(Green-function inverse Fourier transform) (15.15)

In particular, the case of $G^{+}(\tau,0)$ is important, as it gives the time-evolution operator for evolving the system forward from $t=0$ to τ:

$$U(\tau,0) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^{+}(E) \qquad (\tau>0).$$

(Green-function Fourier relation) (15.16)

This relation is particularly useful in that it shows that matrix elements of the evolution operator,

$$K(\beta,t;\alpha,t_0) := \langle\beta|U(t,t_0)|\alpha\rangle, \tag{15.17}$$

(propagator)

collectively called the **propagator**, can be computed from matrix elements of the retarded Green function. The propagator gives the transition amplitude, or the probability amplitude for the system to be in state β at time t, given that it was in state α at the (earlier) time t_0. Note that in the case of forward propagation $t>t_0$, the propagator can also be regarded as the collection of matrix elements of of the retarded Green function $G^{+}(t,t_0)$ in view of the definition (15.10).

15.2.3 Relation to Laplace Transform

We have seen what amounts to this formalism before, when solving the optical Bloch equations via the Laplace-transform method. Recall (Section 5.5.2.1) that the Laplace transform of a time derivative is

$$\mathscr{L}[\dot{y}(t)] = s\mathscr{L}[y(t)] - y(0), \tag{15.18}$$

where the Laplace transform is defined in general by

$$\mathscr{L}[y](s) := \int_0^\infty dt\, e^{-st}\, y(t). \tag{15.19}$$

The Laplace transform of $\partial_t U(t, t_0)$ is then

$$\mathscr{L}[\partial_t U(t, t_0)] = s\mathscr{L}[U(t, t_0)] - U(t_0, t_0) = s\mathscr{L}[U(t, t_0)] - 1. \tag{15.20}$$

Then the Laplace transform of Eq. (15.12) reads

$$i\hbar s\mathscr{L}[U(t, t_0)] - i\hbar = H\mathscr{L}[U(t, t_0)], \tag{15.21}$$

or

$$(i\hbar s - H)\mathscr{L}[U(t, t_0)] = i\hbar, \tag{15.22}$$

so that the Laplace transform of the evolution operator becomes

$$\frac{1}{i\hbar}\mathscr{L}[U(t, t_0)] = \frac{1}{i\hbar s - H}. \tag{15.23}$$

Comparing this to the definition (15.1) of the resolvent, we can identify $z = i\hbar s$ as the rescaled coordinate, and $G(z)$ is proportional to the Laplace transform of the evolution operator, but with a rescaled coordinate, rotated along the imaginary axis. The propagator relation (15.16) is essentially the inverse Laplace transform for $\tau > 0$, representing a convenient method to algebraically solve the initial-value problem for the propagator.

15.3 Transitions Between Discrete States

Consider a quantum system described by Hamiltonian

$$H = H_0 + V, \tag{15.24}$$

where H_0 is the unperturbed Hamiltonian, with interaction V causing transitions among the eigenstates of H_0. We then have two resolvents, one corresponding to the perturbed Hamiltonian and given by the original definition (15.1), and one corresponding to the unperturbed system,

$$G_0(z) := \frac{1}{z - H_0}. \tag{15.25}$$
(unperturbed resolvent operator)

It is convenient to relate these two operators. Starting with $B = A + (B - A)$ for arbitrary A and B, we multiply on the left by B^{-1} and on the right by A^{-1} to obtain the identity

$$\frac{1}{A} = \frac{1}{B} + \frac{1}{B}(B - A)\frac{1}{A}. \tag{15.26}$$

Letting $A = z - H_0 - V$ and $B = z - H_0$, we find

$$G(z) = G_0(z) + G_0(z)VG(z). \tag{15.27}$$
(perturbed resolvent)

This relation may be used directly, or it may be iterated to obtain a perturbation series in terms of the unperturbed resolvent:

$$G = G_0 + G_0VG_0 + G_0VG_0VG_0 + G_0VG_0VG_0VG_0 + \cdots . \tag{15.28}$$
(perturbed resolvent)

In the nonperturbative case, we may take matrix elements of the relation (15.27) in the basis of eigenstates of H_0. The diagonal matrix elements become

$$\begin{aligned}\langle\alpha|G(z)|\alpha\rangle &= \langle\alpha|G_0(z)|\alpha\rangle + \langle\alpha|G_0(z)VG(z)|\alpha\rangle \\ &= \frac{1}{z-E_\alpha} + \frac{1}{z-E_\alpha}\sum_j \langle\alpha|V|j\rangle\langle j|G(z)|\alpha\rangle,\end{aligned} \tag{15.29}$$

while the off-diagonal elements become

$$\begin{aligned}\langle\beta|G(z)|\alpha\rangle &= \langle\beta|G_0(z)|\alpha\rangle + \langle\beta|G_0(z)VG(z)|\alpha\rangle \\ &= \frac{1}{z-E_\beta}\sum_j \langle\beta|V|j\rangle\langle j|G(z)|\alpha\rangle,\end{aligned} \tag{15.30}$$

for $\beta \neq \alpha$. Writing these more compactly,

$$\begin{aligned}(z-E_\alpha)G_{\alpha\alpha}(z) &= 1 + \sum_j V_{\alpha j}G_{j\alpha}(z) \\ (z-E_\beta)G_{\beta\alpha}(z) &= \sum_j V_{\beta j}G_{j\alpha}(z) \qquad (\beta \neq \alpha).\end{aligned} \tag{15.31}$$
(resolvent matrix elements)

The strategy then is to solve the algebraic equations, and then compute the inverse Fourier transform for the resolvent matrix elements to obtain the propagator.

15.3.1 Example: Rabi Oscillations

As an example, we revisit the Rabi-flopping problem in the two-level atom from Section 5.2.2. In the rotating frame of the laser field, the free atomic Hamiltonian is

$$H_0 = -\hbar\Delta|\mathrm{e}\rangle\langle\mathrm{e}|, \tag{15.32}$$

while the laser-driven interaction is

$$V = \frac{\hbar\Omega}{2}\left(\sigma + \sigma^\dagger\right) = \frac{\hbar\Omega}{2}\Big(|\mathrm{g}\rangle\langle\mathrm{e}| + |\mathrm{e}\rangle\langle\mathrm{g}|\Big). \tag{15.33}$$

Then the relations Eqs. (15.31) for the resolvent matrix elements become

$$\begin{aligned}(z+\hbar\Delta)G_{\mathrm{ee}} &= 1 + \frac{\hbar\Omega}{2}G_{\mathrm{ge}} \\ zG_{\mathrm{gg}} &= 1 + \frac{\hbar\Omega}{2}G_{\mathrm{eg}} \\ (z+\hbar\Delta)G_{\mathrm{eg}} &= \frac{\hbar\Omega}{2}G_{\mathrm{gg}} \\ zG_{\mathrm{ge}} &= \frac{\hbar\Omega}{2}G_{\mathrm{ee}}.\end{aligned} \tag{15.34}$$

Taking the third relation and using the second relation to eliminate G_{gg} decouples the relations and gives a relation for G_{eg} alone:

$$(z+\hbar\Delta)G_{\mathrm{eg}} = \frac{\hbar\Omega}{2}G_{\mathrm{gg}} = \frac{\hbar\Omega}{2z}\left(1 + \frac{\hbar\Omega}{2}G_{\mathrm{eg}}\right). \tag{15.35}$$

Solving for $G_{\rm eg}$, we find

$$G_{\rm eg}(z) = \frac{\hbar\Omega/2}{z^2 + \hbar\Delta z - \hbar^2\Omega^2/4}. \tag{15.36}$$

Factoring the quadratic denominator,

$$G_{\rm eg}(z) = \frac{\hbar\Omega/2}{\left(z + \frac{\hbar\Delta}{2} + \frac{\hbar\tilde{\Omega}}{2}\right)\left(z + \frac{\hbar\Delta}{2} - \frac{\hbar\tilde{\Omega}}{2}\right)}, \tag{15.37}$$

we see that this matrix element has poles at

$$z = -\frac{\hbar\Delta}{2} \pm \frac{\hbar\tilde{\Omega}}{2}, \tag{15.38}$$

where as before

$$\tilde{\Omega} := \sqrt{\Omega^2 + \Delta^2} \tag{15.39}$$

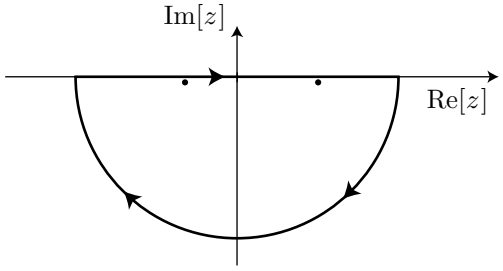
is the generalized Rabi frequency. Now the corresponding matrix element of the retarded Green function is

$$G^+_{\rm eg}(E) = G_{\rm eg}(E + i0^+) = \frac{\hbar\Omega/2}{\left(E + \frac{\hbar\Delta}{2} + \frac{\hbar\tilde{\Omega}}{2} + i0^+\right)\left(E + \frac{\hbar\Delta}{2} - \frac{\hbar\tilde{\Omega}}{2} + i0^+\right)}. \tag{15.40}$$

In this expression, the $i0^+$ terms have the effect of shifting both poles infinitesimally *below* the real axis. Now according to Eq. (15.16), the relevant propagator is given by the inverse Fourier transform

$$\langle {\rm e}|U(\tau,0)|{\rm g}\rangle = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^+_{\rm eg}(E) \qquad (\tau > 0). \tag{15.41}$$

Due to the form of the exponential factor, in the lower complex plane $E = x - iy$ gives a damping factor of the form $e^{-y\tau/\hbar}$. Thus to evaluate this integral, we complete the contour around the lower half-plane.



The contribution from the lower great half-circle vanishes due to the exponential factor and the asymptotic E^{-2} dependence of the remaining part of the integrand. However, the contour encloses both poles, and the Cauchy integral formula [Eq. (14.79)] states that the integral is $2\pi i$ times the sum of the two residues; however, because the contour runs clockwise, we should also include another minus sign.. We thus find

$$\begin{aligned}\langle {\rm e}|U(\tau,0)|{\rm g}\rangle &= \frac{(\hbar\Omega/2)e^{i(\Delta+\tilde{\Omega})\tau/2}}{-\hbar\tilde{\Omega}} + \frac{(\hbar\Omega/2)e^{i(\Delta-\tilde{\Omega})\tau/2}}{\hbar\tilde{\Omega}} \\ &= -i\frac{\Omega}{\tilde{\Omega}}e^{i\Delta\tau/2}\sin\frac{\tilde{\Omega}\tau}{2}.\end{aligned} \tag{15.42}$$

This gives the probability amplitude of finding the atom in the excited state at time τ, given that it was in the ground state at time 0. Thus, it agrees with our earlier solution in Section 5.2.2.1.

15.4 Level-Shift Operator

Again, let's start with a system with Hamiltonian $H = H_0 + V$, where H_0 is the unperturbed Hamiltonian. Working with the eigenstates of H_0, we may be interested in particular in the matrix element $G_{\alpha\alpha}(z)$, which reflects the survival probability of state $|\alpha\rangle$ due to the perturbation V. Now suppose we define projection operators for $|\alpha\rangle$ and everything *but* $|\alpha\rangle$:[2]

$$P_\alpha := |\alpha\rangle\langle\alpha|, \qquad Q_\alpha := 1 - P_\alpha = \sum_{j\neq\alpha} |j\rangle\langle j|. \tag{15.43}$$

Now recalling that the resolvent is defined by

$$(z - H)G(z) = (z - H_0 - V)G(z) = 1, \tag{15.44}$$

we can insert a $(P_\alpha + Q_\alpha)$ before the resolvent, and then operate on the right with P_α:

$$(z - H_0 - V)(P_\alpha + Q_\alpha)G(z)P_\alpha = P_\alpha. \tag{15.45}$$

This then becomes

$$(z - H_0 - V)P_\alpha[P_\alpha G(z)P_\alpha] + (z - H_0 - V)Q_\alpha[Q_\alpha G(z)P_\alpha] = P_\alpha. \tag{15.46}$$

where we used $P_\alpha^2 = P_\alpha$ and $Q_\alpha^2 = Q_\alpha$. Operating on the left with P_α, and using $P_\alpha Q_\alpha = Q_\alpha P_\alpha = 0$ [in particular, $P_\alpha(z - H_0)Q_\alpha = 0$],

$$P_\alpha(z - H)P_\alpha[P_\alpha G(z)P_\alpha] - P_\alpha V Q_\alpha[Q_\alpha G(z)P_\alpha] = P_\alpha. \tag{15.47}$$

Operating on the left of Eq. (15.46) instead with Q_α,

$$-Q_\alpha V P_\alpha[P_\alpha G(z)P_\alpha] + Q_\alpha(z - H)Q_\alpha[Q_\alpha G(z)P_\alpha] = 0. \tag{15.48}$$

Solving for $Q_\alpha G(z)P_\alpha$,

$$Q_\alpha G(z)P_\alpha = \frac{Q_\alpha}{Q_\alpha(z - H)Q_\alpha} V P_\alpha[P_\alpha G(z)P_\alpha], \tag{15.49}$$

and putting this into Eq. (15.47) to decouple these two equations,

$$P_\alpha(z - H)P_\alpha[P_\alpha G(z)P_\alpha] - P_\alpha V Q_\alpha \frac{Q_\alpha}{Q_\alpha(z - H)Q_\alpha} V P_\alpha[P_\alpha G(z)P_\alpha] = P_\alpha. \tag{15.50}$$

Simplifying and expanding out the Hamiltonian, we find

$$P_\alpha \left[z - H_0 - V - V\frac{Q_\alpha}{z - Q_\alpha H_0 Q_\alpha - Q_\alpha V Q_\alpha} V \right] P_\alpha G(z)P_\alpha = P_\alpha. \tag{15.51}$$

The part of the quantity in braces involving the interaction potential is important, and is called the **level-shift operator**:

$$R(z) := V + V\frac{Q_\alpha}{z - Q_\alpha H_0 Q_\alpha - Q_\alpha V Q_\alpha} V. \tag{15.52}$$
(level-shift operator)

Then Eq. (15.51) becomes

$$P_\alpha \left[z - H_0 - R(z) \right] P_\alpha G(z)P_\alpha = P_\alpha, \tag{15.53}$$

or

$$P_\alpha G(z)P_\alpha = \frac{P_\alpha}{z - P_\alpha H_0 P_\alpha - P_\alpha R(z) P_\alpha}. \tag{15.54}$$
(projection of the resolvent)

[2] We are essentially following the derivation of C. Cohen-Tannoudji *et al.*, *op. cit.*, Section III.B.2, p. 174. Essentially the same derivation is also given by P. Lambropoulos and P. Zoller, "Autoionizing states in strong laser fields," *Physical Review A* **24**, 379 (1981) (doi: 10.1103/PhysRevA.24.379).

In particular, by matching the coefficient of P_α, this means that the relevant matrix element of the resolvent is given by

$$G_{\alpha\alpha}(z) = \frac{1}{z - E_\alpha - R_{\alpha\alpha}(z)}.$$

(resolvent matrix element in terms of level-shift operator) (15.55)

Note that so far this is an *exact* expression. Furthermore, this formalism is easy to generalize to the case of survival among a subspace of multiple states: P_α is redefined as the sum of projectors over the relevant states, and we still have $Q_\alpha = 1 - P_\alpha$, with all the algebra so far carrying through except for the last expression above.

15.4.1 Decomposition of the Level-Shift Operator

In analyzing the resolvent $G(z)$, recall that we in general require the resolvent just next to the real axis, $G(E \pm i0^+)$. We will similarly require the level-shift operator at the same locations, $R(E \pm i0^+)$. Using the relation (Problem 14.6)

$$\frac{1}{x \pm i0^+} = \mathrm{P}\,\frac{1}{x} \mp i\pi\delta(x), \tag{15.56}$$

where the ''P'' denotes that an integral taken over that term is interpreted as a Cauchy principal value, we can rewrite Eq. (15.52) just off the real axis as

$$R(E \pm i0^+) = V + \mathrm{P}\,V\frac{Q_\alpha}{E - Q_\alpha H Q_\alpha}V \mp i\pi V Q_\alpha \delta(E - Q_\alpha H Q_\alpha) Q_\alpha V. \tag{15.57}$$

Now defining the Hermitian operators

$$\Delta(E) := \mathrm{P}\,\frac{1}{\hbar}V\frac{Q_\alpha}{E - Q_\alpha H Q_\alpha}V$$

(Hermitian part of level-shift operator) (15.58)

and

$$\Gamma(E) := \frac{2\pi}{\hbar}V Q_\alpha \delta(E - Q_\alpha H Q_\alpha) Q_\alpha V,$$

(anti-Hermitian part of level-shift operator) (15.59)

we see that these are the Hermitian and anti-Hermitian parts of the level-shift operator, which we can now write as

$$R(E \pm i0^+) = V + \hbar\Delta(E) \mp i\frac{\hbar\Gamma(E)}{2}. \tag{15.60}$$

(level-shift operator)

The operators Δ and Γ correspond to dispersive and dissipative effects (i.e., Stark shifts and decay) due to the perturbation V. To see this explicitly, we can write down the retarded Green function as

$$G^+_{\alpha\alpha}(E) = G_{\alpha\alpha}(E + i0^+) = \frac{1}{E - E_\alpha - V_{\alpha\alpha} - \hbar\Delta_{\alpha\alpha}(E) + i[\hbar\Gamma_{\alpha\alpha}(E)/2 + 0^+]}. \tag{15.61}$$

In this form, the resolvent appears to have a pole where E satisfies

$$E = E_\alpha + V_{\alpha\alpha} + \hbar\Delta_{\alpha\alpha}(E) - i\hbar\frac{\Gamma_{\alpha\alpha}(E)}{2}, \tag{15.62}$$

though of course the nature of the singularities of the resolvent depends on the exact forms of $\Delta_{\alpha\alpha}(E)$ or $\Gamma_{\alpha\alpha}(E)$. In the **pole approximation**, we assume that the interaction V leads only to a very weak perturbation and thus we can replace E by E_α on the right-hand side of this expression:

$$E \approx E_\alpha + V_{\alpha\alpha} + \hbar\Delta_{\alpha\alpha}(E_\alpha) - i\hbar\frac{\Gamma_{\alpha\alpha}(E_\alpha)}{2}. \tag{15.63}$$

Then the propagator from Eq. (15.16) becomes

$$\begin{aligned} U_{\alpha\alpha}(\tau,0) &= -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^+(E) \qquad (\tau > 0) \\ &= e^{-i(E_\alpha + V_{\alpha\alpha})\tau/\hbar}\, e^{-i\Delta_{\alpha\alpha}(E_\alpha)\tau}\, e^{-\Gamma_{\alpha\alpha}(E_\alpha)\tau/2}. \end{aligned} \tag{15.64}$$

Thus, we see that the energy of the state $|\alpha\rangle$ has been shifted by the first-order shift $V_{\alpha\alpha}$ as well as the higher-order shift $\Delta_{\alpha\alpha}(E_\alpha)$. The population of the state $|\alpha\rangle$ also *decays* at the rate $\Gamma_{\alpha\alpha}(E_\alpha)$.

15.4.2 Perturbation Expansion

Though exact, the expression we have is not yet so illuminating, particularly in the form of the level-shift operator (15.52). However, we can first note that

$$Q_\alpha \frac{1}{z - H_0} Q_\alpha = \frac{Q_\alpha}{z - Q_\alpha H_0 Q_\alpha} = \frac{Q_\alpha}{z - H_0}. \tag{15.65}$$

That is, since Q_α is a sum of projectors of eigenstates of H_0, the operators Q_α and $(z - H_0)^{-1}$ commute, and $(z - H_0)^{-1}$ acts as a scalar when multiplied by each of the projectors in Q_α. The level-shift operator (15.60) has the form $R = V + VQ_\alpha G Q_\alpha V$; then using the series (15.28) for the resolvent, we see the expanded level-shift operator has the expansion

$$R(z) = V + V\frac{Q_\alpha}{z - H_0}V + V\frac{Q_\alpha}{z - H_0}V\frac{Q_\alpha}{z - H_0}V + \cdots .$$

(series expansion of level-shift operator) (15.66)

Then the matrix element we need for Eq. (15.55) is

$$R_{\alpha\alpha}(z) = V_{\alpha\alpha} + \sum_{\beta\neq\alpha} \frac{V_{\alpha\beta}V_{\beta\alpha}}{z - E_\beta} + \sum_{\substack{\beta\neq\alpha \\ \gamma\neq\alpha}} \frac{V_{\alpha\beta}V_{\beta\gamma}V_{\gamma\alpha}}{(z - E_\beta)(z - E_\gamma)} + \cdots .$$

(series expansion of level-shift matrix element) (15.67)

This series can be truncated at any order in V. Note that, at least up to the second-order term, setting $z = E_\alpha$ gives the energy shift of the unperturbed level α in time-independent perturbation theory.

To gain a bit more insight into this perturbation expansion, note that Eq. (15.55) can be expanded as

$$G_{\alpha\alpha}(z) = \frac{1}{z - E_\alpha}\left[1 + \frac{R_{\alpha\alpha}(z)}{z - E_\alpha} + \frac{R^2_{\alpha\alpha}(z)}{(z - E_\alpha)^2} + \cdots\right]. \tag{15.68}$$

Thus, even when $R_{\alpha\alpha}(z)$ is truncated to some order in the perturbation, $G_{\alpha\alpha}(z)$ in the form (15.55) still contains contributions at all orders in V, and so corresponds to some *nonperturbative* expansion of the resolvent. We can compare this to the direct series expansion (15.28) of the resolvent, which gives

$$G_{\alpha\alpha}(z) = \frac{1}{z - E_\alpha}\left[1 + \frac{V_{\alpha\alpha}}{z - E_\alpha} + \sum_\beta \frac{V_{\alpha\beta}V_{\beta\alpha}}{(z - E_\alpha)(z - E_\beta)} + \sum_{\beta\gamma} \frac{V_{\alpha\beta}V_{\beta\gamma}V_{\gamma\alpha}}{(z - E_\alpha)(z - E_\beta)(z - E_\gamma)} + \cdots\right]. \tag{15.69}$$

[Note that the summation indices are not restricted as in Eq. (15.67).] The two series expansions (15.67) and (15.69) have exactly the same terms, but correspond to different *orderings* of the term. The direct expansion (15.69) is strictly in powers of the perturbation V. However, in the expression (15.67), we note that $R_{\alpha\alpha}$ never contains $(z - E_\alpha)$ in any denominator. Thus, the expansion (15.68) for $G_{\alpha\alpha}(z)$ corresponds to a power series (Laurent series) in $(z - E_\alpha)^{-1}$. This focus on the pole at E_α is sensible if we are interested in the effects of the perturbation on the state $|\alpha\rangle$, whose properties are determined by the resolvent in the vicinity of the perturbed energy of state $|\alpha\rangle$. Given a small perturbation, the pole should not have shifted too much,

so we are still maintaining accuracy in the relevant region.[3] (For example, the residue of the shifted pole gives the perturbed dynamics, and thus accuracy in this neighborhood is important.)

It is also convenient to give the series expansion in terms of the Hermitian and anti-Hermitian parts of $R(E \pm i0^+)$. The Hermitian part $\Delta(E)$ from Eq. (15.58) expands in the same way as the resolvent in Eq. (15.67):

$$\Delta(E) = V + V\frac{Q_\alpha}{E - H_0}V + V\frac{Q_\alpha}{E - H_0}V\frac{Q_\alpha}{E - H_0}V + \cdots .$$

(expansion of level-shift operator, Hermitian part) (15.70)

The diagonal matrix elements are then

$$\hbar\Delta_{\alpha\alpha}(E) = V_{\alpha\alpha} + \sum_{\beta\neq\alpha}\frac{V_{\alpha\beta}V_{\beta\alpha}}{E - E_\beta} + \sum_{\substack{\beta\neq\alpha\\ \gamma\neq\alpha}}\frac{V_{\alpha\beta}V_{\beta\gamma}V_{\gamma\alpha}}{(E - E_\beta)(E - E_\gamma)} + \cdots .$$

(matrix-element expansion of level-shift operator, Hermitian part) (15.71)

The same matrix element of the anti-Hermitian part, from Eq. (15.59), is

$$\Gamma_{\alpha\alpha}(E) = \frac{2\pi}{\hbar}\sum_{\substack{\beta\neq\alpha\\ \gamma\neq\alpha}} V_{\alpha\beta}V_{\gamma\alpha}\delta(E - \delta_{\beta\gamma}E_\beta - V_{\beta\gamma}),$$

(anti-Hermitian part of level-shift operator) (15.72)

Notice the similarity to Fermi's Golden Rule [Eq. (11.58)], except here we are summing over all possible decay paths.

15.5 Spontaneous Decay

15.5.1 Pole Approximation

As in our previous analysis of spontaneous emission in Chapter 11, we take the state $|\mathrm{e}\rangle$ to be coupled to $|\mathrm{g}, 1_{\mathbf{k},\zeta}\rangle$, with the relevant matrix element of the interaction Hamiltonian reading [Eq. (11.63)]

$$\langle \mathrm{e}|H_{\mathrm{AF}}|\mathrm{g}, 1_{\mathbf{k},\zeta}\rangle = \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0 V}}\,(\hat{\varepsilon}_{\mathbf{k},\zeta}\cdot\mathbf{d}_{\mathrm{ge}})\,e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{15.73}$$

Since we wish to examine the survival probability, we consider the diagonal matrix element of the resolvent, which from Eq. (15.55) we can write in terms of the level-shift operator as

$$G_{\mathrm{ee}}(z) = \frac{1}{z - E_{\mathrm{e}} - R_{\mathrm{ee}}(z)}. \tag{15.74}$$

Near the real axis, we can use Eq. (15.60) to write

$$G_{\mathrm{ee}}(E \pm i0^+) = \frac{1}{E - E_{\mathrm{e}} - \hbar\Delta_{\mathrm{ee}}(E) \pm i\hbar\Gamma_{\mathrm{ee}}(E)/2}, \tag{15.75}$$

where we have used $(H_{\mathrm{AF}})_{\mathrm{ee}} = 0$. We showed in Section 15.4.1 that in the pole approximation, $\Gamma_{\mathrm{ee}}(E_{\mathrm{e}})$ is the decay rate of the excited state, and $\Delta_{\mathrm{ee}}(E_{\mathrm{e}})$ is the energy (Stark) shift of the excited state. We now must evaluate the dispersive matrix element from Eqs. (15.70),

$$\hbar\Delta_{\mathrm{ee}}(E) = \sum_{\beta\neq\mathrm{e}}\frac{(H_{\mathrm{AF}})_{\mathrm{e}\beta}(H_{\mathrm{AF}})_{\beta\mathrm{e}}}{E - E_\beta} + \sum_{\substack{\beta\neq\mathrm{e}\\ \gamma\neq\mathrm{e}}}\frac{(H_{\mathrm{AF}})_{\mathrm{e}\beta}(H_{\mathrm{AF}})_{\beta\gamma}(H_{\mathrm{AF}})_{\gamma\mathrm{e}}}{(E - E_\beta)(E - E_\gamma)} + \cdots . \tag{15.76}$$

[3] This point is discussed in detail by Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *op. cit.*, Section III.B.1, p. 172.

Truncating this expression to lowest order and making the pole approximation ($E \approx E_{\rm e}$), we have

$$\hbar\Delta_{\rm ee}(E) \approx \sum_{\beta\neq {\rm e}} \frac{(H_{\rm AF})_{{\rm e}\beta}(H_{\rm AF})_{\beta{\rm e}}}{E_{\rm e} - E_\beta}, \tag{15.77}$$

which is just the Lamb shift of the excited state $|{\rm e}\rangle$ in second-order perturbation theory (evaluated in Section 13.12). Evidently, solving the full expansion (15.76) self-consistently for E leads to the *exact* Lamb-shifted energy

$$\tilde{E}_{\rm e} = E_{\rm e} + \hbar\Delta_{\rm ee}(\tilde{E}_{\rm e}). \tag{15.78}$$

We also need the absorptive matrix element

$$\Gamma_{\rm ee}(E) = \frac{2\pi}{\hbar} \sum_{\substack{\beta\neq {\rm e}\\ \gamma\neq {\rm e}}} (H_{\rm AF})_{{\rm e}\beta}(H_{\rm AF})_{\gamma{\rm e}}\delta(E - \delta_{\beta\gamma}E_\beta - (H_{\rm AF})_{\beta\gamma}). \tag{15.79}$$

Note that in the pole approximation, we can drop the off-diagonal terms where $\beta \neq \gamma$, since the interaction $H_{\rm AF}$ is assumed to be weak enough to justify perturbation theory, and thus the delta function can never reach resonance without the presence of the E_β term in the argument. Thus,

$$\Gamma_{\rm ee}(E_{\rm e}) \approx \frac{2\pi}{\hbar} \sum_{\beta\neq {\rm e}} |(H_{\rm AF})_{{\rm e}\beta}|^2\, \delta(E_{\rm e} - E_\beta - (H_{\rm AF})_{\beta\beta}). \tag{15.80}$$

But there is no first-order shift due to the dipole interaction $H_{\rm AF}$, so

$$\Gamma_{\rm ee}(E_{\rm e}) \approx \frac{2\pi}{\hbar} \sum_{\beta\neq {\rm e}} |(H_{\rm AF})_{{\rm e}\beta}|^2\, \delta(E_{\rm e} - E_\beta). \tag{15.81}$$

This expression is equivalent to Fermi's Golden Rule, as in Eq. (11.58), explicitly summed over the continuum of final states $|{\rm g}, 1_{\mathbf{k},\zeta}\rangle$, and we know from Section 11.6.1 that this leads to the correct spontaneous decay rate in free space. Again, not making any perturbative expansion would yield the same expression for the decay rate, but with $E_{\rm e}$ replaced by $\tilde{E}_{\rm e}$ (i.e., the transition frequency ω_0 in the decay-rate formula (11.68) is the exact value, including the Lamb shift computed to all orders). Henceforth, we will simply absorb the Lamb shift into the bare-state energy E_α, since we will assume that when applying the results of any calculation, we will use the *observed* atomic energies, which already include the correct Lamb shift.

15.5.2 Line Shape of Spontaneous Decay

To compute the line shape of spontaneous decay, we will use the matrix element $\langle {\rm g}, 1_{\mathbf{k},\zeta}|G(z)|{\rm e}\rangle$ of the resolvent operator, which will give the rate to create a photon in mode $(\mathbf{k},\zeta)$, of frequency $\omega_{\mathbf{k}} = ck$. Starting with the second identity in Eqs. (15.31), we have

$$\langle {\rm g}, 1_{\mathbf{k},\zeta}|G(z)|{\rm e}\rangle = \frac{1}{z - \hbar\omega_{\mathbf{k}}} \langle {\rm g}, 1_{\mathbf{k},\zeta}|H_{\rm AF}|{\rm e}\rangle\langle {\rm e}|G(z)|{\rm e}\rangle. \tag{15.82}$$

The excited-state matrix element of the resolvent is given by Eq. (15.61), with the shift and decay rate determined in the last section:

$$\langle {\rm e}|G^+(E)|{\rm e}\rangle = \frac{1}{E - \hbar\omega_0 - \hbar\Delta_{\rm ee} + i\hbar\Gamma/2 + i0^+}. \tag{15.83}$$

In the pole approximation, recall that Γ is the usual decay rate of $|{\rm e}\rangle$, and $\Delta_{\rm ee}$ is the Lamb shift of the excited state, which we will absorb into the excited-state energy $\hbar\omega_0$:

$$\langle {\rm e}|G^+(E)|{\rm e}\rangle = \frac{1}{E - \hbar\omega_0 + i\hbar\Gamma/2 + i0^+}. \tag{15.84}$$

Thus, from (15.82) and (15.84), we can write down the Green function

$$\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|G^+(E)|\mathrm{e}\rangle = \frac{\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle}{(E-\hbar\omega_{\mathbf{k}}+i0^+)(E-\hbar\omega_0+i\hbar\Gamma/2+i0^+)}. \tag{15.85}$$

Transforming to find the propagator,

$$\begin{aligned}\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|U(\tau,0)|\mathrm{e}\rangle &= -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|G^+(E)|\mathrm{e}\rangle \\ &= -\frac{\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle}{2\pi i}\int_{-\infty}^{\infty} dE\, \frac{e^{-iE\tau/\hbar}}{(E-\hbar\omega_{\mathbf{k}}+i0^+)(E-\hbar\omega_0+i\hbar\Gamma/2+i0^+)}.\end{aligned} \tag{15.86}$$

We carry out this integral as in Section 15.3.1 by completing the contour around the lower half-plane. The result is

$$\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|U(\tau,0)|\mathrm{e}\rangle = \langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle\left[\frac{e^{-i\omega_{\mathbf{k}}\tau}}{\hbar(\omega_{\mathbf{k}}-\omega_0)+i\hbar\Gamma/2}+\frac{e^{-i\omega_0\tau}e^{-\Gamma\tau/2}}{-\hbar(\omega_{\mathbf{k}}-\omega_0)-i\hbar\Gamma/2}\right], \tag{15.87}$$

where the first term is the residue of the pole at $E=\hbar\omega_{\mathbf{k}}$ and the second term is the residue of the pole at $E=\hbar\omega_0-i\hbar\Gamma/2$. Simplifying this expression, we find

$$\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|U(\tau,0)|\mathrm{e}\rangle = \frac{\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle}{\hbar(\omega_{\mathbf{k}}-\omega_0+i\Gamma/2)}\left[e^{-i\omega_{\mathbf{k}}\tau}-e^{-i\omega_0\tau}e^{-\Gamma\tau/2}\right]. \tag{15.88}$$

Therefore, the probability to decay into mode $(\mathbf{k},\zeta)$ is the squared modulus of this amplitude:

$$P(\mathbf{k},\zeta,\tau) = |\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|U(\tau,0)|\mathrm{e}\rangle|^2 = \frac{|\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle|^2}{\hbar^2[(\omega_{\mathbf{k}}-\omega_0)^2+\Gamma^2/4]}\left[1+e^{-\Gamma\tau}-2e^{-\Gamma\tau/2}\cos(\omega_{\mathbf{k}}-\omega_0)\tau\right]. \tag{15.89}$$

Let's expand out the matrix element, using Eq. (15.73), which becomes

$$|\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle|^2 = \frac{\hbar\omega_{\mathbf{k}}}{6\epsilon_0 V}|d_{\mathrm{ge}}|^2 \tag{15.90}$$

for an isotropic atom in quantization volume V. In passing to the continuum limit, recall that we make the replacement

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{(2\pi)^3}\int d^3k. \tag{15.91}$$

Converting to spherical coordinates, in free space this is isotropic, so we may carry out the angular integral, and change variables $\omega = ck$, with the result

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{2\pi^2 c^3}\int_0^{\infty} d\omega\, \omega^2. \tag{15.92}$$

Since the summations are implicit in a later calculation of a probability, for now we can let

$$V^{-1} \longrightarrow \frac{\omega^2}{2\pi^2 c^3}d\omega \tag{15.93}$$

so that

$$|\langle \mathrm{g}, 1_{\mathbf{k},\zeta}|H_{\mathrm{AF}}|\mathrm{e}\rangle|^2 = \frac{\hbar\omega^3}{12\pi^2\epsilon_0 c^3}|d_{\mathrm{ge}}|^2 d\omega = \frac{\hbar^2\Gamma}{4\pi}d\omega, \tag{15.94}$$

where

$$\Gamma := \frac{\omega_0^3|d_{\mathrm{ge}}|^2}{3\pi\epsilon_0\hbar c^3} \tag{15.95}$$

is the usual free-space decay rate, and we have taken $\omega \approx \omega_0$. The we can rewrite Eq. (15.89) as a continuous probability density for emission after changing variables: at frequency $\omega = \omega_{\mathbf{k}}$ and polarization ζ at time τ:

$$P(\omega, \zeta, \tau)\, d\omega = \frac{\Gamma}{4\pi[(\omega-\omega_0)^2 + \Gamma^2/4]} \left[1 + e^{-\Gamma\tau} - 2e^{-\Gamma\tau/2}\cos(\omega-\omega_0)\tau\right] d\omega. \tag{15.96}$$

If we don't care about the polarization of the emitted light, we can sum over the two orthogonal polarizations to find

$$P(\omega, \tau)\, d\omega = \frac{\Gamma}{2\pi[(\omega-\omega_0)^2 + \Gamma^2/4]} \left[1 + e^{-\Gamma\tau} - 2e^{-\Gamma\tau/2}\cos(\omega-\omega_0)\tau\right] d\omega.$$

(time-dependent spontaneous-emission line shape) (15.97)

In the long-time limit, when a photon has certainly been emitted, this becomes

$$P(\omega)\, d\omega = \frac{\Gamma}{2\pi[(\omega-\omega_0)^2 + \Gamma^2/4]}\, d\omega, \tag{15.98}$$

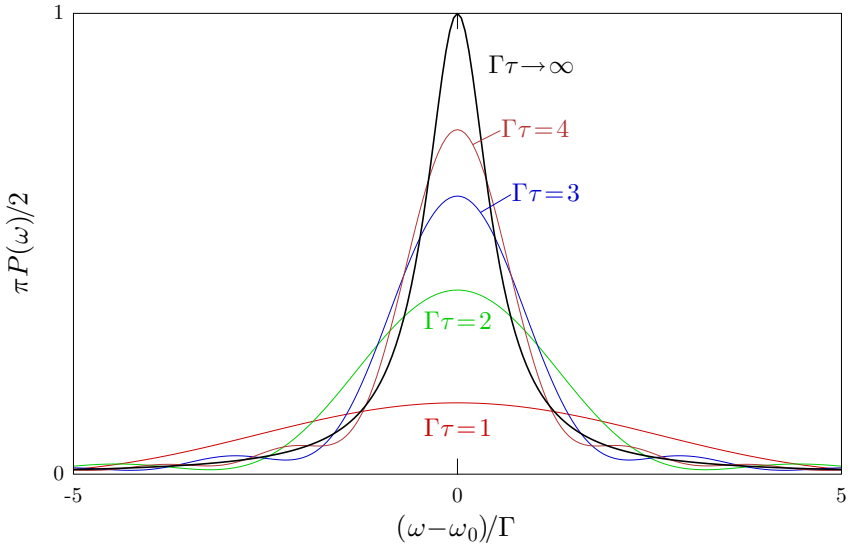
(spontaneous-emission line shape)

which is a properly normalized Lorentzian line shape of full width at half maximum of Γ. Alternately, integrating the transition probability density over all frequencies gives

$$P(\tau) = \int_0^\infty d\omega\, P(\omega, \tau) \approx \int_{-\infty}^\infty d\omega\, P(\omega, \tau) = 1 - e^{-\Gamma\tau}, \tag{15.99}$$

which is the expected exponential behavior.

Of course, the above treatment predicts some nontrivial time dependence to the line shape, including an oscillatory component. The line shape is shown here for several interaction times.



The departures from the long-time Lorentzian only occur if $\Gamma\tau$ is not too large. However, to resolve any such difference would require a measurement time much longer than Γ^{-1}, so these differences would be difficult to detect experimentally. Intuitively, the oscillations arise here when the exponential decay is "interrupted" at time τ. The spectral response should thus be the convolution of the long-time Lorentzian with a sinc function that becomes increasingly narrow in time. Unless, of course, there is a way to "freeze" the interaction after some short time. This is not normally possible in spontaneous emission, but is possible in the spontaneous Raman problem we consider below.

15.5.3 Branches of the Resolvent

In the previous section, we found the following situation: an atom in the excited state $|\mathrm{e}\rangle$ decays to the ground state $|\mathrm{g}\rangle$ an energy $\hbar\omega_0$ lower, but the energy photon emitted has an uncertainty that doesn't necessarily

match, as it can be emitted in a range of width $\hbar\Gamma$ around the transition energy $\hbar\omega_0$. Clearly, energy is conserved on *average*, but we have a time-independent Hamiltonian for the coupled quantum atom–field system, so energy should be conserved in *each* process individually. So what gives? Well, first of all, we will see below that the coupled state $|\mathrm{e}\rangle$ no longer has a well-defined energy. We might expect this: since it decays, it is not an eigenstate of the full Hamiltonian, so it should not have a definite energy. Furthermore, the energy that goes into *preparing* the atom in $|\mathrm{e}\rangle$ must be yet *more* uncertain, since in view of the unstable nature of the state, the preparation must take place on a time scale much shorter than $1/\Gamma$. Thus, the uncertainty in the energy of the emitted photon is easily accounted for by the uncertainty in the energy of setting up the problem.

But now let's explore the idea that the coupled state $|\mathrm{e}\rangle$ has no well-defined energy. The matrix element

$$\langle \mathrm{e}|G_0(z)|\mathrm{e}\rangle = \frac{1}{z-\hbar\omega_0} \tag{15.100}$$

of the *unperturbed* resolvent has a single pole at the excited-state energy $\hbar\omega_0$. However, in Eq. (15.75) we wrote an expression for the same matrix element of the *coupled* resolvent

$$\langle \mathrm{e}|G(E\pm i0^+)|\mathrm{e}\rangle = \frac{1}{E-\hbar[\omega_0+\Delta_{\mathrm{ee}}(E)]\pm i\hbar\Gamma_{\mathrm{ee}}(E)/2}, \tag{15.101}$$

in the vicinity of the real axis, where we see that the value of $G(z)$ jumps as we cross the real axis. In coupling the atom to the continuum, the pole at the bare atomic energy appears to have changed into a branch cut, reflecting the unstable nature of the state, and the fact that it has no well-defined energy. Note that this was not the case in the Rabi-flopping example in Eq. (15.35)—rather, it is a consequence of coupling the excited state to a continuum. In particular, note from Eq. (15.75) that the retarded Green function $G^+_{\mathrm{ee}}(E) = G_{\mathrm{ee}}(E+i0^+)$ appears to have a pole at $E=\hbar[\omega_0+\Delta_{\mathrm{ee}}(E)]-i\hbar\Gamma_{\mathrm{ee}}(E)/2$, *below* the real axis. However, $G^+_{\mathrm{ee}}(E)$ is only defined *above* the real axis; just below the real axis, the resolvent is given by the different value $G^-_{\mathrm{ee}}(E) = G_{\mathrm{ee}}(E-i0^+)$. So while the pole appears to have changed into a branch cut, we can view it as having ''disappeared'' behind the branch cut that formed. That is, we may still find the pole if we analytically continue $G^+_{\mathrm{ee}}(E) = G_{\mathrm{ee}}(E+i0^+)$ into the lower half plane. In this case, rather than suffer the discontinuity in the resolvent in crossing the branch cut, one can think of crossing continuously through the real axis, but ending up in a different **Riemann sheet**, or the **second Riemann sheet**, since the function value in the lower half plane defined in this way differs from the function value given with the branch cut. Then we have the function

$$G^{\mathrm{II}}_{\mathrm{ee}}(z) = \frac{1}{z-\hbar[\omega_0+\Delta_{\mathrm{ee}}(z)]+i\hbar\Gamma_{\mathrm{ee}}(z)/2}, \tag{15.102}$$

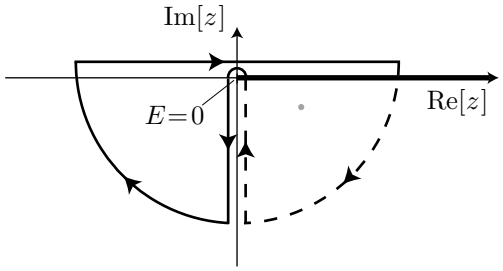
which has the same functional form as $G^+_{\mathrm{ee}}(E) = G_{\mathrm{ee}}(E+i0^+)$, but unlike $G_{\mathrm{ee}}(z)$, it is defined in this way in the lower half of the complex plane $\mathrm{Im}[z]<0$, and thus corresponds to the function extended to the second Riemann sheet.

The usual example for extending through branch cuts is $\log(z)$, which, when defined as a function, has a branch cut along the negative real axis. Thus, $\log(-x\pm i\delta) = \log(x)\pm i(\pi+\delta)$ if $x>0$, which is a difference of $2\pi i + 2\delta$ that does not vanish as $\delta \longrightarrow 0$. This preserves the continuity (analyticity) and single-valued nature of $\log z$, basically by excluding the discontinuity from the function's domain to $|\arg z|<\pi$ (i.e., such that the discontinuity is never ''detected'' along a continuous path). Of course, adding any integer multiple of $2\pi i$ to $\log z$ is still valid as a logarithm, since when inverted, $\exp(\log z + 2\pi i n) = z$. Each n thus corresponds to an ''extension'' of the logarithm function to a different Riemann sheet.

15.5.4 Nonexponential Decay

Thus, in making the pole approximation in Section 15.5.1 to arrive at the rate Γ of exponential decay of the excited state, we were somewhat sloppy, as it turns out. Implicitly, we solved an integral by closing a contour around the lower half-plane as in Section 15.3.1. But to do so, we needed to cross a branch cut, and we need to be more careful about this.

To locate the branch cut in the resolvent, recall Eq. (15.81) for $\Gamma_{\text{ee}}(E)$. There we can see that the branch cut exists anywhere that $\Gamma_{\text{ee}}(E)$ is nonzero, since it represents the discontinuity across the branch cut. This function is in fact nonzero for any energy where there exists a possible decay energy E_β. But since the decay is to state $|\text{g}, 1_{\mathbf{k},\zeta}\rangle$, where the atomic ground state has zero energy, the possible decay energies are $\hbar\omega_{\mathbf{k}}$, that is to say, any positive real number. So the branch cut extends from the origin, and along the entire positive real axis. When we close the contour, including the branch cut, we can do so as shown here.



The contour starts into the lower half of the complex plane, but in crossing the branch cut, we enter the second Riemann sheet (where the contour is shown by a dotted line). To compensate for this, we return along the negative real axis, turn around the branch point at the origin, and then continue along the great semicircle. This contour still encloses the pole, which is again in the second Riemann sheet (and thus shown as a grey dot), and the residue here results in the exponential decay that we have already derived. The extra contribution is due to the two portions along the negative imaginary axis, which comes from extending Eq. (15.16) to these portions, along with the Eqs. (15.101) and (15.102) for the appropriate Green-function expressions, with the result

$$\begin{aligned}\langle \text{e}|U^{\parallel}(\tau,0)|\text{e}\rangle &= -\frac{1}{2\pi i}\int_{-\infty}^{0} d(iy)\, e^{y\tau/\hbar}\left[G^{\text{II}}_{\text{ee}}(iy) - G_{\text{ee}}(iy)\right] \\ &= -\frac{1}{2\pi}\int_{-\infty}^{0} dy\, e^{y\tau/\hbar}\left[\frac{1}{iy - \hbar\omega_0 + i\hbar\Gamma/2} - \frac{1}{iy - \hbar\omega_0 - i\hbar\Gamma/2}\right] \\ &= \frac{1}{2\pi i}\int_{0}^{\infty} dy\, e^{-y\tau/\hbar}\left[\frac{1}{y - i\hbar\omega_0 - \hbar\Gamma/2} - \frac{1}{y - i\hbar\omega_0 + \hbar\Gamma/2}\right],\end{aligned} \tag{15.103}$$

where we have absorbed the Lamb shift into ω_0. To compute the integral here, first note that the exponential integral (Problem 13.2) $E_1(z)$ is defined by

$$E_1(z) := \int_z^\infty dy\, \frac{e^{-y}}{y}. \tag{15.104}$$

Letting $y \longrightarrow \alpha y$ (with $\text{Re}[\alpha] > 0$ to guarantee convergence),

$$E_1(z) = \int_{z/\alpha}^\infty dy\, \frac{e^{-\alpha y}}{y}, \tag{15.105}$$

and letting $y \longrightarrow y + \beta$ (β not on the negative real axis, and $\beta \neq 0$, as we will see from the expansion for $E_1(z)$ later on),

$$E_1(z) = e^{-\alpha\beta}\int_{z/\alpha-\beta}^\infty dy\, \frac{e^{-\alpha y}}{y+\beta}. \tag{15.106}$$

Now taking $z = \alpha\beta$, we find the integral formula

$$\int_0^\infty \frac{e^{-\alpha y}}{y+\beta}\, dy = e^{\alpha\beta}E_1(\alpha\beta) \qquad (\text{Re}[\alpha] > 0,\ \beta \notin \mathbb{R}^-,\ \beta \neq 0). \tag{15.107}$$

Then Eq. (15.103) becomes

$$\langle \text{e}|U^{\parallel}(\tau,0)|\text{e}\rangle = \frac{e^{-i\omega_0\tau}}{2\pi i}\left[e^{-\Gamma\tau/2}E_1(-i\omega_0\tau - \Gamma\tau/2) - e^{\Gamma\tau/2}E_1(-i\omega_0\tau + \Gamma\tau/2)\right] \tag{15.108}$$

Using the asymptotic expansion (Problem 15.3)

$$E_1(z) = \frac{e^{-z}}{z}\left\{1 - \frac{1}{z} + \frac{2}{z^2} - \frac{3!}{z^3} + \cdots + \frac{n!}{(-z)^n} + \cdots\right\} \tag{15.109}$$

to lowest order,

$$\begin{aligned}\langle \mathrm{e}|U^{\parallel}(\tau,0)|\mathrm{e}\rangle &= \frac{1}{2\pi i}\left[\frac{1}{-i\omega_0\tau - \Gamma\tau/2} - \frac{1}{-i\omega_0\tau + \Gamma\tau/2}\right] + O(\tau^{-2}) \\ &= \frac{i\Gamma}{2\pi(\omega_0^2 + \Gamma^2/4)\tau} + O(\tau^{-2}).\end{aligned} \tag{15.110}$$

Thus, we see that at long times, this calculation predicts a slow power-law decay of the excited-state population as t^{-2}, which dominates the exponential decay at long times. Unfortunately, this calculation is not quite right, as we have ignored the dependence of $\Gamma_{\mathrm{ee}}(E)$ on E, replacing it by its pole-approximation value of Γ.

15.5.5 Frequency-Dependent Decay Rate

To derive an improved expression for $\Gamma_{\mathrm{ee}}(E)$, we return to Eq. (15.81), but without making the pole approximation:

$$\Gamma_{\mathrm{ee}}(E) \approx \frac{2\pi}{\hbar}\sum_{\beta\neq \mathrm{e}} |(H_{\mathrm{AF}})_{\mathrm{e}\beta}|^2\, \delta(E - E_\beta). \tag{15.111}$$

In our treatment of spontaneous emission using Fermi's Golden rule (Section 11.6), we showed that the sum over the continuum modes and the delta function are replaced by the density of states $\rho(E)$, so that

$$\Gamma_{\mathrm{ee}}(E) = \frac{2\pi}{\hbar}\,|(H_{\mathrm{AF}})_{\mathrm{e}\beta}|^2\,\rho(E). \tag{15.112}$$

The density of states is [Eq. (11.66)]

$$\rho(E) = \frac{E^2 V}{\pi^2\hbar^3 c^3}, \tag{15.113}$$

and we have the usual matrix element [Eq. (11.63)]

$$|(H_{\mathrm{AF}})_{\mathrm{e}\beta}|^2 = |\langle \mathrm{e}|H_{\mathrm{AF}}|\mathrm{g}, 1_{\mathbf{k},\zeta}\rangle|^2 = \frac{\hbar\omega_{\mathbf{k}} d_{\mathrm{ge}}^2}{6\epsilon_0 V} = \frac{E d_{\mathrm{ge}}^2}{6\epsilon_0 V}, \tag{15.114}$$

upon identifying the initial and final-state energies as the same (due to the delta function), so that $E = \hbar\omega_{\mathbf{k}}$. Combining Eqs. (15.111)-(15.114), we find

$$\Gamma_{\mathrm{ee}}(E) = \frac{E^3 d_{\mathrm{ge}}^2}{3\pi\epsilon_0\hbar^4 c^3} = \Gamma\frac{E^3}{(\hbar\omega_0)^3}, \tag{15.115}$$

after using the usual expression

$$\Gamma = \frac{\omega_0^3 d_{\mathrm{ge}}^2}{3\pi\epsilon_0\hbar c^3}. \tag{15.116}$$

This is, in fact, *not* the answer we want. This result comes from using the electric-dipole Hamiltonian $H_{\mathrm{AF}} = -\mathbf{d}\cdot\mathbf{E}$. On the other hand, we can use the alternative Coulomb-gauge interaction Hamiltonian $H_{\mathrm{AF}} = (e/m_{\mathrm{e}})\mathbf{p}_{\mathrm{e}}\cdot\mathbf{A}$ from Section 9.3. As we discussed before in Section 9.3.2, the ratio of the dipole to the Coulomb Hamiltonians is ω/ω_0, or in this case $E/\hbar\omega_0$. Since the square of the matrix element enters our calculation, Eq. (15.115) instead becomes

$$\Gamma_{\mathrm{ee}}(E) = \frac{\Gamma}{\hbar\omega_0}E \tag{15.117}$$

(decay-function matrix element)

in the Coulomb gauge. Of course, in the pole approximation none of this matters, since we identify $\omega \approx \omega_0$ anyway.

Why prefer the Coulomb gauge to the dipole gauge now? Recall that they should give the same result, but provided that *all* atomic levels are included in the interaction, which we are not doing. But in a practical sense, putting the cubic expression (15.115) into Eq. (15.103) in place of Γ means we have to factor a cubic polynomial to find the poles and thus carry out the contour integration, and frankly, who wants to do *that*? The other reason comes from the origin of $\Gamma(E)$ as the singular part of $\Delta(E)$, as in the factorization starting with Eq. (15.56). To see this, we can start with the second-order truncation of Eq. (15.66)

$$R_{\rm ee}(z) = V_{\rm ee} + \sum_{\beta\neq {\rm e}} \frac{V_{{\rm e}\beta}V_{\beta{\rm e}}}{z-E_\beta} = \sum_{\beta\neq {\rm e}} \frac{V_{{\rm e}\beta}V_{\beta{\rm e}}}{z-E_\beta}, \tag{15.118}$$

where we have dropped the first-order term, which we will comment on in a bit. The usual mode sum becomes

$$\sum_{\beta\neq {\rm e}} |V_{{\rm e}\beta}|^2 \longrightarrow \frac{\hbar^2\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\omega^3, \tag{15.119}$$

with $E_\beta \longrightarrow \hbar\omega$, so that

$$R_{\rm ee}(z) = \frac{\hbar^2\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3}{z-\hbar\omega}. \tag{15.120}$$

This is the (divergent) expression for the Lamb shift in the rotating-wave approximation (before separating out the decay). Then applying Eq. (15.56),

$$\begin{aligned} R_{\rm ee}(E\pm i0^+) &= \frac{\hbar^2\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3}{E-\hbar\omega} \mp i\frac{\hbar^2\Gamma}{2\omega_0^3}\int_0^\infty d\omega\,\omega^3\delta(E-\hbar\omega) \\ &= \frac{\hbar^2\Gamma}{2\pi\omega_0^3}\int_0^\infty d\omega\,\frac{\omega^3}{E-\hbar\omega} \mp i\frac{\hbar\Gamma E^3}{2(\hbar\omega_0)^3}, \end{aligned} \tag{15.121}$$

where we identify the last term as $(\mp i\hbar/2)\Gamma_{\rm ee}(E)$, gives us the result (15.115). But recall that the first (Lamb-shift) term here, which diverges as ω^2, must be renormalized by adding the dipole self-energy contribution [Eq. (13.153), Section 13.12.2.1]

$$H_{\mathbf{P}^\perp} = \frac{1}{2\epsilon_0}\int d^3r\, P^{\perp 2}(\mathbf{r}), \tag{15.122}$$

which comes in at first order here, and then the result must be mass-renormalized. The result then agrees with the Coulomb-gauge calculation, which is only linearly divergent:

$$R_{\rm ee}(E\pm i0^+) = \frac{\hbar^2\Gamma}{2\pi\omega_0}\int_0^\infty d\omega\,\frac{\omega}{E-\hbar\omega} \mp i\frac{\hbar\Gamma E}{2\hbar\omega_0}. \tag{15.123}$$

Since the renormalization is necessary to get a physical answer anyway, here we will prefer the Coulomb gauge (particularly in that the renormalization is not straightforward when z is not fixed to ω_0).

15.5.6 Branch Contribution

Now we return to calculating corrections to exponential decay, using our improved expression (15.117) for $\Gamma_{\rm ee}(E)$. Retracing the derivation above, starting with Eq. (15.103), we find

$$\begin{aligned}
\langle {\rm e}|U^{\parallel}(\tau,0)|{\rm e}\rangle &= -\frac{1}{2\pi i}\int_{-\infty}^{0} d(iy)\, e^{y\tau/\hbar}\left[G_{\rm ee}^{\rm II}(iy) - G_{\rm ee}(iy)\right] \\
&= -\frac{1}{2\pi}\int_{-\infty}^{0} dy\, e^{y\tau/\hbar}\left[\frac{1}{iy-\hbar\omega_0+i\hbar\Gamma(iy)/2} - \frac{1}{iy-\hbar\omega_0-i\hbar\Gamma(iy)/2}\right] \\
&= \frac{1}{2\pi i}\int_{0}^{\infty} dy\, e^{-y\tau/\hbar}\left[\frac{1}{y-i\hbar\omega_0-iy\Gamma/2\omega_0} - \frac{1}{y-i\hbar\omega_0+iy\Gamma/2\omega_0}\right] \\
&= \frac{\beta_-}{2\pi i}\int_{0}^{\infty} dy\, \frac{e^{-y\tau/\hbar}}{y-i\hbar\omega_0\beta_-} - \frac{\beta_+}{2\pi i}\int_{0}^{\infty} dy\, \frac{e^{-y\tau/\hbar}}{y-i\hbar\omega_0\beta_+} \\
&= \frac{1}{2\pi i}\Big[\beta_- e^{i\beta_-\omega_0\tau}E_1(i\beta_-\omega_0\tau) - \beta_+ e^{i\beta_+\omega_0\tau}E_1(i\beta_+\omega_0\tau)\Big],
\end{aligned} \tag{15.124}$$

where we have introduced the notation

$$\beta_\pm := \frac{1}{1\pm i\dfrac{\Gamma}{2\omega_0}} = \frac{\omega_0(\omega_0\mp i\Gamma/2)}{\omega_0^2+\Gamma^2/4}, \tag{15.125}$$

where $\beta_\pm \approx 1$ for the typical case of $\omega_0 \gg \Gamma$. The asymptotic expansion to lowest order vanishes,

$$\langle {\rm e}|U^{\parallel}(\tau,0)|{\rm e}\rangle = \frac{1}{2\pi i}\left[\frac{\beta_-}{i\beta_-\omega_0\tau} - \frac{\beta_+}{i\beta_+\omega_0\tau}\right] + O(\tau^{-2}) = O(\tau^{-2}), \tag{15.126}$$

so keeping the second-order term in the expansion (15.109), we find

$$\begin{aligned}
\langle {\rm e}|U^{\parallel}(\tau,0)|{\rm e}\rangle &= \frac{1}{2\pi i}\left[-\frac{\beta_-}{(i\beta_-\omega_0\tau)^2} + \frac{\beta_+}{(i\beta_+\omega_0\tau)^2}\right] + O(\tau^{-3}) \\
&= \frac{1}{2\pi i\omega_0^2\tau^2}\left[\frac{1}{\beta_-} - \frac{1}{\beta_+}\right] + O(\tau^{-3}) \\
&= -\frac{\Gamma}{2\pi\omega_0^3\tau^2} + O(\tau^{-3}).
\end{aligned} \tag{15.127}$$

Thus, this dominates the exponential decay at long times, so the probability amplitude decreases asymptotically as t^{-4}. That is, the survival probability $P(t)$ eventually becomes[4]

$$P(t) \sim \left(\frac{\Gamma}{2\pi\omega_0^3}\right)^2 \frac{1}{t^4}, \tag{15.128}$$

but for typical transitions where $\omega_0 \gg \Gamma$, the crossover from exponential to this power-law behavior happens only at extremely long times, where the survival probability is essentially undetectably small.

15.5.7 Pole Contribution

Returning to Eq. (15.75) for the retarded Green function,

$$G_{\rm ee}^{+}(E) = \frac{1}{E - E_{\rm e} + i\hbar\Gamma_{\rm ee}(E)/2}, \tag{15.129}$$

[4] J. Mostowski and K. Wódkiewicz, ''On the Decay Law of Unstable States,'' *Bulletin L'Académie Polonaise des Science, Série des Sciences Mathématiques, Astronomiques et Physiques* **21**, 1027 (1973); P. L. Knight and P. W. Milonni, ''Long-Time Deviations from Exponential Decay in Atomic Spontaneous Emission Theory,'' *Physics Letters* **56A** 275 (1976) (doi: 10.1016/0375-9601(76)90306-6).

we can use the expression (15.117) for the function $\Gamma_{ee}(E)$ to go beyond the pole approximation for the contribution of the pole, which previously just gave exponential decay at rate Γ. We now have

$$G^+_{ee}(E) = \frac{1}{E - \hbar\omega_0 + i\Gamma E/2\omega_0} = \frac{\beta_+}{E - \beta_+\hbar\omega_0}, \tag{15.130}$$

and thus, with Eq. (15.16), the pole contribution to the propagator becomes

$$\begin{aligned}\langle e|U^\bullet(\tau,0)|e\rangle &= -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^+_{ee}(E) \\ &= \beta_+ e^{-i\beta_+\omega_0\tau} \\ &= \frac{\omega_0(\omega_0 - i\Gamma/2)}{\omega_0^2 + \Gamma^2/4} e^{-i\tilde{\omega}_0\tau} e^{-\tilde{\Gamma}\tau/2} = \frac{(\tilde{\omega}_0 - i\tilde{\Gamma}/2)}{\omega_0} e^{-i\tilde{\omega}_0\tau} e^{-\tilde{\Gamma}\tau/2},\end{aligned} \tag{15.131}$$

where the contour completed around the lower half-plane encloses the single pole at β_+, and we have defined the shifted resonance frequency

$$\tilde{\omega}_0 := \frac{\omega_0^3}{\omega_0^2 + \Gamma^2/4} = \frac{\omega_0}{1 + (\Gamma/2\omega_0)^2} \tag{15.132}$$

(shifted resonance frequency)

and the shifted decay rate

$$\tilde{\Gamma} := \frac{\Gamma}{1 + (\Gamma/2\omega_0)^2}, \tag{15.133}$$

(shifted decay rate)

both of which have very small corrections of order $(\Gamma/\omega_0)^2$ (typically $\sim 10^{-16}$ for alkali dipole transitions) as a result of the more precise treatment of the pole, accounting for the frequency dependence of the decay rate (15.117).

15.5.8 Short Times

To focus on short times, we will need the series expansion of the exponential integral around $z = 0$ (see Problem 13.2),

$$E_1(x) = -\gamma - \log x - \sum_{j=1}^{\infty} \frac{(-1)^j x^j}{j j!}, \tag{15.134}$$

where $\gamma \approx 0.577\,215\,664\,901\,532\,860\,607$ is the Euler–Mascheroni constant. From Eq. (15.124), we had the branch contribution

$$\langle e|U^\parallel(\tau,0)|e\rangle = \frac{1}{2\pi i}\Big[\beta_- e^{i\beta_-\omega_0\tau} E_1(i\beta_-\omega_0\tau) - \beta_+ e^{i\beta_+\omega_0\tau} E_1(i\beta_+\omega_0\tau)\Big] \tag{15.135}$$

to the propagator, where

$$\beta_\pm := \frac{1}{1 \pm i\dfrac{\Gamma}{2\omega_0}} = \frac{\omega_0(\omega_0 \mp i\Gamma/2)}{\omega_0^2 + \Gamma^2/4} = \frac{\tilde{\omega}_0 \mp i\tilde{\Gamma}/2}{\omega_0}. \tag{15.136}$$

Sadly, this expression diverges at short times. Namely, the logarithmic term in Eq. (15.134) leads to a short-time scaling of

$$\langle e|U^\parallel(\tau,0)|e\rangle = -\frac{1}{2\pi i}\Big[\beta_- \log(i\beta_-\omega_0\tau) - \beta_+ \log(i\beta_+\omega_0\tau)\Big]. \tag{15.137}$$

Since the terms do not exactly cancel, there is a logarithmic divergence at short times. Evidently, there is a problem at short times with taking the integral over all frequencies.

15.5.8.1 Hard Cutoff

To handle this, we can introduce a high-energy cutoff Λ for the integrals. This echoes the strategy in the Lamb shift (Section 13.12), where the argument was that the logarithmically divergent integral should be cut off at the large energy $\Lambda = m_e c^2$, where relativistic effects should take over and naturally cut off the integral.[5] To evaluate the integral with the frequency cutoff,

$$\begin{aligned}\int_0^\Lambda \frac{e^{-\alpha y}}{y+\beta}\,dy &= \int_0^\infty \frac{e^{-\alpha y}}{y+\beta}\,dy - \int_\Lambda^\infty \frac{e^{-\alpha y}}{y+\beta}\,dy \\ &= \int_0^\infty \frac{e^{-\alpha y}}{y+\beta}\,dy - e^{-\alpha\Lambda}\int_0^\infty \frac{e^{-\alpha y}}{y+\beta+\Lambda}\,dy\end{aligned} \tag{15.138}$$

where we have let $y \longrightarrow y+\Lambda$ in the second step. Then using Eq. (15.138), we have the integral formula

$$\int_0^\Lambda \frac{e^{-\alpha y}}{y+\beta}\,dy = e^{\alpha\beta}E_1(\alpha\beta) - e^{\alpha(\beta+\Lambda)}E_1[\alpha(\beta+\Lambda)] \qquad (\mathrm{Re}[\alpha]>0,\ \beta\notin\mathbb{R}^-,\ \beta+\Lambda\notin\mathbb{R}^-,\ \beta\neq 0,\ \beta+\Lambda\neq 0). \tag{15.139}$$

Then retracing the derivation of Eq. (15.124), but cutting off the integral at energy Λ,

$$\begin{aligned}\langle \mathrm{e}|U^{\parallel}(\tau,0)|\mathrm{e}\rangle &= \frac{\beta_-}{2\pi i}\int_0^\Lambda dy\,\frac{e^{-y\tau/\hbar}}{y-i\hbar\omega_0\beta_-} - \frac{\beta_+}{2\pi i}\int_0^\Lambda dy\,\frac{e^{-y\tau/\hbar}}{y-i\hbar\omega_0\beta_+} \\ &= \frac{1}{2\pi i}\Big[\beta_- e^{i\beta_-\omega_0\tau}E_1(i\beta_-\omega_0\tau) - \beta_+ e^{i\beta_+\omega_0\tau}E_1(i\beta_+\omega_0\tau)\Big] \\ &\qquad - \frac{1}{2\pi i}\Big[\beta_- e^{(i\beta_-\omega_0-\Lambda/\hbar)\tau}E_1[(i\beta_-\omega_0-\Lambda/\hbar)\tau] - \beta_+ e^{(i\beta_+\omega_0-\Lambda/\hbar)\tau}E_1[(i\beta_+\omega_0-\Lambda/\hbar)\tau]\Big].\end{aligned} \tag{15.140}$$

at short times, using the expansion (15.134), this becomes

$$\langle \mathrm{e}|U^{\parallel}(\tau,0)|\mathrm{e}\rangle = \frac{1}{2\pi i}\Big[\beta_-\log(1-\Lambda/i\beta_-\hbar\omega_0) - \beta_+\log(1-\Lambda/i\beta_+\hbar\omega_0)\Big] + O(\tau), \tag{15.141}$$

which is finite. Notice that even for large Λ, the logarithms are still comparatively of order unity and almost the same because the $\beta_\pm$ are both close to unity. The two terms then nearly cancel, with the difference of order

$$\beta_- - \beta_+ = \frac{\omega_0\Gamma}{\omega_0^2+\Gamma^2/4} \approx \frac{\Gamma}{\omega_0}, \tag{15.142}$$

which is much smaller than unity. Thus, the contribution at $\tau = 0$ is negligible compared to the pole contribution. In principle, the decay rate should vanish at $\tau = 0$ (Section 11.7.1), but this does not appear to be the case here with this cutoff or the soft cutoff below.

The problem with this cutoff procedure is that it modifies the long-time scaling. The asymptotic calculation to $O(\tau^{-1})$ from Eq. (15.126) should now have the extra contribution

$$\langle \mathrm{e}|U^{\parallel\Lambda}(\tau,0)|\mathrm{e}\rangle = \frac{1}{2\pi i}\left[\frac{\beta_-}{i\beta_-\omega_0\tau-\Lambda/\hbar} - \frac{\beta_+}{i\beta_+\omega_0\tau-\Lambda/\hbar}\right] + O(\tau^{-2}) \tag{15.143}$$

from the new cutoff terms. However, due to the presence of Λ, these terms no longer cancel, but

$$\begin{aligned}\langle \mathrm{e}|U^{\parallel\Lambda}(\tau,0)|\mathrm{e}\rangle &= \frac{\chi}{\pi[\chi^2+(1-i\xi^2)](\Lambda/\hbar)\tau} + O(\tau^{-2}) \\ &\approx \frac{\chi}{\pi(\Lambda/\hbar)\tau} + O(\tau^{-2}),\end{aligned} \tag{15.144}$$

[5] For issues regarding this hard cutoff and the dipole approximation in this calculation, see J. Seke and W. N. Herfort, "Deviations from exponential decay in the case of spontaneous emission from a two-level atom," *Physical Review A* **38**, 833 (1988) (doi: 10.1103/PhysRevA.38.833); J. Seke and W. Herfort, "Finite-time deviations from exponential decay in the case of spontaneous emission from a two-level hydrogenic atom," *Physical Review A* **40**, 1926 (1989) (doi: 10.1103/PhysRevA.40.1926). Note that they obtain a slightly different, cutoff-dependent asymptotic scaling.

where $\chi := \Gamma/2\omega_0 \ll 1$ and $\xi := \hbar\omega_0/\Lambda \ll 1$. Thus, the long-time scaling behavior is

$$P(t) \sim \frac{\chi^2}{\pi^2(\Lambda/\hbar)^2\tau^2} = \frac{\Gamma^2}{4\pi^2\omega_0^2(\Lambda/\hbar)^2\tau^2}, \tag{15.145}$$

which is not the τ^{-4} behavior we expect from Eq. (15.128). Note that there was no divergence problem at long times before, so this new scaling is a symptom that indicates this cutoff procedure is not quite right: the long-time scaling behavior should be cutoff-independent. Since it scales as Λ^{-2}, we might expect the numerical coefficient to be small, but it will still eventually dominate.

15.5.8.2 Soft Cutoff

A different scenario for cutting off the integral is to smoothly bring the integral to zero at large frequencies. Physically, this represents the fact that the effects of short wavelengths should be attenuated by the finite size of the atom, since the atom ''smooths'' out the wave on this length scale. Thus, it is appropriate to take a cutoff energy of $\Lambda \sim 2\pi c/a$, where a is the atomic radius (Bohr radius). We explicitly miss this effect in the dipole approximation, which treats the atom as a point. A simple functional form for the cutoff is an exponential of the form $e^{-y/\Lambda}$ to cut off large energies y. This corresponds to assuming a Lorentzian shape for the atom, with the cutoff modeling the convolution of the atomic profile with the field modes of different frequencies. Thus, Eq. (15.124) becomes

$$\begin{aligned}
\langle \mathrm{e}|U^{\parallel}(\tau,0)|\mathrm{e}\rangle &= \frac{1}{2\pi i}\int_0^\infty dy\, e^{-y\tau/\hbar}\left[\frac{1}{y - i\hbar\omega_0 - iy\Gamma/2\omega_0} - \frac{1}{y - i\hbar\omega_0 + iy\Gamma/2\omega_0}\right] e^{-y/\Lambda} \\
&= \frac{1}{2\pi i}\int_0^\infty dy\, e^{-y(\tau+\hbar/\Lambda)/\hbar}\left[\frac{1}{y - i\hbar\omega_0 - iy\Gamma/2\omega_0} - \frac{1}{y - i\hbar\omega_0 + iy\Gamma/2\omega_0}\right] \\
&= \frac{1}{2\pi i}\Big\{\beta_- e^{i\beta_-\omega_0(\tau+\hbar/\Lambda)}E_1[i\beta_-\omega_0(\tau+\hbar/\Lambda)] - \beta_+ e^{i\beta_+\omega_0(\tau+\hbar/\Lambda)}E_1[i\beta_+\omega_0(\tau+\hbar/\Lambda)]\Big\},
\end{aligned} \tag{15.146}$$

which is exactly the same as the result without any cutoff, but with the time displaced forward $\tau \longrightarrow \tau+\hbar/\Lambda$. This avoids the singularity at $\tau = 0$ since there

$$\langle \mathrm{e}|U^{\parallel}(\tau=0,0)|\mathrm{e}\rangle = -\frac{1}{2\pi i}\Big[\beta_- \log[i\beta_-\hbar\omega_0/\Lambda] - \beta_+ \log[i\beta_+\hbar\omega_0/\Lambda]\Big], \tag{15.147}$$

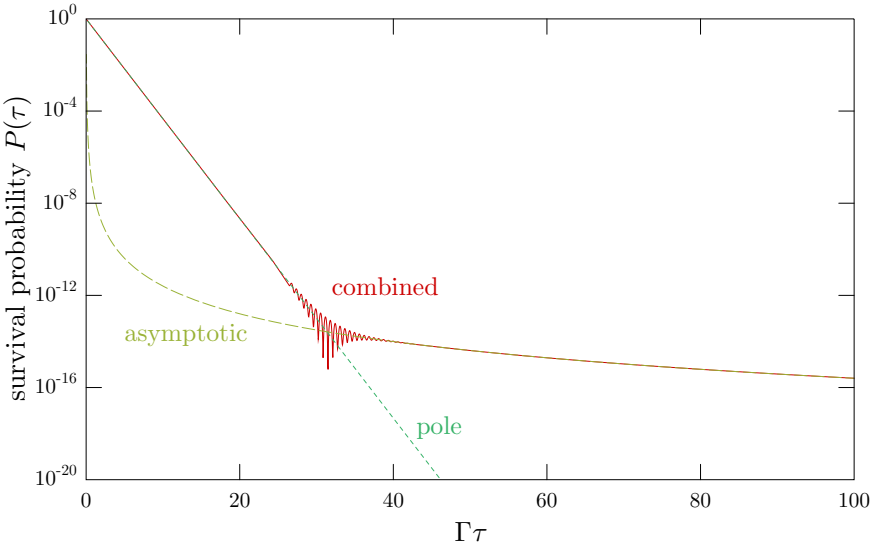
which is again finite and negligible compared to unity. At long times, $\tau + \hbar/\Lambda \approx \tau$, so we obtain the correct τ^{-4} scaling at long times.

15.5.9 Intermediate Times

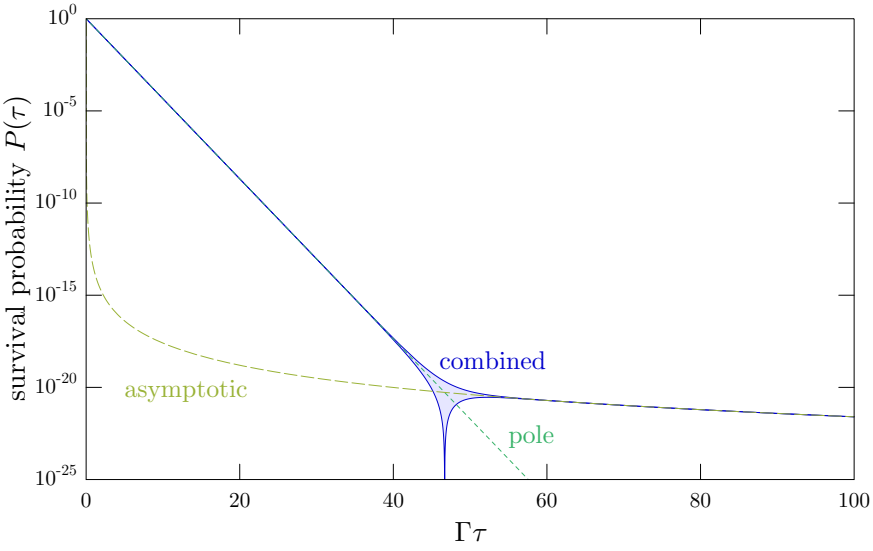
The total survival probability is then given by combining the pole amplitude from Eq. (15.131) and the branch amplitude from Eq. (15.146):

$$\begin{aligned}
P(\tau) &= |\langle \mathrm{e}|U(\tau,0)|\mathrm{e}\rangle|^2 = \big|\langle \mathrm{e}|U^{\bullet}(\tau,0)|\mathrm{e}\rangle + \langle \mathrm{e}|U^{\parallel}(\tau,0)|\mathrm{e}\rangle\big|^2 \\
&= \left|\beta_+ e^{-i\beta_+\omega_0\tau} + \frac{1}{2\pi i}\Big\{\beta_- e^{i\beta_-\omega_0(\tau+\hbar/\Lambda)}E_1[i\beta_-\omega_0(\tau+\hbar/\Lambda)] - \beta_+ e^{i\beta_+\omega_0(\tau+\hbar/\Lambda)}E_1[i\beta_+\omega_0(\tau+\hbar/\Lambda)]\Big\}\right|^2.
\end{aligned} \tag{15.148}$$

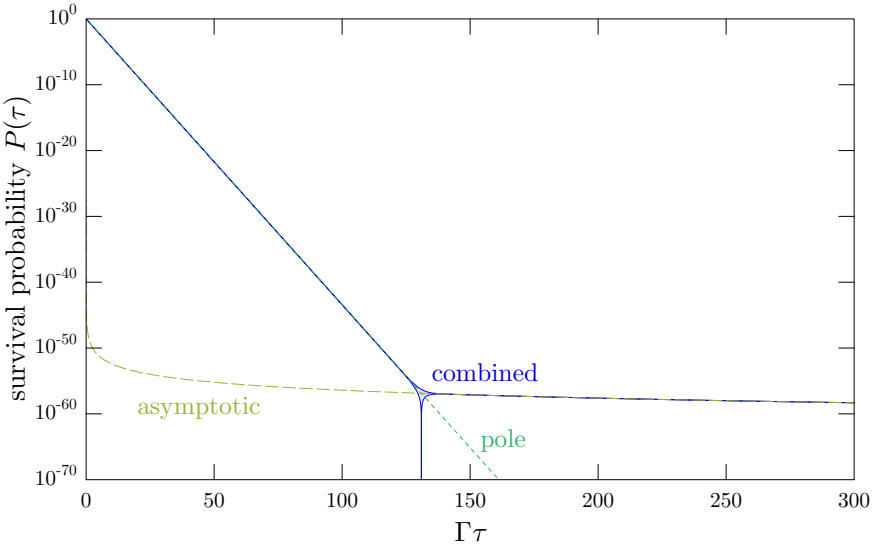
We have seen that for short times, the pole contribution dominates, and thus the decay is exponential. For long times, the branch contribution dominates, and the decay crosses over to a power law. At intermediate times, when both contributions are important, the behavior is somewhat more complicated. The pole contribution always oscillates at optical frequencies, but we have seen that asymptotically, the branch contribution does not. Thus, there can be optical-frequency beating between the two contributions. Unfortunately, it is difficult to visualize these high-frequency beats on the long decay time scales, except for the unrealistic case where Γ is not too different from ω_0. This plot shows the case $\Gamma/\omega_0 = 10^{-1}$, with a cutoff $\Lambda/\hbar\Gamma = 10^2$, along with the exponential pole decay alone and the asymptotic τ^{-4} decay from Eq. (15.128). The oscillations at the crossover are clear here.



This next plot shows the case $\Gamma/\omega_0 = 10^{-2}$, with a cutoff $\Lambda/\hbar\Gamma = 10^3$, again, along with the exponential pole decay alone and the asymptotic τ^{-4} decay from Eq. (15.128). Here, the oscillations are already too rapid to meaningfully plot, so we instead plot the envelope of the oscillations (calculated from the sum and difference of the absolute values of the two contributions). The increases in both parameters shift the crossover to a later time and smaller survival probability.



For the realistic case of ^{87}Rb on the 780 nm D_2 transition, we have $\Gamma/2\pi = 6.07$ MHz and $\omega_0/2\pi = 384.23$ THz. We can also take $\Lambda = hc/a$, where the atomic radius $a = 2.99\ a_0$ comes from the dipole moment of $2.99\ ea_0$ for the $|F = 2, m_F = 2\rangle \longrightarrow |F' = 3, m'_F = 3\rangle$ stretched-state hyperfine transition, and $a_0 \approx 0.529$ Å is the Bohr radius. Thus, the parameters we need are $\Gamma/\omega_0 = 1.58 \times 10^{-8}$ and $\Lambda/\hbar\Gamma = 5 \times 10^{10}$. Note that to obtain the correct asymptotic behavior numerically, a cancellation between the different terms is necessary to get a smaller number, so arbitrary-precision arithmetic is required in this regime (standard double-precision, floating-point arithmetic gives an error-dominated asymptotic scaling as τ^{-2}.) Also, note that had we instead used the hard cutoff, the asymptotic scaling (15.145) makes a substantial difference in the long-time region of this plot, even with the relativistic cutoff.



The crossover occurs after some 130 lifetimes, with a survival probability well below 10^{-50}. Long-time nonexponential decay of atomic spontaneous emission is unlikely to ever be seen in an experiment. This is the way the numbers work out in this problem but keep in mind that long-time nonexponential decay is a *generic* phenomenon, and the method here is a good way to get the *full* time dependence of the decay.

15.5.10 Interpretation of Nonexponential Decay

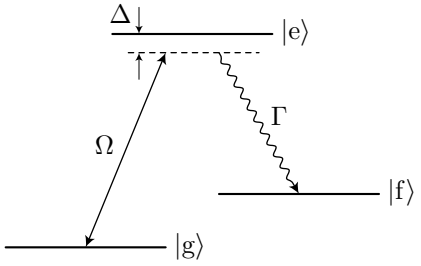
What is the meaning of this long, nonexponential tail of the decay curve? First of all recall that exponential decay follows from having a *constant* rate of decay,

$$\partial_t P = -\Gamma P(t) \quad \longrightarrow \quad P(t) = P(0)\, e^{-\Gamma t}. \tag{15.149}$$

This fundamentally means that the system decays in exactly the same way at each instant in time, *independent of its past history*. This solution is unique, so any deviation from exponential decay points to a "memory" in the system, or a breakdown of the Markov approximation (the Born–Markov master equation of Section 4.5, or equivalently the Lindblad master equation of Section 19.1 assume the Markov approximation and thus cannot predict this kind of nonexponential decay). The "memory" of the atom of the emitted photon is somewhat counterintuitive, however: evidently the photon-emission amplitude, even though it propagates rapidly away from the atom to infinity, has some long tail that interacts with the atom and interferes with the remaining decay amplitude.

15.6 Spontaneous Raman Scattering

As another example of the resolvent method, consider spontaneous Raman scattering in a three-level Λ atom, where the transition $|\mathrm{g}\rangle \longrightarrow |\mathrm{e}\rangle$ is coupled by a laser field with detuning Δ, and spontaneous decay occurs on the $|\mathrm{e}\rangle \longrightarrow |\mathrm{f}\rangle$ transition at rate Γ.



Technically, $|\mathrm{e}\rangle$ must also decay to $|\mathrm{g}\rangle$ if the transition can be coupled by the laser, but we assume that this decay route is much slower than the decay to $|\mathrm{f}\rangle$ (see Problem 15.5). This model also describes quenching of a metastable state by coupling to a quickly decaying state, or influence on the metastability of a state by coupling to another decaying level.[6]

The free Hamiltonian is

$$H_0 = \hbar\Delta|\mathrm{g}\rangle\langle\mathrm{g}| - \hbar\omega_{\mathrm{ef}}|\mathrm{f}\rangle\langle\mathrm{f}| \tag{15.150}$$

in the rotating frame of the laser field (Section 5.1.5), taking the energy E_e of $|\mathrm{e}\rangle$ to be zero, and where $\omega_{\mathrm{ef}} = (E_\mathrm{e} - E_\mathrm{f})/\hbar$. The atom–field coupling is given in the rotating-wave approximation by

$$V = \frac{\hbar\Omega}{2}\left(\sigma + \sigma^\dagger\right) + \sum_{\mathbf{k},\zeta} \hbar\Big[g_{\mathbf{k},\zeta}(\mathbf{r})\sigma_\mathrm{f}^\dagger a_{\mathbf{k},\zeta} + \mathrm{H.c.}\Big], \tag{15.151}$$

where $\sigma := |\mathrm{g}\rangle\langle\mathrm{e}|$, $\sigma_\mathrm{f} := |\mathrm{f}\rangle\langle\mathrm{e}|$, Ω is the usual Rabi frequency for the laser field, and $g_{\mathbf{k},\zeta}$ are the free-space coupling coefficients for the vacuum field [Eq. (11.7)].

Now we can focus on the coupling between $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$. Defining the projector $P := |\mathrm{g}\rangle\langle\mathrm{g}| + |\mathrm{e}\rangle\langle\mathrm{e}|$ and the orthogonal projector $Q := 1 - P$, we can use the result (15.54) in terms of the level-shift operator,

$$PG(z)P = \frac{P}{z - PH_0P - PR(z)P}. \tag{15.152}$$

The resolvent in the subspace of $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ can then be written in matrix form as

$$\begin{bmatrix} G_{\mathrm{ee}}(z) & G_{\mathrm{eg}}(z) \\ G_{\mathrm{ge}}(z) & G_{\mathrm{gg}}(z) \end{bmatrix} = \begin{bmatrix} z - E_\mathrm{e} - R_{\mathrm{ee}}(z) & -R_{\mathrm{eg}}(z) \\ -R_{\mathrm{ge}}(z) & z - E_\mathrm{g} - R_{\mathrm{gg}}(z) \end{bmatrix}^{-1}. \tag{15.153}$$

Since we want to analyze the survival probability of $|\mathrm{g}\rangle$, we can use the inversion formula

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{15.154}$$

to write

$$G_{\mathrm{gg}}(z) = \frac{z - E_\mathrm{e} - R_{\mathrm{ee}}(z)}{[z - E_\mathrm{e} - R_{\mathrm{ee}}(z)][z - E_\mathrm{g} - R_{\mathrm{gg}}(z)] - R_{\mathrm{ge}}(z)R_{\mathrm{eg}}(z)}, \tag{15.155}$$

which we will now evaluate.

Using Eq. (15.60) for the level-shift operator, we can compute the matrix element

$$R_{\mathrm{ee}}(E + i0^+) = V_{\mathrm{ee}} + \hbar\Delta_{\mathrm{ee}}(E) - i\frac{\hbar\Gamma_{\mathrm{ee}}(E)}{2}. \tag{15.156}$$

In the pole approximation, we take $E = E_\mathrm{e}$, and then $\hbar\Delta_{\mathrm{ee}}(E_\mathrm{e})$ is the Lamb shift—which we absorb into the excited-state energy—of $|\mathrm{e}\rangle$ due to the coupling to the vacuum continuum, and $\Gamma = \Gamma_{\mathrm{ee}}(E_\mathrm{e})$ represents the spontaneous decay of $|\mathrm{e}\rangle \longrightarrow |\mathrm{f}\rangle$ due to the vacuum coupling, and thus

$$R_{\mathrm{ee}}(E + i0^+) = -i\frac{\hbar\Gamma}{2}. \tag{15.157}$$

Similarly,

$$R_{\mathrm{gg}}(E + i0^+) = 0, \tag{15.158}$$

since $|\mathrm{g}\rangle$ is not coupled to the vacuum continuum. To get the off-diagonal matrix elements, we can use the perturbative expansion (15.66) up to second order,

$$R(z) = V + V\frac{Q}{z - H_0}V, \tag{15.159}$$

[6] Claude Cohen-Tannoudji, Jacques Dupont-Roc, and Gilbert Grynberg, *Atom–Photon Interactions: Basic Processes and Applications* (Wiley, 1992), Section III.C.3.

so that to second order

$$R_{\rm eg}(z) = V_{\rm eg} = \frac{\hbar\Omega}{2}, \tag{15.160}$$

with the same result for $R_{\rm ge}(z)$, since

$$\langle {\rm e}|VQV|{\rm g}\rangle = \sum_{\mathbf{k},\zeta} \langle {\rm e}|V|{\rm f},1_{\mathbf{k},\zeta}\rangle\langle {\rm f},1_{\mathbf{k},\zeta}|V|{\rm g}\rangle = 0, \tag{15.161}$$

again since $|{\rm g}\rangle$ is not coupled to the vacuum.

Now that we have the level-shift operator in the subspace of $|{\rm g}\rangle$ and $|{\rm e}\rangle$, we note that a nice interpretation of Eq. (15.152) is that $PG(z)P$ is the resolvent operator of the effective subspace Hamiltonian

$$P[H_0 - R(z)]P = \begin{bmatrix} E_{\rm e} - i\hbar\Gamma/2 & \hbar\Omega/2 \\ \hbar\Omega/2 & E_{\rm g} \end{bmatrix}, \tag{15.162}$$

which is now no longer Hermitian due to the decay. Returning now to the resolvent matrix element (15.155), which now becomes

$$G^+_{\rm gg}(E) = G_{\rm gg}(E + i0^+) = \frac{E - E_{\rm e} + i\hbar\Gamma/2}{(E - E_{\rm e} + i\hbar\Gamma/2)(E - E_{\rm g}) - (\hbar\Omega/2)^2}, \tag{15.163}$$

which has poles

$$E_\pm = \frac{1}{2}\left[E_{\rm e} + E_{\rm g} - \frac{i\hbar\Gamma}{2} \pm \sqrt{\left(E_{\rm e} - E_{\rm g} - \frac{i\hbar\Gamma}{2}\right)^2 + (\hbar\Omega)^2}\right] \tag{15.164}$$

(shifted energies)

corresponding to the eigenvalues of the effective Hamiltonian (15.162).

Thus, the propagator from the inversion formula (15.16) gives the survival amplitude

$$\langle {\rm g}|U(\tau,0)|{\rm g}\rangle = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} dE\, e^{-iE\tau/\hbar}\, G^+_{\rm gg}(E). \tag{15.165}$$

We can do this integral via a contour around the lower half-plane, which encloses both poles, since the square root of Eq. (15.164) always has an imaginary part smaller in magnitude than $i\hbar\Gamma/2$ (this is apparent when visualizing the squaring and square root operations as respectively doubling and halving the complex angle). Then with

$$G^+_{\rm gg}(E) = \frac{E - E_{\rm e} + i\hbar\Gamma/2}{(E - E_+)(E - E_-)}, \tag{15.166}$$

the propagator becomes

$$\langle {\rm g}|U(\tau,0)|{\rm g}\rangle = \frac{1}{E_+ - E_-}\left[\left(E_+ - E_{\rm e} + \frac{i\hbar\Gamma}{2}\right)e^{-iE_+\tau/\hbar} - \left(E_- - E_{\rm e} + \frac{i\hbar\Gamma}{2}\right)e^{-iE_-\tau/\hbar}\right], \tag{15.167}$$

(survival amplitude)

which is a fairly complicated expression, which we can analyze more intuitively in the the limits of weak and strong pumping.

15.6.1 Weak Pumping

For weak pumping, Ω is small, and thus we can expand the square root in Eq. (15.164) to lowest order in Ω:

$$E_\pm \approx \frac{1}{2}\left[E_{\rm e} + E_{\rm g} - \frac{i\hbar\Gamma}{2} \pm \left(E_{\rm e} - E_{\rm g} - \frac{i\hbar\Gamma}{2}\right)\left(1 + \frac{(\hbar\Omega)^2}{2\left(E_{\rm e} - E_{\rm g} - \frac{i\hbar\Gamma}{2}\right)^2}\right)\right], \tag{15.168}$$

or

$$E_+ \approx E_\mathrm{e} - \frac{i\hbar\Gamma}{2} + \frac{(\hbar\Omega)^2}{4\left(E_\mathrm{e} - E_\mathrm{g} - \dfrac{i\hbar\Gamma}{2}\right)}$$
$$E_- \approx E_\mathrm{g} - \frac{(\hbar\Omega)^2}{4\left(E_\mathrm{e} - E_\mathrm{g} - \dfrac{i\hbar\Gamma}{2}\right)}. \tag{15.169}$$

Note that the eigenvalues here are only small corrections to the original eigenvalues. Recalling that $E_\mathrm{e} = 0$ and $E_\mathrm{g} = \hbar\Delta$,

$$\frac{E_+}{\hbar} \approx -\frac{i\Gamma}{2} - \frac{\Omega^2}{4\left(\Delta + \dfrac{i\Gamma}{2}\right)} = -\frac{i\Gamma}{2} - \frac{\Omega^2\,(\Delta - i\Gamma/2)}{4\left(\Delta^2 + \Gamma^2/4\right)} = -\frac{i\Gamma}{2} - \tilde{\Delta} + \frac{i\tilde{\Gamma}}{2}$$
$$\frac{E_-}{\hbar} \approx \Delta + \frac{\Omega^2}{4\left(\Delta + \dfrac{i\Gamma}{2}\right)} = \Delta + \frac{\Omega^2\,(\Delta - i\Gamma/2)}{4\left(\Delta^2 + \Gamma^2/4\right)} = \Delta + \tilde{\Delta} - \frac{i\tilde{\Gamma}}{2}, \tag{15.170}$$

where we have defined

$$\tilde{\Delta} := \left[\frac{\Omega^2}{4\left(\Delta^2 + \Gamma^2/4\right)}\right]\Delta \tag{15.171}$$
(shift of $|\mathrm{g}\rangle$)

and

$$\tilde{\Gamma} := \left[\frac{\Omega^2}{4\left(\Delta^2 + \Gamma^2/4\right)}\right]\Gamma. \tag{15.172}$$
(decay rate of $|\mathrm{g}\rangle$)

Thus, the survival amplitude (15.167) becomes

$$\langle \mathrm{g}|U(\tau,0)|\mathrm{g}\rangle = \frac{\left(\tilde{\Delta} - \dfrac{i\tilde{\Gamma}}{2}\right)e^{i\tilde{\Delta}\tau}e^{-(\Gamma-\tilde{\Gamma})\tau/2} + \left(\Delta + \tilde{\Delta} + \dfrac{i(\Gamma-\tilde{\Gamma})}{2}\right)e^{-i(\Delta+\tilde{\Delta})\tau}e^{-\tilde{\Gamma}\tau/2}}{(\Delta + i\Gamma/2) + 2(\tilde{\Delta} - i\tilde{\Gamma}/2)}, \tag{15.173}$$

or noting that $\tilde{\Gamma} \ll \Gamma$ and $\tilde{\Delta} \ll \Delta$,

$$\langle \mathrm{g}|U(\tau,0)|\mathrm{g}\rangle \approx e^{-i(\Delta+\tilde{\Delta})\tau}e^{-\tilde{\Gamma}\tau/2}, \tag{15.174}$$
(weak-pumping survival amplitude)

This expression shows that the survival amplitude rotates at the natural (unperturbed) frequency of Δ, plus an ac Stark shift $\tilde{\Delta}$ due to the pumping laser. There is also the slow decay of $|g\rangle$ at rate $\tilde{\Gamma}$. Note that in the full expression (15.173) there is also a fast-decaying term, decaying at rate $\Gamma - \tilde{\Gamma}$, and shifted by $-\tilde{\Delta}$ from zero energy. This is because the weak field mixes the ground and excited states slightly, so the part of $|\mathrm{e}\rangle$ mixed into $|\mathrm{g}\rangle$ decays essentially at the decay rate for $|\mathrm{e}\rangle$, and has the opposite Stark shift as expected for a two-level system.

One curious effect is that $\tilde{\Gamma} \longrightarrow 0$ as $\Gamma \longrightarrow \infty$. Since a decay from $|\mathrm{e}\rangle$ to $|\mathrm{f}\rangle$ is a measurement of whether or not the atom is in $|\mathrm{e}\rangle$ (indicated by the detection of an emitted photon), Γ is essentially the rate at which the measurement is taking place. If this measurement is strong enough, the atom can never be promoted from $|\mathrm{g}\rangle$ to $|\mathrm{e}\rangle$ in the first place—an example of the quantum Zeno effect.

15.6.2 Strong Pumping

In the limit of strong pumping ($\Omega \gg \Gamma$), the eigenvalues/poles from Eq. (15.164) become

$$E_\pm \approx \frac{1}{2}\left[E_\mathrm{e} + E_\mathrm{g} - \frac{i\hbar\Gamma}{2} \pm \hbar\tilde{\Omega}\right] = \frac{\hbar}{2}\left[\Delta - \frac{i\Gamma}{2} \pm \tilde{\Omega}\right], \tag{15.175}$$

where

$$\tilde{\Omega} := \sqrt{\Omega^2 + \Delta^2} \tag{15.176}$$

is the usual generalized Rabi frequency. Then the survival amplitude (15.167) becomes

$$\langle \mathrm{g}|U(\tau,0)|\mathrm{g}\rangle = \frac{1}{\tilde{\Omega}}\left[\left(\Delta+\tilde{\Omega}\right)e^{-i\Delta\tau/2}e^{-\Gamma\tau/4}e^{-i\tilde{\Omega}\tau} - \left(\Delta-\tilde{\Omega}\right)e^{-i\Delta\tau/2}e^{-\Gamma\tau/4}e^{i\tilde{\Omega}\tau}\right], \tag{15.177}$$

or

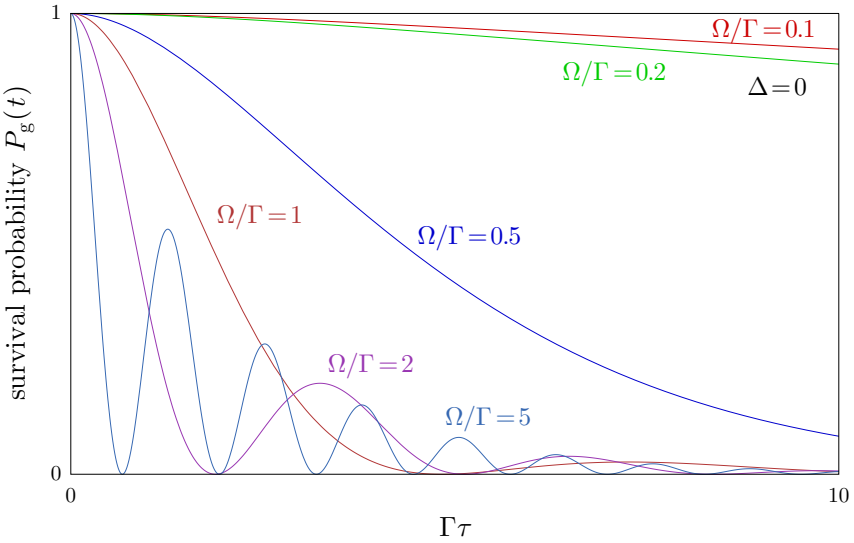
$$\langle \mathrm{g}|U(\tau,0)|\mathrm{g}\rangle = e^{-i\Delta\tau/2}e^{-\Gamma\tau/4}\left[\cos\frac{\tilde{\Omega}\tau}{2} - i\frac{\Delta}{\tilde{\Omega}}\sin\frac{\tilde{\Omega}\tau}{2}\right].$$

(strong-pumping survival amplitude) (15.178)

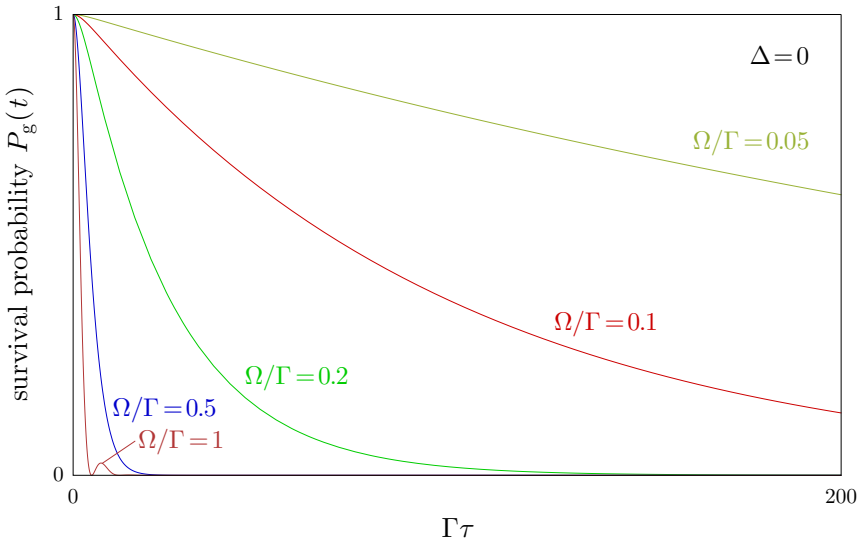
These are the usual generalized Rabi oscillations [cf. Eq. (5.59)], noting the sign difference in Δ], but now damped at rate $\Gamma/2$. Here the field mixes $|\mathrm{g}\rangle$ and $|\mathrm{e}\rangle$ together in equal parts, so $|\mathrm{g}\rangle$ decays at half the decay rate of $|\mathrm{e}\rangle$.

15.6.3 General Case

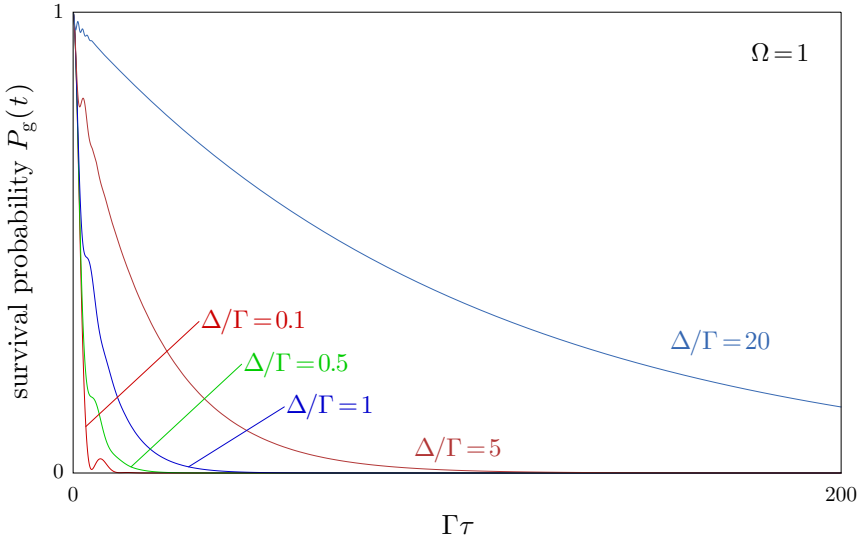
The general case interpolates between simple exponential decay and damped Rabi oscillations in a reasonable obvious way, as shown here for the on-resonance case $\Delta = 0$.



Even for relatively weak pumping $\Omega/\Gamma = 0.1$, when the decay is essentially exponential, the one obvious feature is the nonexponential decay at short times, since the whole process must start via a part of a Rabi oscillation from $|\mathrm{g}\rangle$ to $|\mathrm{e}\rangle$. Of course, we already know that the decay must be nonexponential at short times in any case (Section 11.7.1).



For the off-resonance case (with $\Omega = 1$), the Rabi oscillations are incomplete, and become more rapid, since the oscillations occur around the generalized Rabi frequency. Obviously, the decay becomes slower for larger detunings, but also note that the fast oscillations damp out before a smooth decay takes over.



15.7 Exercises

Problem 15.1

Show that the resolvent operator

$$G(z) := \frac{1}{z - H} \tag{15.179}$$

for the Hamiltonian H is analytic off the real axis, in the sense that every matrix element $\langle\psi|G(z)|\psi'\rangle$ for arbitrary states $|\psi\rangle, |\psi'\rangle$ is an analytic function anywhere away from the real axis. State explicitly your criteria for analyticity.

Problem 15.2

The inhomogeneous Helmholtz equation

$$\left(\nabla^2 + k^2\right)\psi(\mathbf{r}) = -f(\mathbf{r}), \tag{15.180}$$

where $f(\mathbf{r})$ is an arbitrary source function, has a Green function (resolvent) defined by [see Eq. (14.57), noting that we are ditching the ϵ_0 but keeping the minus sign]

$$\left(-\nabla^2 - k^2\right)G(\mathbf{r}, \mathbf{r}'; k^2) = \delta^d(\mathbf{r} - \mathbf{r}') \tag{15.181}$$

in d spatial dimensions.

(a) If we assume the Helmholtz equation to be defined on a compact domain, show that the retarded "energy-space" Green function $G(\mathbf{r}, \mathbf{r}'; k^2)$ can be written in the form[7]

$$G^+(\mathbf{r}, \mathbf{r}'; k^2) = \sum_n \frac{\psi_n(\mathbf{r})\psi_n^*(\mathbf{r}')}{k_n^2 - k^2 - i0^+}, \tag{15.182}$$

where $\psi_n(x)$ are the eigenfunctions of the homogeneous version of Eq. (15.180) with (discrete) eigenvalues $k = k_n$. Be careful with the sign of the imaginary deformation here!

(b) Show that in the continuum limit where $k_n \longrightarrow p$, the (retarded) Green function may be written as

$$G^+(\mathbf{r}, \mathbf{r}'; k^2) = \frac{1}{(2\pi)^d}\int d^d p\, \frac{\psi_p(\mathbf{r})\psi_p^*(\mathbf{r}')}{p^2 - k^2 - i0^+} \tag{15.183}$$

in d spatial dimensions.

Problem 15.3

Derive the asymptotic expansion

$$E_1(z) = \frac{e^{-z}}{z}\left\{1 - \frac{1}{z} + \frac{2}{z^2} - \frac{3!}{z^3} + \cdots + \frac{n!}{(-z)^n} + \cdots\right\}. \tag{15.184}$$

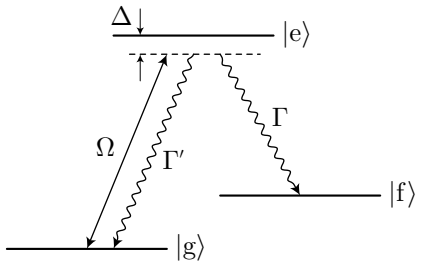
Problem 15.4

Work out a formula for the inverse Laplace transform, using the integral formula for the propagator in terms of the resolvent operator. State any restrictions on the validity of your formula.

Problem 15.5

In analyzing the spontaneous Raman problem, we ignored any decay back to the initial (ground) state $|\mathrm{g}\rangle$. Suppose we modify the setup to explicitly include a decay rate of Γ' from $|\mathrm{e}\rangle \longrightarrow |\mathrm{g}\rangle$.

[7] see, e.g., Marco Schäfer, Idrish Huet, and Holger Gies, "Energy-momentum tensors with worldline numerics," *International Journal of Modern Physics Conference Series* **14**, 511 (2012) (doi: 10.1142/S2010194512007647), arXiv.org preprint (arXiv: quant-ph/0605180v3).



(a) Why is the resolvent method not a natural approach to handle this new problem?

(b) Derive a corrected formula for the decay rate of $|g\rangle$ in the weak pumping limit, accounting for the new decay path. *Hint:* set up and solve Einstein-type rate equations for this system, generalizing the results from the resolvent approach as appropriate (e.g., introducing an auxiliary decay path). You need not retrace the derivation using the resolvent method if you can just indicate the appropriate changes.

Problem 15.6

Consider an atom at a fixed location in an optical cavity. The optical cavity is initially in the vacuum state, and its resonance frequency ω does not necessarily coincide with the atomic resonance frequency ω_0. The atom starts in the excited state.

(a) Compute the decay rate for the atom, assuming the "bad-cavity" limit of large κ. Ignore decay into non-cavity modes. *Hint:* what is the level structure of this problem?

(b) The enhancement of the atomic spontaneous emission rate by a cavity is called the **Purcell effect**. What is now known as the **Purcell factor** was given by Purcell[8] as

$$\eta_{\mathrm{P}} = \frac{3Q\lambda^3}{4\pi^2 V}, \tag{15.185}$$

where Q is the quality factor of the cavity, λ is the emission wavelength, and V is the cavity volume. Purcell's result was that multiplying the atomic decay rate by this factor gives the cavity-modified decay rate. Show that your result is consistent with Purcell's for a cavity whose resonance matches that of the atom, under the assumption that the atomic dipole is aligned with the cavity-mode polarization ($\hat{\varepsilon} \cdot \mathbf{d}_{\mathrm{ge}} = d_{\mathrm{ge}}$, *without* the factor of $\sqrt{3}$).

Problem 15.7

Suppose the intensity of an optical cavity of resonant frequency ω decays exponentially at rate κ. The cavity spectrum is bounded from below, and thus should decay nonexponentially at long times. For example, given that the cavity begins with exactly one photon, the photon's survival probability should become nonexponential at long times.

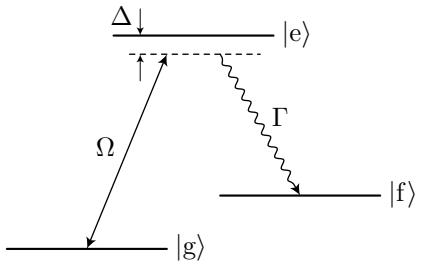
(a) Treating the cavity decay rate as approximately independent of frequency, give an expression for the asymptotic survival probability for long times.

(b) Estimate the scaled time κt of crossover to nonexponential decay for a linear, two-mirror cavity of length 10 cm, assuming identical mirrors with 99% intensity reflection coefficients and a resonance wavelength of 532 nm. Also, estimate the survival probability at this crossover time.

Problem 15.8

Consider the spontaneous-Raman problem, for which we derived the survival probability of $|g\rangle$ in Section 15.6.

[8] E. M. Purcell, "Spontaneous Emission Probabilities at Radio Frequencies," *Physical Review* **69**, 681 (1946) (doi: 10.1103/PhysRev.69.674.2).



Under the condition of weak excitation (small Ω or large $|\Delta|$), derive an expression for the spectral lineshape of the emitted light (assuming the long-time limit). Interpret your solution.

Chapter 16

Photodetection

In Chapter 2, we considered the coherence of classical light, and, for example, what this meant for the classical Hanbury-Brown–Twiss experiment. We will now reconsider this experiment using the quantum description of the fields, and examine the most dramatic departures from the classical-field predictions, which occur for fields with one or two photons.

16.1 Counting Photons

Let's start by developing a bit of formalism to handle detectors interacting with the quantum electromagnetic field.[1] Recall that the quantized electric field has the form [Eq. (8.61)]

$$\mathbf{E}(\mathbf{r},t) = -\sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}} \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) a_{\mathbf{k},\zeta}(t) + \text{H.c.}, \tag{16.1}$$

or identifying the two terms with $E^{(+)}$ and $E^{(-)}$, we can isolate the annihilation component of the field:

$$\mathbf{E}^{(+)}(\mathbf{r},t) = -\sum_{\mathbf{k},\zeta} \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}} \mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r}) a_{\mathbf{k},\zeta}(t). \tag{16.2}$$

We may regard a photodetection event, as a *transition* in the state of the field, $|\text{i}\rangle \longrightarrow |\text{f}\rangle$, where $|\text{i}\rangle$ is the initial state of the field before the detection event, and $|\text{f}\rangle$ is the final state of the system afterward, where one photon in mode $(\mathbf{k},\zeta)$ was removed from the field at time t. The transition amplitude for this process is proportional to

$$\langle \text{f}|U(t+0^+, t-0^+)|\text{i}\rangle \propto \langle \text{f}|a_{\mathbf{k},\zeta}(t)|\text{i}\rangle. \tag{16.3}$$

Including the spatial profile of the field,

$$\langle \text{f}|U(t+0^+, t-0^+)|\text{i}\rangle \propto \langle \text{f}|\hat{\varepsilon}_\zeta \cdot \mathbf{E}^{(+)}_{\mathbf{k},\zeta}(\mathbf{r},t)|\text{i}\rangle, \tag{16.4}$$

where the field here is the single, relevant term in the mode sum (16.2), and $\hat{\varepsilon}_\zeta$ is the unit polarization vector of the mode at the location of the detector. Of course, we can include the entire field,

$$\langle \text{f}|U(t+0^+, t-0^+)|\text{i}\rangle \propto \langle \text{f}|\hat{\varepsilon} \cdot \mathbf{E}^{(+)}(\mathbf{r},t)|\text{i}\rangle, \tag{16.5}$$

since in view of the definition of $|\text{f}\rangle$ only one of the field annihilation operators contributes to the matrix element. We have also dropped the subscript on the polarization vector, so that $\hat{\varepsilon}$ represents the field polarization sensed by the detector. Now the probability for detecting a photon is given by summing the

[1] Here we are following Roy J. Glauber, "The Quantum Theory of Optical Coherence," *Physical Review* **130**, 2529 (1963) (doi: 10.1103/PhysRev.130.2529).

squares of the amplitudes (each amplitude corresponds to a final state where a photon is lost from a particular mode; each mode is orthogonal and thus each final state is distinguishable):

$$\begin{aligned} P(t) &\propto \sum_{\mathrm{f}} |\langle \mathrm{f}|\hat{\varepsilon}\cdot\mathbf{E}^{(+)}(\mathbf{r},t)|\mathrm{i}\rangle|^2 \\ &= \sum_{\mathrm{f}} \langle \mathrm{i}|\hat{\varepsilon}^*\cdot\mathbf{E}^{(-)}|\mathrm{f}\rangle\langle \mathrm{f}|\hat{\varepsilon}\cdot\mathbf{E}^{(+)}|\mathrm{i}\rangle \\ &= \langle \mathrm{i}|\hat{\varepsilon}^*\cdot\mathbf{E}^{(-)}\hat{\varepsilon}\cdot\mathbf{E}^{(+)}|\mathrm{i}\rangle \end{aligned} \tag{16.6}$$

The initial state $|\mathrm{i}\rangle$ is arbitrary, and we can think of the expectation value even for a mixed state by performing an ensemble average over initial states. We can also carry out a sum over polarizations, so that

$$P(t) \propto \left\langle E_\alpha^{(-)}(\mathbf{r},t)E_\alpha^{(+)}(\mathbf{r},t)\right\rangle, \tag{16.7}$$
(photodetection probability)

with an implied sum over α.

This expression for the photodetection probability motivates the definition of a field correlation function, the **degree of first-order coherence**:

$$G^{(1)}(\mathbf{r}_1,t_1,\mathbf{r}_2,t_2) := \left\langle E_\alpha^{(-)}(\mathbf{r}_1,t_1)E_\alpha^{(+)}(\mathbf{r}_2,t_2)\right\rangle,$$
(degree of first-order coherence) (16.8)

which is the quantum analog of the classical field correlation function, which we saw in normalized form in Eq. (2.23), which gives the fringe visibility in an interference experiment. Note the particular ordering of the field operators in the correlation function and the detection probability, which is of the general form

$$\left\langle a^\dagger a\right\rangle. \tag{16.9}$$

This ordering is called **normal ordering**, which refers to having all annihilation operators to the right of all creation operators. This particular ordering is important, as in the vacuum state, the expectation value $\langle 0|a^\dagger a|0\rangle = 0$ gives a zero detection probability or zero correlation, both of which are appropriate for the vacuum. The other ordering here would correspond to detection of photons from the vacuum, which is physically nonsensical.

Similarly, the *joint* probability amplitude to detect one photon at $(\mathbf{r},t)$ and $(\mathbf{r}',t')$ is

$$\langle \mathrm{f}|U(t+0^+,t-0^+)|\mathrm{i}\rangle \propto \langle \mathrm{f}|\hat{\varepsilon}_\zeta\cdot\mathbf{E}^{(+)}_{\mathbf{k},\zeta}(\mathbf{r}',t')\,\hat{\varepsilon}_\zeta\cdot\mathbf{E}^{(+)}_{\mathbf{k},\zeta}(\mathbf{r},t)|\mathrm{i}\rangle. \tag{16.10}$$

To compute the detection probability, we square this, sum over all final states, and consider any polarization as before, with the result

$$P(t) \propto \left\langle E_\alpha^{(-)}(\mathbf{r},t)E_\alpha^{(-)}(\mathbf{r}',t')E_\alpha^{(+)}(\mathbf{r}',t')E_\alpha^{(+)}(\mathbf{r},t)\right\rangle,$$
(joint photodetection probability) (16.11)

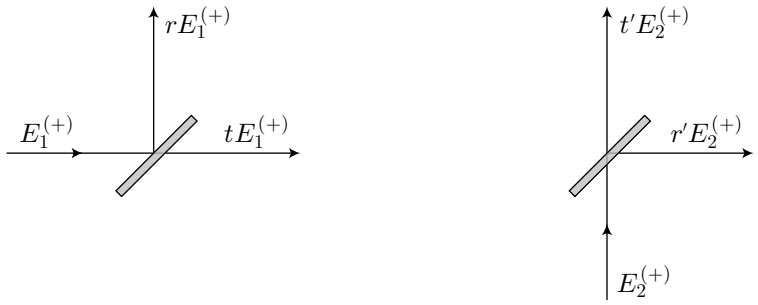
again with an implied sum over α. This joint detection probability motivates the definition of a higher-order correlation function, the **degree of second-order coherence**:

$$G^{(2)}(\mathbf{r}_1,t_1,\mathbf{r}_2,t_2) := \left\langle E_\alpha^{(-)}(\mathbf{r}_1,t_1)E_\alpha^{(-)}(\mathbf{r}_2,t_2)E_\alpha^{(+)}(\mathbf{r}_2,t_2)E_\alpha^{(+)}(\mathbf{r}_1,t_1)\right\rangle.$$
(degree of second-order coherence) (16.12)

This is the quantum analog of the classical *intensity* correlation function, e.g., $\langle I(t)I(t+\tau)\rangle$, which we saw in normalized form in Eq. (2.68). Note that these joint expectation values are still in normal form with the general form $\left\langle a_1^\dagger a_2^\dagger a_2 a_1\right\rangle$, so that either joint expectation value vanishes unless there are at least two photons around somewhere to be detected. We are also ignoring some subtleties regarding the two field annihilation operators; recall that $E_\alpha^{(+)}(\mathbf{r}_1,t_1)$ and $E_\alpha^{(+)}(\mathbf{r}_2,t_2)$ commute only if the two respective spacetime points lie on the same light cone. In practice this does not matter, since for example the two detectors will monitor different outputs of a beam splitter, such that they cannot influence each other.

16.2 Beam Splitters

For our purposes, a beam splitter is an optical element that transforms two input modes or ''beams'' into two output modes. We have treated the beam splitter before in Section 12.1.2, but as it is central to our discussion here, we will review the setup with a slightly different notation. Labeling the field at the first and second inputs as $E_1^{(+)}$ and $E_2^{(+)}$, respectively, the transformation properties are characterized by field reflection and transmission coefficients r and t, representing reflection and transmission of $E_1^{(+)}$, and coefficients r' and t', representing reflection and transmission of $E_2^{(+)}$, as shown here.



Assuming the beam splitter is lossless, it must induce a unitary transformation on the two input modes, which we can represent by the matrix

$$U = \begin{bmatrix} t & r' \\ r & t' \end{bmatrix}, \tag{16.13}$$

in the sense that the output modes are related to the input modes by this operator:

$$\begin{bmatrix} E_{\text{out},1} \\ E_{\text{out},2} \end{bmatrix} = U \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}. \tag{16.14}$$

However, the fact that U is unitary constrains its form; in fact the general form for a 2×2 unitary matrix is

$$U = \begin{bmatrix} t & -r^* \\ r & t^* \end{bmatrix}, \tag{16.15}$$

from which we conclude that

$$|r|^2 + |t|^2 = 1, \tag{16.16}$$

which expresses the lossless property of the beam splitter, and

$$r' = -r^*, \qquad t' = t^*, \tag{16.17}$$

so that the reflection and transmission coefficients from either direction only differ by phases (which we have somewhat arbitrarily fixed here).

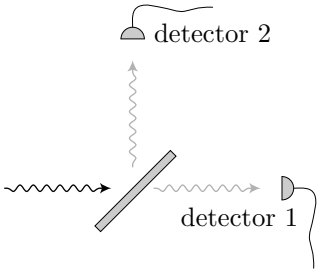
If we consider only monochromatic fields, with inputs and outputs at the same frequency, then from Eq. (16.2) we see that the fields $E^{(+)}$ differs from a lowering operator a only by a constant factor, which is the same for every mode here. Quantum mechanically, then, we may write Eq. (16.14) in terms of operators as

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = U \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} t & -r^* \\ r & t^* \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{16.18}$$
(beam splitter transformation)

where $a_{1,2}$ are the annihilation operators for the input modes, and $b_{1,2}$ are the annihilation operators for the output modes.

16.3 Collision of One Photon and a Beam Splitter

Consider a single photon incident on a beam splitter, where we monitor each output of the beam splitter with a detector.



To the extent that it makes sense to do so, we will consider the input and output fields to be monochromatic as in Eq. (16.18). If we begin with a single photon in input 1, then the initial state is

$$|1,0\rangle = a_1^\dagger |0,0\rangle, \tag{16.19}$$

where the states $|n,m\rangle$ are joint Fock states for the two inputs. If we solve Eq. (16.18) for the input operators in terms of the output operators, we have

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = U^\dagger \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} t^* & r^* \\ -r & t \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \tag{16.20}$$

so that

$$a_1 = t^* b_1 + r^* b_2, \tag{16.21}$$

or

$$a_1^\dagger = t b_1^\dagger + r b_2^\dagger. \tag{16.22}$$

To find the output state after the beam splitter, we can use this relation to eliminate the input-field operator in Eq. (16.19)

$$|\psi\rangle_{\text{out}} = \left(t b_1^\dagger + r b_2^\dagger\right)|0,0\rangle = t|1,0\rangle + r|0,1\rangle.$$

(output state for single-photon input) (16.23)

The output state is thus an entangled state, with a superposition of having a single photon in each mode. The result is now fairly obvious, but from Eq. (16.11), the joint photodetection probability is

$$P(t) \propto \left\langle b_1^\dagger b_2^\dagger b_2 b_1 \right\rangle = 0, \tag{16.24}$$

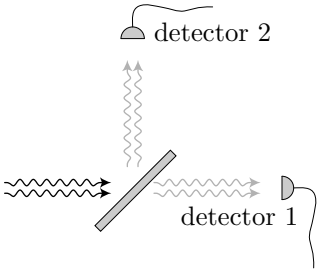
which of course vanishes for the output state (16.23), since there is only one photon. Experimentally, it is difficult to prepare an input state of exactly a single photon. In practice, a highly attenuated classical field (coherent state) is used, which has the form $|0\rangle + \varepsilon|1\rangle$, where $\varepsilon \ll 1$. If this field is an output of a nonlinear crystal, where correlated pairs of photons are generated from a laser field via **parametric downconversion**, then the detection of a photon in the other output can be used to ''herald'' the presence of a single photon in the beam-splitter setup. Thus measurement is used to convert the coherent state into a one-photon state. The real situation is more complicated due to ''accidental'' coincidences (since there is a small probability of having two photon pairs present simultaneously), finite detection times, and ''dark counts,'' or spurious photodetection events due to thermal fluctuations in the detectors. However, this can be done, and is one of the simplest demonstrations of the manifestly quantum-mechanical nature of the electromagnetic field:[2] a *classical* field can be divided arbitrarily, so a classical Hanbury-Brown–Twiss experiment always gives a signal for arbitrarily weak fields. In terms of the normalized degree of second-order coherence, the quantum version of this experiment violates the inequality (2.72), since $g^{(2)}$ can be much smaller than unity.

[2] P. Grangier , G. Roger and A. Aspect, ''Experimental Evidence for a Photon Anticorrelation Effect on a Beam Splitter: A New Light on Single-Photon Interferences,'' *Europhysics Letters* **1**, 173 (1986) (doi: 10.1209/0295-5075/1/4/004); J. J. Thorn, M. S. Neel, V. W. Donato, G. S. Bergreen, R. E. Davies, and M. Beck, ''Observing the quantum behavior of light in an undergraduate laboratory,'' *American Journal of Physics* **72**, 1210 (2004) (doi: 10.1119/1.1737397).

16.4 Two-Photon Interference

16.4.1 Simple Theory

Suppose we now treat the case of *two* incident photons on a beam splitter. Again treating the fields as monochromatic (and identical), we can model a photodetection experiment with this particular input. There are two general cases we can consider. The first is when both photons are incident in the same input—here, input 1.



Then the input state is

$$|2,0\rangle = \frac{(a_1^\dagger)^2}{\sqrt{2}}|0,0\rangle, \tag{16.25}$$

and again using Eq. (16.22) to eliminate the input operator, we find the output state

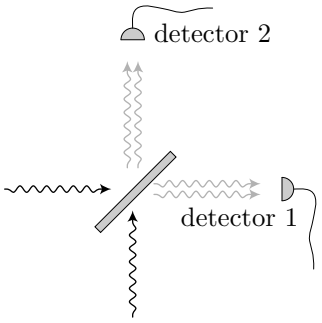
$$|\psi\rangle_{\text{out}} = \frac{\left(tb_1^\dagger + rb_2^\dagger\right)^2}{\sqrt{2}}|0,0\rangle. \tag{16.26}$$

We can write this out to obtain

$$|\psi\rangle_{\text{out}} = t^2|2,0\rangle + \sqrt{2}rt|1,1\rangle + r^2|0,2\rangle. \tag{16.27}$$

This is not too surprising. Identifying the probabilities for two photon transmitted, one photon transmitted and zero photons transmitted as $|t|^4$, $2|rt|^2$, and $|r|^4$, respectively, this is just the *classical* transmission probability of two independent particles according to the binomial distribution, where the "success probability" for a single particle is $|t|^2$.

The other case, where one photon is incident in each input, is quite different, however.



Here, the input state is

$$|1,1\rangle = a_1^\dagger a_2^\dagger|0,0\rangle. \tag{16.28}$$

We can again use Eq. (16.22) to eliminate $a_1^\dagger$, and to eliminate $a_2^\dagger$, we can use Eq. (16.20) to write

$$a_2^\dagger = -r^* b_1^\dagger + t^* b_2^\dagger. \tag{16.29}$$

Thus, the output state is

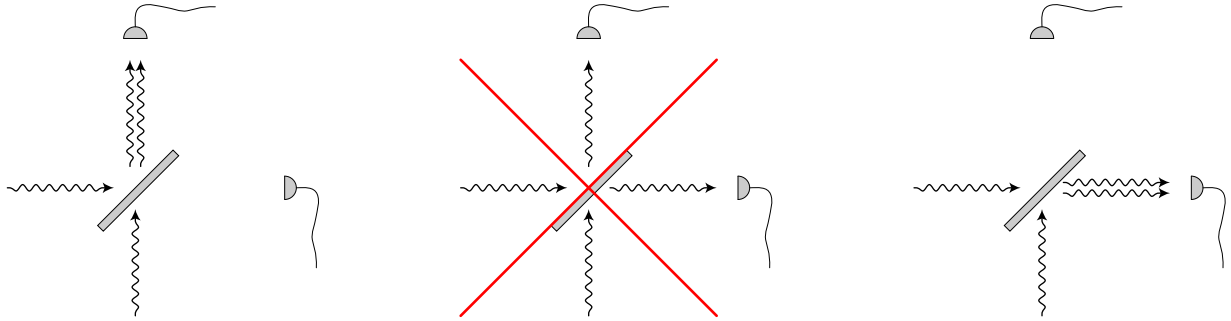
$$|\psi\rangle_{\text{out}} = \left(tb_1^\dagger + rb_2^\dagger\right)\left(-r^* b_1^\dagger + t^* b_2^\dagger\right)|0,0\rangle. \tag{16.30}$$

Multiplying all this out,

$$|\psi\rangle_{\text{out}} = -\sqrt{2}r^*t|2,0\rangle + \left(|t|^2 - |r|^2\right)|1,1\rangle + \sqrt{2}rt^*|0,2\rangle$$

(output state, for one photon in each input) (16.31)

The difference here is that the $|1,1\rangle$ term exhibits destructive interference. The classical probabilities for two photons in output 1 is $|rt|^2$, which is the same as the probability for two photons in output 2; the remaining probability for one photon in each direction is $1 - 2|rt|^2 = |r|^4 + |t|^4$. In both the classical and quantum case, the probability for coincidence detection—corresponding to one photon in each output—is minimized for an equal beam splitter with $|r| = |t| = 1/\sqrt{2}$. In this case, the classical probability is 1/2, while the quantum probability is zero. This is thus a quantum interference effect between the two photons, which rules out the photons leaving the beam splitter in different outputs.



This tendency of the photons to "stick" together is a nice demonstration of the bosonic nature of the quantum electromagnetic field. Had the photons been fermions, the requirement of antisymmetry of the state would have actually produced the *opposite* prediction: coincidences would happen with unit probabilities, since the outcome must always have one photon in each output. The classical prediction is somewhere in between. In this sense, the bosonic case can be regarded as *constructive* interference for the two non-coincidence outcomes, increasing the probability of finding the photons to be together; this is consistent with our discussion of the exchange interaction in Section 4.4.4.1. This effect is known as the **Hong–Ou–Mandel effect**, after the first experimental demonstration.[3]

16.4.2 Coherence Effects

Though the Hong–Ou–Mandel effect is due to interference, it turns out that it does not sensitively depend on the relative phase of the two input photons. That is, varying the relative phase by π does not necessarily cause a large change in the interference effect, as it would in an interferometer. To see this, we must relax the monochromatic idealization of the input light.[4]

16.4.2.1 Quantum Beam

Recall again from Eq. (16.2) that the annihilation part of the electromagnetic field has the form

$$\mathbf{E}^{(+)}(\mathbf{r},t) = -\sum_{\mathbf{k},\zeta}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\epsilon_0}}\mathbf{f}_{\mathbf{k},\zeta}(\mathbf{r})a_{\mathbf{k},\zeta}(t). \tag{16.32}$$

If we consider the light to be in the form of a "beam," as in the output of a laser, then we should regard the wave vector $\mathbf{k}$ to point along a particular direction, say the x-direction. Thus, $k_y = k_z = 0$ and we have only a one-dimensional set of modes. Recall that in calculations with th e three-dimensional field, in the

[3] C. K. Hong, Z. Y. Ou, and L. Mandel, "Measurement of subpicosecond time intervals between two photons by interference," *Physical Review Letters* **59**, 2044 (1987) (doi: 10.1103/PhysRevLett.59.2044).

[4] Here, we are following H. Fearn and R. Loudon, "Theory of two-photon interference," *Journal of the Optical Society of America B* **6**, 917 (1989) (doi: 10.1364/JOSAB.6.000917).

continuum limit we made the replacement

$$\sum_{\mathbf{k}} \longrightarrow \frac{V}{(2\pi)^3} \int_{\text{all space}} d^3k, \tag{16.33}$$

since the spacing between modes in any direction in $\mathbf{k}$-space was $2\pi/L$, where $L^3 = V$. Then the quantization volume V canceled the corresponding factor from the squared mode functions $|f_{\mathbf{k},\zeta}|^2 \propto 1/V$ for the free-sapce modes

$$f_{\mathbf{k},\zeta}(\mathbf{r}) = \hat{\varepsilon}_{\mathbf{k},\zeta} \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{V}}. \tag{16.34}$$

In the one-dimensional case, we will similarly have

$$\sum_{k_x \geq 0} \longrightarrow \frac{L}{2\pi} \int_0^\infty dk, \tag{16.35}$$

taking the beam to point along the positive x-axis. For calculations second order in the field, we can modify the field by changing the sum to an integral, tacking on the square root of the discretization factor $L/2\pi$, and assume a particular polarization along $\hat{\varepsilon}$:

$$\mathbf{E}^{(+)}(\mathbf{r},t) = -\hat{\varepsilon} \int_0^\infty d\omega \sqrt{\frac{\hbar\omega}{4\pi\epsilon_0 cA}}\, a(\omega)\, e^{i(kx-\omega t)}. \tag{16.36}$$
(quantized beam)

Here, we have changed the integration variable to $\omega = \omega_{\mathbf{k}} = ck$, defined the mode area $A = V/L$, and written out explicitly the time dependence of the mode annihilation operator. Recall that in the continuum limit, we have $[a(\omega), a^\dagger(\omega')] = \delta(\omega - \omega')$. If a beam is narrowband, the excitations represented by $a(\omega)$ will be tightly localized near some ''laser frequency'' ω_{L}. Since the factor of $\sqrt{\omega}$ should vary slowly over this spectrum, we can replace it by its value at the laser frequency, so that

$$\mathbf{E}^{(+)}(\mathbf{r},t) = -\hat{\varepsilon} \sqrt{\frac{\hbar\omega}{4\pi\epsilon_0 cA}} \int_0^\infty d\omega\, a(\omega)\, e^{i(kx-\omega t)}.$$
(quantized, quasi-monochromatic beam) (16.37)

We thus have essentially a Fourier transform of the monochromatic mode operators $a(\omega)$.

16.4.2.2 Pulse-Annihilation Operators

Again, the above expression (16.37) shows that the time-dependent electric-field operator for a quasi-monochromatic beam appears as a one-dimensional Fourier transform of the field operators $a(\omega)$. We can take this as a motivation to define the creation operator

$$A^\dagger(\alpha) := \int_0^\infty d\omega\, \alpha(\omega)\, a^\dagger(\omega),$$
(creation operator, pulsed excitation) (16.38)

where $\alpha(\omega)$ represents the spectrum of the excitation, which is normalized according to

$$\int_0^\infty d\omega\, |\alpha(\omega)|^2 = 1. \tag{16.39}$$

Since we are assuming a quasi-monochromatic beam, whose spectral width is much smaller than ω_{L} (as in a laser field), we can extend the lower limit of the integral, so that

$$\int_{-\infty}^\infty d\omega\, |\alpha(\omega)|^2 \approx 1. \tag{16.40}$$

Thus, this creation operator creates a photon similar to $a^\dagger(\omega)$ in the sense that $A^\dagger(\alpha)|0\rangle$ represents a normalized, one-photon state, but in a superposition of different frequencies. Thus, emulating the form of the field operator (16.37), we can define a temporal envelope

$$\alpha(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, \alpha(\omega)\, e^{-i\omega t}, \tag{16.41}$$
(pulse envelope)

and it is not hard to show by direct substitution that

$$\int_{-\infty}^{\infty} dt\, |\alpha(t)|^2 = 1, \tag{16.42}$$

so that the envelope function created by $A^\dagger(\alpha)$ is also normalized.

16.4.2.3 Detection

We can now also replace the full field (16.37) with the normalized, time-dependent annihilation operator

$$a(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\omega\, a(\omega)\, e^{-i\omega t}, \tag{16.43}$$
(pulse-annihilation operator)

within the same narrowband approximation. This is proportional to the full field except that we have dropped the dependence on the spatial coordinate x, since for propagation in vacuum it can be absorbed into the temporal phase factor. We need not assume any particular frequency dependence for the annihilation operator, and in fact we will need this operator for detection. For a wide-bandwidth detector, this operator corresponds to annihilating a photon at the particular time t. (A finite detector bandwidth corresponds to some uncertainty in the time of annihilation.) Then we can use this operator in place of the full field in Eq. (16.7) for the detection probability, and integrate over the detection time interval T to find the total (average) number of detected photons:

$$\langle N \rangle = \int_0^T dt\, \left\langle a^\dagger(t)\, a(t) \right\rangle. \tag{16.44}$$
(mean number of detected photons)

We have replaced the proportionality by an equality here; this expression is scaled properly, as we can see by considering the state of n excitations $|n\rangle$, assuming a sufficiently long detection time T (and assuming the excitations are created after $t = 0$). Similarly, based on Eq. (16.11), we can write down the mean cross-correlation for the photocounts for two detectors:

$$\langle N_1 N_2 \rangle = \int_0^T dt \int_0^T dt' \left\langle a_1^\dagger(t)\, a_2^\dagger(t')\, a_2(t')\, a_1(t) \right\rangle.$$
(joint detection average) (16.45)

This expression is normalized properly as in the average number of detected photons, and for the two-photon input states, that we will consider below, corresponds to the joint detection probability over the (long) detection time.

16.4.2.4 Interference of Coherence

Now back to the problem of two-photon interference. The input mode, now with two (possibly different) quasi-monochromatic photons, is

$$|1,1\rangle = A_1^\dagger(\alpha_1) A_2^\dagger(\alpha_2) |0,0\rangle, \tag{16.46}$$

where the subscripts $A_\beta^\dagger$ label the mode on which the creation operator acts. This expression generalizes the monochromatic expression (16.28). The same beam-splitter-transformation relations (16.22) and (16.29)

hold here (assuming the action of the beam splitter is frequency-independent and nondispersive), so that

$$\begin{aligned} A_1^\dagger(\alpha_1) &= tB_1^\dagger(\alpha_1) + rB_2^\dagger(\alpha_1) \\ A_2^\dagger(\alpha_2) &= -r^* B_1^\dagger(\alpha_2) + t^* B_2^\dagger(\alpha_2). \end{aligned} \tag{16.47}$$

We have thus connected the input operators $A_\beta^\dagger$ to the output operators $B_\beta^\dagger$, which are defined in exactly the same way. We can obtain the output mode by using these relations in the input state (16.46):

$$|\psi\rangle_{\text{out}} = \left[-r^* t B_1^\dagger(\alpha_1) B_1^\dagger(\alpha_2) + |t|^2 B_1^\dagger(\alpha_1) B_2^\dagger(\alpha_2) - |r|^2 B_2^\dagger(\alpha_1) B_1^\dagger(\alpha_2) + r t^* B_2^\dagger(\alpha_1) B_2^\dagger(\alpha_2)\right] |0,0\rangle. \tag{16.48}$$

Only the middle two terms correspond to one output photon in each mode, and thus these will give the only contribution to $\langle N_1 N_2\rangle$.

To compute the detector cross-correlation (16.45), we can simply consider the post-detection state,

$$b_2(t')\, b_1(t) |\psi\rangle_{\text{out}} = b_2(t')\, b_1(t) \left[|t|^2 B_1^\dagger(\alpha_1) B_2^\dagger(\alpha_2) - |r|^2 B_2^\dagger(\alpha_1) B_1^\dagger(\alpha_2)\right] |0,0\rangle, \tag{16.49}$$

where again we need only the middle two terms of Eq. (16.48). In this state, we have parts that refer to either mode; for example, the part of this state that refers to wave packet 1 in mode 1 is

$$\begin{aligned} b_1(t) B_1^\dagger(\alpha_1)|0\rangle &= \frac{1}{\sqrt{2\pi}} \int d\omega \int d\omega'\, b(\omega)\, e^{-i\omega t}\, \alpha_1(\omega')\, b^\dagger(\omega')|0\rangle \\ &= \frac{1}{\sqrt{2\pi}} \int d\omega \int d\omega'\, \delta(\omega - \omega')\, \alpha_1(\omega')\, e^{-i\omega t}\, |0\rangle \\ &= \frac{1}{\sqrt{2\pi}} \int d\omega\, \alpha_1(\omega)\, e^{-i\omega t}\, |0\rangle \\ &= \alpha_1(t)|0\rangle, \end{aligned} \tag{16.50}$$

where in the first step we used Eq. (16.43) for $b_1(t)$ and Eq. (16.38) for $B_1^\dagger(\alpha_1)$; in the second step we used the commutation relation $b(\omega)b^\dagger(\omega')|0\rangle = [b(\omega), b^\dagger(\omega')]|0\rangle = \delta(\omega - \omega')|0\rangle$; and we used Eq. (16.41) for the pulse envelope $\alpha(t)$. The other parts of Eq. (16.49) follow from this result simply by relabeling the arbitrary indices, and finally we may use the norm of the resulting post-detection state to write

$$\begin{aligned} \langle N_1 N_2\rangle &= \int_0^T dt \int_0^T dt' \left\langle b_1^\dagger(t)\, b_2^\dagger(t')\, b_2(t')\, b_1(t)\right\rangle \\ &= \int_0^T dt \int_0^T dt' \left||t|^2 \alpha_1(t)\alpha_2(t') - |r|^2 \alpha_2(t)\alpha_1(t')\right|^2. \end{aligned} \tag{16.51}$$

Multiplying out the square, we obtain the squares of each of the terms in the absolute value, which have time dependence of the form $|\alpha_1(t)|^2|\alpha_2(t')|^2$; due to the normalization of these pulse profiles, the integrals give $|t|^4$ and $|4|^4$ for these two terms. The remaining two cross terms give $-|r|^2|t|^2\alpha_1(t)\alpha_2^*(t)\alpha_1^*(t')\alpha_2(t') + \text{c.c.}$, which when integrated, combine to give

$$-2|r|^2|t|^2 \left|\int dt\, \alpha_2^*(t)\alpha_1(t)\right|^2. \tag{16.52}$$

Combining terms, we finally find

$$\langle N_1 N_2\rangle = |r|^4 + |t|^4 - 2|r|^2|t|^2 \left|\int dt\, \alpha_2^*(t)\alpha_1(t)\right|^2, \tag{16.53}$$

(two-photon cross-correlation)

where the overall result is automatically positive, since the modulus of the integral at at most unity, since the pulse profiles are normalized.

The last term in the cross correlation is the overlap integral of the two input pulses. For identical, perfectly overlapping pulses, the integral reduces to unity, and thus

$$\langle N_1 N_2 \rangle = |r|^4 + |t|^4 - 2|r|^2|t|^2 = \left(|r|^2 - |t|^2\right)^2 , \tag{16.54}$$

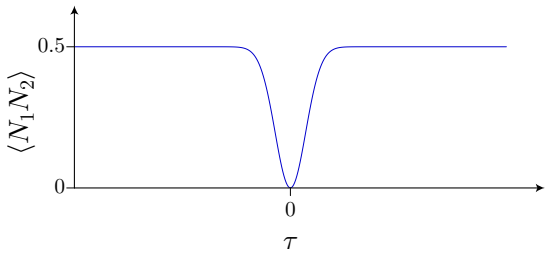
which recovers the simple result (16.31) from the monochromatic theory. If the pulses are widely separated, then the overlap integral vanishes, and we recover the classical expectation

$$\langle N_1 N_2 \rangle = |r|^4 + |t|^4, \tag{16.55}$$

as is appropriate for distinguishable pulses. Finally, if the two input pulses are *identical*, but one is delayed by time τ with respect to the other, then $\alpha_2(t) = \alpha_1(t - \tau)$, and

$$\langle N_1 N_2 \rangle = |r|^4 + |t|^4 - 2|r|^2|t|^2 \left| \int dt\, \alpha^*(t)\alpha(t+\tau) \right|^2 , \tag{16.56}$$

and thus the interference term reduces to the degree of first-order coherence (normalized autocorrelation function) for the input pulse. While this modulates the *fringe visibility* in an interferometer, it represents the entire interference in the Hong–Ou–Mandel experiment. Thus as a function of the time delay of one of the pulses, the coincidence probability exhibits a "dip," konwn as the **Hong–Ou–Mandel dip**, whose profile is the pulse-field autocorrelation function (something like the convolution of the pulse with itself). This is illustrated below for the case of a symmetric beam splitter and a Gaussian pulse envelope.



The width of the dip is of the order of the coherence length, which can be very long for narrow-line lasers, or much shorter for pulsed lasers.

Part III

Stochastic Trajectories in Quantum Mechanics

Chapter 17

Stochastic Processes

Before tackling quantum measurements head-on, we will first examine some of the basic mathematics for handling measurements, and in particular *continuous* quantum measurements. We will thus need to look at some basics in the area of **stochastic processes**—that is, the mathematics for modeling systems as having underlying randomness influencing the dynamics.[1] Historically, the term ''stochastic'' has also been used to refer to low-dimensional, chaotic dynamics of Hamiltonian systems, which is not what we mean here. By *stochastic* we are referring to a truly random element that is not predictable even in principle. This is sensible for modeling quantum measurements, which are considered to be sources of true randomness. However, despite the inherent unpredictability, we can fruitfully model stochastic systems by building on the basic formalism introduced here.

17.1 Finite Random Walks, Diffusion, and the Central Limit Theorem

One central problem in statistical mechanics that is useful in quantum optics—and indeed underlies much of the formalism of quantum measurement that we will develop—is the random-walk process. Suppose a random walker takes a random step of size X with probability density $f(x)$ between periodic intervals of duration Δt. Let's assume that all the steps are statistically independent, and the probability distribution is characterized by

$$\langle X\rangle = 0, \quad \mathrm{Var}[X] = \sigma^2. \tag{17.1}$$

After N steps (N large), where has the walker ended up? The **central limit theorem** says that the probability density of the accumulated displacement

$$S_N := \sum_{j=1}^{N} X_j \tag{17.2}$$

for N steps is Gaussian with zero mean and variance $N\sigma^2$. That is, the width (standard deviation) is $\sigma\sqrt{N}$.[2] The probability distribution thus becomes asymptotically Gaussian with a time-dependent width of

$$\sigma(t) = \sigma\sqrt{\frac{t}{\Delta t}}. \tag{17.3}$$

This random-walk behavior is characteristic of a **diffusion process**, which is a transport process by which the distribution grows as $t^{1/2}$,

$$\Delta x \sim D\, t^{1/2}, \tag{17.4}$$

[1] Note that we will be giving just an introductory overview, and will sacrifice rigor in favor of intuition; a good rigorous introduction is W. Horsthemke and R. Lefever, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology* (Springer, 1984). Another good introduction is the classic C. W. Gardiner, *Handbook of Stochastic Methods* 1st ed. (Springer, 1983).

[2] Recall that the variance of X is defined by $\mathrm{Var}[X] := \left\langle (X - \langle X\rangle)^2 \right\rangle = \langle X^2\rangle - \langle X\rangle^2$, and the standard deviation is the square root of the variance.

where for the random walker the **diffusion coefficient** is $D = \sigma/\sqrt{\Delta t}$. Note that within certain restrictions, the final distribution is Gaussian, *independent of the one-step distribution*.

17.1.1 Two-Step Distribution

Before proving the full central limit theorem, we will examine the probability density after exactly two steps. The mathematical problem is as follows: let X_1 and X_2 be *independent* random variables with probability density functions $f_1(x)$ and $f_2(x)$, respectively. That is, the probability that $X_{1,2}$ is between x and $x + dx$ is $f_{1,2}(x)\,dx$. Then we can ask, what is the probability density of $X_1 + X_2$?

To answer this, we can note that $X_1 + X_2 = x$ for any pair of values of X_1 and X_2 that happen to add up to x. But then we must sum over all such pairs. The probability that both X_1 and X_2 will *both* have particular probabilities is the product of the individual probabilities since the variables are independent. Thus, expressing what we said in equation form,

$$\begin{aligned}&\text{Prob}(X_1 + X_2 \text{ between } x \text{ and } x + dx) = \\ &\qquad \sum_{x',x''} \text{Prob}(X_1 \text{ between } x' \text{ and } x' + dx') \times \text{Prob}(X_2 \text{ between } x'' \text{ and } x'' + dx'' \,|\, x = x' + x'').\end{aligned} \tag{17.5}$$

We can translate this statement in terms of the probability densities and implement the constraint as a δ-function (with a factor of dx, so that the δ-function registers unity when the condition is met). Letting $f_+(x)$ denote the probability density of $X_1 + X_2$,

$$f_+(x)\,dx = \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx''\, f_1(x')\, f_2(x'')\, \delta(x' + x'' - x)\, dx. \tag{17.6}$$

Evaluating the x'' integral, we see that the probability density of the sum is the **convolution** of the individual densities,

$$f_+(x)\,dx = \int_{-\infty}^{\infty} dx'\, f_1(x')\, f_2(x - x')\, dx =: (f_1 * f_2)(x)\, dx, \tag{17.7}$$

where we use the $*$ symbol to denote convolution of two functions. Note that this result is general in that it doesn't assume any particular form for $f_1(x)$ or $f_2(x)$.

For the random walk, we assumed *identical*, indepedent steps, so that $f_1(x) = f_2(x) = f(x)$. Thus, the probability density for two steps is

$$f_{S_2}(x) = (f * f)(x), \tag{17.8}$$
(two-step probability density)

i.e., the convolution of the one-step distribution with itself. Recall that the convolution ''smears'' one function with another, and so as the effect of the second step is to smooth the one-step distribution. The idea behind the central limit theorem is that this smoothing continues until the distribution is Gaussian after many steps.

17.1.1.1 Example 1: Convolution with a Delta Function

As an example of the general idea of the convolution of two functions f and g,

$$(f * g)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, g(x - x'), \tag{17.9}$$

consider the convolution of $f(x)$ with the perfectly localized delta function $g(x) = \delta(x)$. The convolution is then

$$(f * \delta)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, \delta(x - x') = f(x). \tag{17.10}$$

The effect of convolution with a delta function is thus simply to do nothing: convolution with a delta function is just the identity operation.

In terms of the random walk, $\delta(x)$ as a one-step probability function simply corresponds to a step of zero length, or just taking no step at all. Thus, it makes intuitive sense that the distribution isn't changed by convolution with $\delta(x)$. In general, when $g(x)$ is some other function, the convolution "smears" $f(x)$ with the **convolution kernel** $g(x)$. Typically, we will use centered kernels; the effect of a displaced kernel is simply to displace the convolution by the same amount. For example, if

$$g(x) = \delta(x - x_0), \tag{17.11}$$

then

$$(f * g)(x) = \int_{-\infty}^{\infty} dx'\, f(x')\, \delta(x - x_0 - x') = f(x - x_0), \tag{17.12}$$

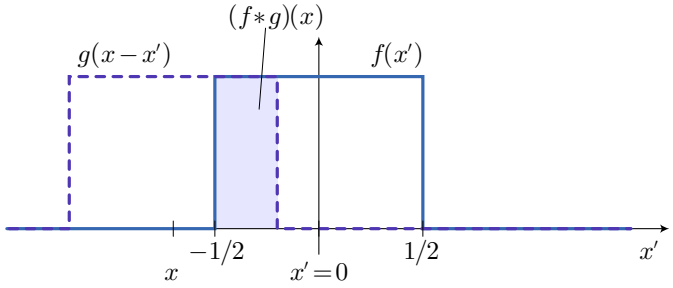
which is just the displaced version of the original.

17.1.1.2 Example 2: Convolution of Box Functions

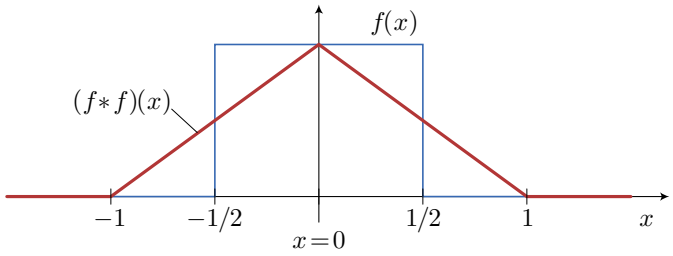
As a slightly more complicated example, consider the convolution of box functions, both given by

$$f(x) = g(x) = \begin{cases} 1, & |x| \le 1/2 \\ 0 & \text{elsewhere,} \end{cases} \tag{17.13}$$

which here are properly normalized to correspond to probability distributions. The convolution consists of displacing $g(x')$ by x, multiplying the functions together, and integrating. For this simple case (box functions of unit height), the convolution (product) just turns out to be the area where the two functions overlap.



When the displacement is large, $|x| > 1$, the boxes don't overlap at all, so the convolution is zero. Otherwise, the overlap area varies linearly with the displacement, so the convolution is a triangle function.



(Note that $f * f$ in the figure is the same as $f * g$ for this special case of $f = g$.) We see now the "smoothing" or "blurring" effect of the convolution. The original functions were discontinuous, but the convolution is continuous. The convolution is also wider than the original functions. As we will see, continued, successive convolutions will make the distribution look Gaussian.

17.1.2 Convolution Theorem

Now that we brought up the convolution, we may as well discuss how to compute it. The **convolution theorem** gives an easy way to evaluate the convolution integral in Eq. (17.7), both in an intuitive and a

computational sense. The convolution theorem states that the Fourier transform of the convolution is the product of the Fourier transforms of the individual functions:

$$\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g]. \tag{17.14}$$
(convolution theorem)

To prove this, we'll just compute the explicit form of $\mathscr{F}[f * g]$. This will be very much a physicist's proof, not a mathematician's proof, in that we'll just assume the functions are nice enough that all the integrals simply exist.

First of all, in our notation here, the Fourier and inverse transforms have the form

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, \tilde{f}(k)\, e^{ikx}, \quad \tilde{f}(k) = \int_{-\infty}^{\infty} dx\, f(x)\, e^{-ikx}, \tag{17.15}$$

where $\tilde{f}(k) \equiv \mathscr{F}[f(x)]$. It's important to make this explicit, since the result depends on the normalization convention we choose for the Fourier transform. Then computing the Fourier transform of $f * g$,

$$\begin{aligned}
\mathscr{F}[f * g] &= \mathscr{F}\left[\int_{-\infty}^{\infty} dx'\, f(x')\, g(x - x')\right] \\
&= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, f(x')\, g(x - x')\, e^{-ikx} \\
&= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx'\, f(x')\, e^{-ikx'} g(x - x')\, e^{-ik(x-x')}.
\end{aligned} \tag{17.16}$$

Letting $x \longrightarrow x + x'$,

$$\begin{aligned}
\mathscr{F}[f * g] &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dx' f(x')\, e^{-ikx'} g(x)\, e^{-ikx} \\
&= \int_{-\infty}^{\infty} dx'\, f(x')\, e^{-ikx'} \int_{-\infty}^{\infty} dx\, g(x)\, e^{-ikx} \\
&= \mathscr{F}[f]\mathscr{F}[g].
\end{aligned} \tag{17.17}$$

Thus, to convolve two functions, just follow this recipe: Fourier transform both functions, multiply them together, then compute the inverse Fourier transform. Mathematically, we can write

$$f * g = \mathscr{F}^{-1}\left\{\mathscr{F}[f]\mathscr{F}[g]\right\}. \tag{17.18}$$

Since Fourier transforms of common function are usually already known, the convolution theorem provides a shortcut for evaluating the full convolution integral.

17.1.2.1 Example: Convolution of Two Gaussians

Since it's easy to compute the Fourier transform of Gaussian distributions, let's use the convolution theorem to convolve two Gaussians. Let's write the two functions as

$$f(x) = Ae^{-x^2/\alpha^2}, \quad g(x) = A'e^{-x^2/\beta^2}. \tag{17.19}$$

The Fourier transform of a Gaussian is also a Gaussian, and in particular

$$\mathscr{F}[f](k) = \tilde{f}(k) = A\alpha\sqrt{\pi}e^{-\alpha^2k^2/4}, \quad \mathscr{F}[g](k) = \tilde{g}(k) = A'\beta\sqrt{\pi}e^{-\beta^2k^2/4}. \tag{17.20}$$

Then the product of the Fourier transforms is

$$\left(\mathscr{F}[f]\mathscr{F}[g]\right)(k) = AA'\alpha\beta\pi e^{-(\alpha^2+\beta^2)k^2/4}. \tag{17.21}$$

Finally, we invert the Fourier transform to obtain the convolution:

$$(f * g)(x) = \mathscr{F}^{-1}\left[AA'\alpha\beta\pi e^{-(\alpha^2+\beta^2)k^2/4}\right] = \frac{AA'\alpha\beta\sqrt{\pi}}{\sqrt{\alpha^2+\beta^2}} \exp\left(-\frac{x^2}{\alpha^2+\beta^2}\right). \tag{17.22}$$

Recall that the standard (normalized) form of the Gaussian is

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{17.23}$$

where the μ is the mean and σ is the standard deviation (σ^2 is the variance). The standard deviation is a common measure of the width of a Gaussian function. Note that $f(x)$ has standard deviation $\alpha/\sqrt{2}$, $g(x)$ has standard deviation $\beta/\sqrt{2}$, and $(f * g)(x)$ has standard deviation $\sqrt{(\alpha^2+\beta^2)/2}$, so that the standard deviations add in quadrature as a result of the convolution. Thus, the convolution of Gaussians is still Gaussian, but the blurring effect of the convolution makes the convolved Gaussian wider than the original functions.

17.1.3 Proof of the Central Limit Theorem

Now we extend the two-step analysis above analysis to N steps. Let $X_1, \ldots, X_N$ be independent, identically distributed random variables. Let $f(x)$ be the probability density function of each of the X_j. Defining the sum by

$$S_N := \sum_{j=1}^{N} X_j, \tag{17.24}$$

we will now ask, what is the probability density $f_{S_N}(x)$ of S_N? Evidently, we can iterate Eq. (17.8) to obtain

$$f_{S_N}(x) = (f * f * \cdots * f)(x), \tag{17.25}$$

where the result is the successive convolution of N copies of f (for $N-1$ total convolution operations). However, it turns out that this distribution becomes simple for large enough N.

The **central limit theorem** states that, provided that the mean and variance of the X_j exist, with the mean $\mu = \langle X_j \rangle$ and variance $\sigma^2 = \mathrm{Var}[X_j]$, the distribution $f_{S_N}(x)$ becomes asymptotically Gaussian for large N with

$$\langle S_N \rangle = N\mu, \quad \mathrm{Var}[S_N] = N\sigma^2. \tag{17.26}$$

(central limit theorem)

(The mean and variance are in fact exact, whereas the form of the distribution is valid for large N.) This is a rough statement, since ''becomes asymptotically Gaussian'' is an imprecise statement. So let's clean this up a bit.

The central limit theorem states that the probability density function $f_{Z_N}(x)$ of the centered, scaled statistic

$$Z_N := \frac{S_N - N\mu}{\sigma\sqrt{N}} \tag{17.27}$$

converges to the ''standard normal'' (Gaussian) distribution

$$f_{Z_N}(x) \longrightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{17.28}$$

which is the special Gaussian with mean 0 and unit variance.

Let's prove this now[3]. To evaluate the convolutions in Eq. (17.25), we need to employ the convolution

[3] This is the physicist's proof; the rigorous version is in T. W. Körner, *Fourier Analysis* (Cambridge, 1988), starting on p. 349.

theorem. Taking the Fourier transform of $f(x)$,

$$\begin{aligned}\tilde{f}(k) &= \int_{-\infty}^{\infty} dx\, f(x)\, e^{-ikx} \\ &= \sum_{j=0}^{\infty} \int_{-\infty}^{\infty} dx\, f(x) \frac{(-ikx)^j}{j!} \\ &= 1 - ik\mu - \frac{k^2(\sigma^2+\mu^2)}{2} + O(k^3).\end{aligned} \tag{17.29}$$

Here, we Taylor-expanded e^{-ikx} and then used the fact that the terms of the expansion were proportional to expectation values $\langle X^j \rangle$. In particular, note that in probability theory the **characteristic function** of a probability density, given by

$$\tilde{f}(-k) = \langle e^{ikX} \rangle, \tag{17.30}$$

is an important tool for manipulating probabilities.

This is more cumbersome than necessary, so let's recompute the expansion in Eq. (17.29) for the centered, scaled variable

$$Z_j = \frac{X_j - \mu}{\sigma\sqrt{N}}, \tag{17.31}$$

with corresponding probability density $f_Z(x)$. The centering effectively zeroes the mean, and the rescaling changes the factor in front of the variance, with the result

$$\tilde{f}_Z(k) = 1 - \frac{k^2}{2N} + O\left[\left(\frac{k}{\sqrt{N}}\right)^3\right]. \tag{17.32}$$

The convolution theorem says that to calculate the transform of the N-fold convolution, we just compute $\tilde{f}_Z(k)$ to the Nth power:

$$\tilde{f}_{Z_N}(k) = \left[\tilde{f}_Z(k)\right]^N = \left(1 - \frac{k^2}{2N} + O\left[\left(\frac{k}{\sqrt{N}}\right)^3\right]\right)^N. \tag{17.33}$$

As N becomes large, we can neglect the higher order terms beyond the first, and then use the formula

$$\lim_{n\longrightarrow\infty} \left(1 + \frac{x}{n}\right)^n = e^x \tag{17.34}$$

to see that for large N, the transform becomes

$$\tilde{f}_{Z_N}(k) = \exp\left(-\frac{k^2}{2}\right). \tag{17.35}$$

But now the inverse Fourier transform of $\exp(-k^2/2)$ is $\exp(-x^2/2)/\sqrt{2\pi}$, so f_{Z_N} converges to a standard normal distribution as $N \longrightarrow \infty$.

17.1.3.1 Example: Square Distribution

As a simple example of the central limit theorem, let's try out the unit box function as the one-step distribution, as we tried out in Section 17.1.1.2:

$$f(x) = \begin{cases} 1, & |x| \le 1/2 \\ 0 & \text{elsewhere.} \end{cases} \tag{17.36}$$

First note that the this function is normalized, so it represents a proper probability distribution. Thus, so do all of its self-convolutions. Let $f^{*N}(x)$ denote the convolution of $f(x)$ with itself $N-1$ times. This is the

same as $f_{S_N}(x)$ for the random-walk interpretation of this distribution. The central limit theorem says that asymptotically, the self-convolution becomes Gaussian,

$$f^{*N}(x) = \frac{1}{\sqrt{2\pi}\sigma_N} e^{-x^2/2\sigma_N^2}, \tag{17.37}$$

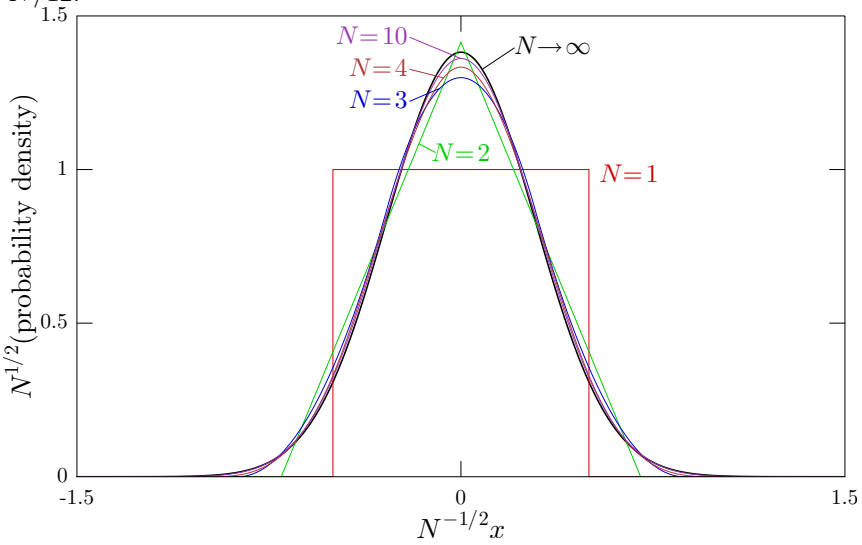
with zero mean, since $f(x)$ is centered. The variance of $f(x)$ is

$$\int_{-\infty}^{\infty} dx\, x^2\, f(x) = \int_{-1/2}^{1/2} x^2\, dx = \frac{1}{12}, \tag{17.38}$$

so that the width of the asymptotic Gaussian is

$$\sigma_N = \sqrt{\frac{N}{12}}. \tag{17.39}$$

Here, $f(x)$ is plotted with several self-convolutions $f^{*N}(x)$, along with the asymptotic form, the Gaussian of width $\sigma_N = \sqrt{N/12}$.



As N increases, the widths of the distributions increase and the peak values decrease, but we have rescaled the axes by appropriate factors of $\sqrt{N}$ to keep the distributions comparable at each step. The box function is very different from the asymptotic Gaussian. However, even the first self-convolution (a triangle function) is already pretty close to the Gaussian, and the successive self-convolutions converge fairly rapidly to the asymptotic form.

17.1.3.2 Application: Standard Deviation of the Mean

Returning again to error analysis, suppose we make independent measurements $X_1, \ldots, X_N$ of some quantity in the laboratory. The **sample mean** is

$$\mu_N := \frac{1}{N} \sum_{j=1}^{N} X_j. \tag{17.40}$$

We can rewrite this as

$$\mu_N = \frac{S_N}{N} = \mu + \frac{\sigma Z_N}{\sqrt{N}}, \tag{17.41}$$

where the first term represents the *true mean*, and the second is the experimental error (statistical fluctuation in the sample mean). Applying the central limit theorem, Z_N is approximately standard normal for large N, so μ_N is Gaussian with mean μ and standard deviation $\sigma/\sqrt{N}$, where σ is the standard deviation of a single measurement. Thus, the **standard deviation of the mean** (also called the **standard error**) is $\sigma/\sqrt{N}$. This is why, by making many measurements, it is possible to increase the accuracy of a measured quantity.

17.1.4 Variances Add in Quadrature

In discussing random walks so far, we have been discussing the asymptotic, N-step probability distribution, and its scaling with time. However, we can make a simpler statement that does not explicitly refer to the distribution. Let $X_1, \ldots, X_N$ be independent random variables, but now we won't even require them to be identically distributed. For the moment, let's also assume $\langle X_n \rangle = 0$. Now consider the sum $X_1 + X_2$. Clearly the mean vanishes, and thus the variance becomes

$$\begin{aligned} \mathrm{Var}[X_1 + X_2] &= \left\langle (X_1 + X_2)^2 \right\rangle \\ &= \left\langle X_1^2 \right\rangle + \left\langle X_2^2 \right\rangle + 2\langle X_1 X_2 \rangle \\ &= \left\langle X_1^2 \right\rangle + \left\langle X_2^2 \right\rangle + 2\langle X_1 \rangle \langle X_2 \rangle \\ &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2], \end{aligned} \tag{17.42}$$

where we used the fact that X_1 and X_2 are independent, and thus their correlation function $\langle X_1 X_2 \rangle$ factorizes into $\langle X_1 \rangle \langle X_2 \rangle$ [the joint probability density $f(x_1, x_2)$ for independent processes must have the factored form $f_1(x_1) f_2(x_2)$]. Thus, the *variances* of independent random variables, and regarding the variance as the square of the ''width'' of the corresponding probability distributions, we see that the widths add *in quadrature* when we add together the random variables. By subtracting $\langle X_1 + X_2 \rangle^2 = \langle X_1 \rangle^2 + \langle X_2 \rangle^2 + 2\langle X_1 \rangle \langle X_2 \rangle$ from each intermediate expression, it isn't hard to see that the same result holds when $\langle X_n \rangle \neq 0$.

Iterating this process, we see that the variance of the sum defined as before,

$$S_N := \sum_{j=1}^{N} X_j, \tag{17.43}$$

is simply

$$\mathrm{Var}[S_N] = \sum_{j=1}^{N} \mathrm{Var}[X_j]. \tag{17.44}$$

Again, if we take each X_n to be identical as for the random walk, and we take the variance as the square of the width σ (i.e., $\mathrm{Var}[X_n] = \sigma^2$), then

$$\mathrm{Var}[S_N] = N\sigma^2, \tag{17.45}$$
(variance of the sum)

or

$$\sigma_N := \sqrt{\mathrm{Var}[S_N]} = \sqrt{N}\,\sigma. \tag{17.46}$$
(standard deviation of the sum)

This is the same as one of the results of the central limit theorem (17.26), but this is not an asymptotic statement, it is *exact*. Thus, we expect the width of the sum to be precisely $\sqrt{N}\,\sigma$. Nevertheless, we often expect this scaling of the distribution width to hold only asymptotically, since in general the ensemble of walkers will have an initial distribution that does not match the one-step distribution (or any N-step distribution), and thus we also need to include the convolution with this initial state.

17.1.5 A Walk on the Cauchy Side

Consider independent, identically distributed random variables $X_1, \ldots, X_N$ with **Cauchy** (Lorentzian) probability density functions

$$f(x) = \frac{1}{\pi(1 + x^2)}. \tag{17.47}$$
(Cauchy distribution)

The Fourier transform is given by

$$\tilde{f}(k) = e^{-|k|}, \tag{17.48}$$

as we can see by computing the *inverse* Fourier transform of $\tilde{f}(k)$:

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-|k|}e^{ikx} \\ &= \frac{1}{2\pi}\left[\int_0^{\infty} dk\, e^{-k(1-ix)} + \int_0^{\infty} dk\, e^{-k(1+ix)}\right] \\ &= \frac{1}{2\pi}\int_0^{\infty} dk\, e^{-k(1-ix)} + \text{c.c.} \\ &= \frac{1}{2\pi(1-ix)} + \text{c.c.} \\ &= \frac{1+ix}{2\pi(1+x^2)} + \text{c.c.} \\ &= \frac{1}{\pi(1+x^2)}. \end{aligned} \tag{17.49}$$

We have now shown that $\mathscr{F}^{-1}[\tilde{f}(k)] = f(x)$. Both $f(x)$ and $\tilde{f}(k)$ are continuous and bounded, so the Fourier transform is invertible; thus, $\mathscr{F}[f(x)] = \tilde{f}(k)$.

Now let's compute the probability density of the mean

$$\mu_N := \frac{1}{N}\sum_{j=1}^{N} X_j. \tag{17.50}$$

The probability density function of the sum

$$S_N := \sum_{j=1}^{N} X_j \tag{17.51}$$

is $(f * f * \ldots * f)(x)$ (N copies of f or $N-1$ convolutions), so using the convolution theorem,

$$\tilde{f}_{S_N}(k) = \left[\tilde{f}(k)\right]^N = \left[e^{-|k|}\right]^N = e^{-|Nk|} = \tilde{f}(Nk). \tag{17.52}$$

In general, if $f(x)$ and $\tilde{f}(k)$ are a Fourier transform pair, then so are $\alpha f(\alpha x)$ and $\tilde{f}(k/\alpha)$. Thus, the inverse transform of $\tilde{f}(Nk)$ is $f(x/N)/N$. The variable μ_N is the same as S_N except for a scaling factor of $1/N$, so the probability density must be the same, but N times wider. So to get the probability density of μ_N, we make the replacement $x \longrightarrow Nx$ in the expression $f_{S_N}(x)\,dx = f(x/N)\,dx/N$ for the probability density of S_N, which gives $f(x)\,dx$. Thus, $f(x)$ is also the probability density of μ_N.

This is different than what we expect from the central limit theorem: there, we expect the mean to have a width that is smaller than that of the one-step distribution by a factor of $\sqrt{N}$. Stated otherwise, what we have shown is that the width of the sum distribution $f_{S_N}(x) = f(x/N)$ is N times larger than that of the one-step distribution, which says that the *widths add* for the Cauchy random walk. But the central limit theorem said that *variances add*, or the widths should add *in quadrature*. Is there a contradiction here?

Obviously there should be some simple resolution. The problem is that variance of X_j does not exist for a Cauchy distribution. This is because the Cauchy distribution only falls off as $1/x^2$ for large $|x|$, and so the variance integral

$$\int_{-\infty}^{\infty} dx\, f(x)\, x^2 \tag{17.53}$$

diverges. The central limit theorem implicitly assumes that the variance exists; thus, the central limit theorem does not apply to this case. This is one case of **anomalous diffusion**, where the diffusion coefficient diverges, because the width of the N-step distribution does not scale diffusively (i.e, it scales as t rather than $\sqrt{t}$).

One important application of such ''fat-tailed'' distributions is in the area of financial modeling,[4] where Gaussian random walks do not adequately model the large jumps observed, for example, in the histories of stock prices.

17.1.6 Arbitrary Combinations of Random Variables

In considering random walks, we have been considering the probability distribution corresponding to the sum (17.2) of independent random variables, in terms of the probability distributions of the separate variables—it turned out to be just the convolution of the individual distributions. But how far can we push this? Here we will develop some concepts seemingly unrelated to probability, and use them to deduce the probability density for an arbitrary (scalar) function of a set of independent random variables.

17.1.6.1 Divergence Theorem

The divergence theorem is fundamental in the study of electrostatics, and the standard form states that for a vector field $\mathbf{A}$,

$$\int_V (\nabla \cdot \mathbf{A})\, d^3r = \oint_S \mathbf{A} \cdot \hat{n}\, dS, \tag{17.54}$$

where V is the volume of integration, S is the surface of the volume, $\hat{n}$ is the (outward-pointing) normal vector to the surface, and dS is the surface-area element for integration over the surface of S.

Let's briefly derive this in d dimensions. Consider a box of infinitesimal volume, given by

$$dV = dx_1 \cdots dx_d. \tag{17.55}$$

The flux of $\mathbf{A}(\mathbf{x})$ through the surface of this volume is given by summing over the fluxes of the two sides bounding each dimension; the ''area'' of the jth side is $dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_d = dV/dx_j$, and thus we have a total flux

$$\begin{aligned} \mathbf{A} \cdot \hat{n}\, dS &= \sum_j \left[\mathbf{A}(\mathbf{x} + dx_j \hat{x}_j) \cdot \hat{x}_j \frac{dV}{dx_j} - \mathbf{A}(\mathbf{x}) \cdot \hat{x}_j \frac{dV}{dx_j} \right] \\ &= \sum_j \frac{\partial \mathbf{A}(\mathbf{x})}{\partial x_j} \cdot \hat{x}_j\, dV \\ &= (\nabla \cdot \mathbf{A})\, dV, \end{aligned} \tag{17.56}$$

where the divergence is interpreted in d dimensions. Now we integrate over the volume, adding up the volume and surface contributions due to all the infinitesimal elements in the integration volume. Whenever two elements contact, their fluxes on their common surface cancel, so the only contribution in the surface integral is due to the flux on the outer surface of the integration volume, and thus we have

$$\int_V d^d x\, (\nabla \cdot \mathbf{A}) = \oint_S \mathbf{A} \cdot \hat{n}\, dS \tag{17.57}$$

(divergence theorem, n dimensions)

as the generalized divergence theorem, converting a volume integral in n dimensions to an integral over the bounding hypersurface, a manifold in $n-1$ dimensions.

17.1.6.2 Transformation of Surface Delta Functions

This divergence theorem is useful in establishing a chain-rule formula for a delta function. Recall that given a function $f(x)$, the delta function obeys the chain rule

$$\delta[f(x-a)] = \sum_{x \in f^{-1}(a)} \frac{\delta(x-a)}{|f'(a)|}, \tag{17.58}$$

[4] Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes* (Chapman & Hall/CRC, 2004), Chapter 1.

which follows from requiring that the delta function be normalized under integration over either $y = f(x)$ or x. For a coordinate change between coordinates $\mathbf{x}$ and $\mathbf{y}$ (both in d dimensions), this formula generalizes to

$$\delta^d(\mathbf{y}) = \frac{\delta^d(\mathbf{x})}{|\partial \mathbf{y}/\partial \mathbf{x}|}, \tag{17.59}$$

so that the scaling factor is now the Jacobian determinant of the coordinate transformation. But what if we have a *scalar* function $y = h(\mathbf{x})$ of a d-dimensional vector $\mathbf{x}$? The appropriate generalization of the determinant turns out to be the Euclidean norm of the vector of partial derivatives of h, as we will now show.

We will start by considering arbitrary scalar functions $h(\mathbf{x})$ and $\mathbf{A}(\mathbf{x})$, respectively, on $\mathbb{R}^d$. Now consider the step function $\Theta(h)$, which defines a volume (or possibly multiple, unconnected volumes). Then consider the integral

$$\int d^d x \, \nabla \cdot \{\Theta[h(\mathbf{x})]\mathbf{A}(\mathbf{x})\} = 0, \tag{17.60}$$

which follows from changing to a surface integral via the divergence theorem, and assuming that either the surface at infinity is outside the volume defined by h (any physically relevant volume should be finite), or that $\mathbf{A}$ vanishes when necessary at infinity. Then using

$$\nabla \cdot [\Theta(h)\mathbf{A}] = \mathbf{A} \cdot \nabla\Theta(h) + \Theta(h)\nabla \cdot \mathbf{A}, \tag{17.61}$$

we have

$$\begin{aligned} \int d^d x \, \Theta(h) \nabla \cdot \mathbf{A} &= -\int d^d x \, \mathbf{A} \cdot \nabla\Theta(h) \\ &= -\int d^d x \, \delta(h) \, \mathbf{A} \cdot \nabla h. \end{aligned} \tag{17.62}$$

Now put

$$f(\mathbf{x}) = \mathbf{A} \cdot \nabla h, \tag{17.63}$$

and consider

$$\begin{aligned} \int d^d x \, \delta[h(\mathbf{x})] \, f(\mathbf{x}) &= -\int d^d x \, \Theta[h(\mathbf{x})] \, \nabla \cdot \mathbf{A}(\mathbf{x}) \\ &= -\int_V d^d x \, (\nabla \cdot \mathbf{A}) \\ &= -\oint_S \mathbf{A} \cdot \hat{n} \, dS \\ &= \oint_{h^{-1}(0)} \mathbf{A} \cdot \frac{\nabla h}{|\nabla h|} \, dS \\ &= \oint_{h^{-1}(0)} \frac{f(\mathbf{x})}{|\nabla h|} \, dS, \end{aligned} \tag{17.64}$$

where we have identified V as the volume defined by $\Theta(h)$; S as the surface, defined by the locus of points where h vanishes; and the normal vector $\hat{n} = -\nabla h/|\nabla h|$, since the gradient is normal to the surface, but points towards the *interior* of the volume (where h is increasing away from the surface). To summarize, we have identified

$$\int d^d x \, \delta[h(\mathbf{x})] \, f(\mathbf{x}) = \oint_{h^{-1}(0)} \frac{f(\mathbf{x})}{|\nabla h|} \, dS, \tag{17.65}$$

(δ-function constraint in integration)

as the effect of a δ-function constraint in a volume integral, where $|\nabla h|$ is the Euclidean norm of the gradient vector:

$$|\nabla h| = \sqrt{\sum_j (\partial_{x_j} h)^2}. \tag{17.66}$$

Again, the constraint here changes the volume integral to an integral over the bounding hypersurface. In this derivation, we have assumed that any function $f(\mathbf{x})$ can be represented as $\mathbf{A}\cdot\nabla h$, but since $\mathbf{A}$ is arbitrary, this is always possible provided $\nabla h \neq 0$ anywhere (and we are free to assume this while still representing any surface manifold we like).

17.1.6.3 Direct Derivation

A more direct derivation of Eq. (17.65) proceeds as follows.[5] Recall that $h(\mathbf{x}) = 0$ defines the surface on which the delta function "fires," and we are considering vector coordinates $\mathbf{x} = (x_1, \ldots, x_d)$. Consider a neighborhood of a point $\mathbf{x}_0$ on the surface, in which one of the partial derivatives is nonvanishing, say $\partial h/\partial x_1 \neq 0$. Then we can consider the action of the δ function along this coordinate, as

$$\int dx_1\, \delta[h(\mathbf{x})]\, f(\mathbf{x}) = \int dx_1\, \frac{\delta(x_1 - x_{01})\, f(\mathbf{x})}{\left|\dfrac{\partial h}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0}} = \frac{f(x_{01}, x_2, \ldots, x_d)}{\left|\dfrac{\partial h}{\partial x_1}\right|_{\mathbf{x}=\mathbf{x}_0}}, \tag{17.67}$$

where $x_{10} = x_{10}(x_2, \ldots, x_d)$ is the x_1-component of $\mathbf{x}_0$, and we have used the ordinary δ-function chain rule (17.58). Then using

$$|\nabla h| = \sqrt{\sum_{j=1}^{d}\left(\frac{\partial h}{\partial x_j}\right)^2} = \left|\frac{\partial h}{\partial x_1}\right| \sqrt{1 + \sum_{j=2}^{d}\left(\frac{\partial x_1}{\partial x_j}\right)^2}, \tag{17.68}$$

and then identifying

$$dS = \sqrt{1 + \sum_{j=2}^{d}\left(\frac{\partial x_1}{\partial x_j}\right)^2}\, dx_2 \cdots dx_d = \sqrt{\sum_{j=1}^{d}\left(\frac{\partial x_1}{\partial x_j}\right)^2}\, dx_2 \cdots dx_d \tag{17.69}$$

as the local surface element (this bears more explanation), and finally integrating over the remaining coordinates [which now locally parameterize the surface $h(\mathbf{x}) = 0$], Eq. (17.65) then results.

Now to return to the business of identifying the surface element (17.69). We can define the surface element in terms of the volume element as

$$dV = dS\, d\ell, \tag{17.70}$$

where $d\ell$ is the line element normal to the surface. Then

$$dS = \frac{dV}{d\ell} = \frac{dx_1 \cdots dx_d}{d\ell}. \tag{17.71}$$

Now suppose that we express $d\ell$ in terms of x_1:

$$d\ell = (\hat{n}\cdot\hat{x}_1)\, dx_1, \tag{17.72}$$

where $\hat{n}$ is normal to the surface, so that $d\ell$ represents only the component of dx_1 normal to the surface. Then dividing the volume by $d\ell$, we only remove the dimension normal to the surface. Using $\hat{n} = \nabla h/|\nabla h|$, we have

$$d\ell = \frac{1}{|\nabla h|}\frac{\partial h}{\partial x_1}\, dx_1 = \frac{dx_1}{\sqrt{1 + \sum_{j=2}^{d}\left(\frac{\partial x_1}{\partial x_j}\right)^2}}, \tag{17.73}$$

after using Eq. (17.68), and then putting this expression into Eq. (17.71), we obtain the surface element (17.69).

[5] This derivation and the connection to the divergence theorem in the previous section are adapted from (the much more terse treatment in) Lars Hörmander, *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis* (Springer-Verlag, 1983), p. 136.

17.1.6.4 Chain Rule for Coordinate Transformations

Using Eq. (17.65), we can also derive a chain rule for integrals of the form of the left-hand side, when we compare equivalent constraints specified by different functions. First, by rescaling the function $f(\mathbf{x})$, we can write

$$\int d^d x\, \delta[h(\mathbf{x})]\, |\nabla h|\, f(\mathbf{x}) = \oint_{h^{-1}(0)} f(\mathbf{x})\, dS(\mathbf{x}). \tag{17.74}$$

Now consider the constraint $h(\mathbf{x})$ and the alternate but equivalent constraint $k(\mathbf{x})$:

$$h(\mathbf{x}) = 0 \iff k(\mathbf{x}) = 0. \tag{17.75}$$

Then the translation of Eq. (17.74) to the alternate constraint function is

$$\int d^d x\, \delta[k(\mathbf{x})]\, |\nabla k|\, f(\mathbf{x}) = \oint_{k^{-1}(0)} f(\mathbf{x})\, dS(\mathbf{x}). \tag{17.76}$$

The right-hand side here is equivalent to that of Eq. (17.74), so eliminating the surface integral, we have

$$\int d^d x\, \delta[h(\mathbf{x})]\, |\nabla h|\, f(\mathbf{x}) = \int d^d x\, \delta[k(\mathbf{x})]\, |\nabla k|\, f(\mathbf{x}) \tag{17.77}$$

This hold for any test function $f(\mathbf{x})$, so

$$\delta[h(\mathbf{x})] = \delta[k(\mathbf{x})]\, \frac{|\nabla k|}{|\nabla h|}.$$

(δ-function chain rule for scalar constraints) (17.78)

Therefore, the transformation just involves the ratio of Euclidean vector-gradient lengths for the two constraint functions, evaluated in each case at the boundaries. These should both exist for a sensible surface-constraint function, since these act to define the normal vector.

17.1.6.5 Probability Density for Combinations of Random Variables

Now we can return to the main point. Suppose we have random variables $X_1, \ldots, X_N$, with probability density function $f_{\mathbf{x}}(x_1, \ldots, x_N) = f_{\mathbf{x}}(\mathbf{x})$. Now suppose we define a combination of the random variables

$$Y = h(X_1, \ldots, X_N). \tag{17.79}$$

Then in the same way as in the result (17.6) for the two-step random walk, we can compute the probability density $f_y(y)$ for Y by integrating over all possible values of $\mathbf{x}$, using a delta function to enforce the relation between the variables:

$$f_y(y) = \int d^N x\, f(\mathbf{x})\, \delta[y - h(\mathbf{x})]. \tag{17.80}$$

Then using Eq. (17.65), we have

$$f_y(y) = \oint_{h(\mathbf{x})=y} \frac{f(\mathbf{x})}{|\nabla h|}\, dS = \oint_{h(\mathbf{x})=y} \frac{f(x_1, \ldots, x_N)}{\sqrt{\sum_{j=1}^{N} \left(\frac{\partial h(\mathbf{x})}{\partial x_j}\right)^2}}\, dS,$$

(transformation law, arbitrary function of random variables) (17.81)

where the integration is over all $\mathbf{x}$ such that $y = h(\mathbf{x})$ [i.e., all possible sets of $(X_1, \ldots, X_N)$ that give the desired value Y], a set that forms a hypersurface S of dimension $N - 1$.

17.1.6.6 Example: Convolution

For example, taking

$$Y = X_1 + X_2, \tag{17.82}$$

we have $|\nabla h| = \sqrt{2}$, taking $x_2 = y - x_1$, and parameterizing the ''surface'' integral with $s = x_1 + x_2$, so that

$$ds = \sqrt{dx_1^2 + dx_2^2} = \sqrt{2}\, dx_1, \tag{17.83}$$

Eq. (17.81) becomes

$$f_y(y) = \int ds\, \frac{f(x_1, y - x_1)}{\sqrt{2}} = \int dx_1\, f_1(x_1)\, f_2(y - x_1), \tag{17.84}$$

where in the last equality we have assumed independence of X_1 and X_2. This result recovers the convolution (17.7).

17.1.6.7 Example: Quotient of Normal Deviates

As a slightly more complicated example, consider two standard-normal deviates X_1 and X_2, with (separable) joint distribution

$$f(x_1, x_2) = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}. \tag{17.85}$$

Now consider the *quotient* $Y = X_1/X_2$ of the two variables, such that the transformation function is

$$h(x_1, x_2) = \frac{x_1}{x_2}. \tag{17.86}$$

Then the gradient norm is

$$|\nabla h| = \sqrt{\frac{1}{x_2^2} + \frac{x_1^2}{x_2^4}} = \frac{\sqrt{x_1^2 + x_2^2}}{x_2^2} = \frac{\sqrt{1 + y^2}}{|x_2|}, \tag{17.87}$$

where we have set $x_1 = y x_2$, taking x_2 as the independent variable. The line element for the ''surface'' integration is

$$ds = \sqrt{dx_1^2 + dx_2^2} = \sqrt{1 + y^2}\, dx_2, \tag{17.88}$$

and thus

$$\begin{aligned} f_y(y) &= \int ds\, \frac{|x_2|\, f(y x_2, x_2)}{\sqrt{1 + y^2}} \\ &= \frac{1}{2\pi} \int dx_2\, |x_2|\, e^{-x_2^2 (1 + y^2)/2} \\ &= \frac{1}{2\pi} \left(\frac{2}{1 + y^2} \right). \end{aligned} \tag{17.89}$$

Then we see that the distribution function for the quotient is

$$f_y(y) = \frac{1}{\pi(1 + y^2)}, \tag{17.90}$$

which is a standard Cauchy distribution.

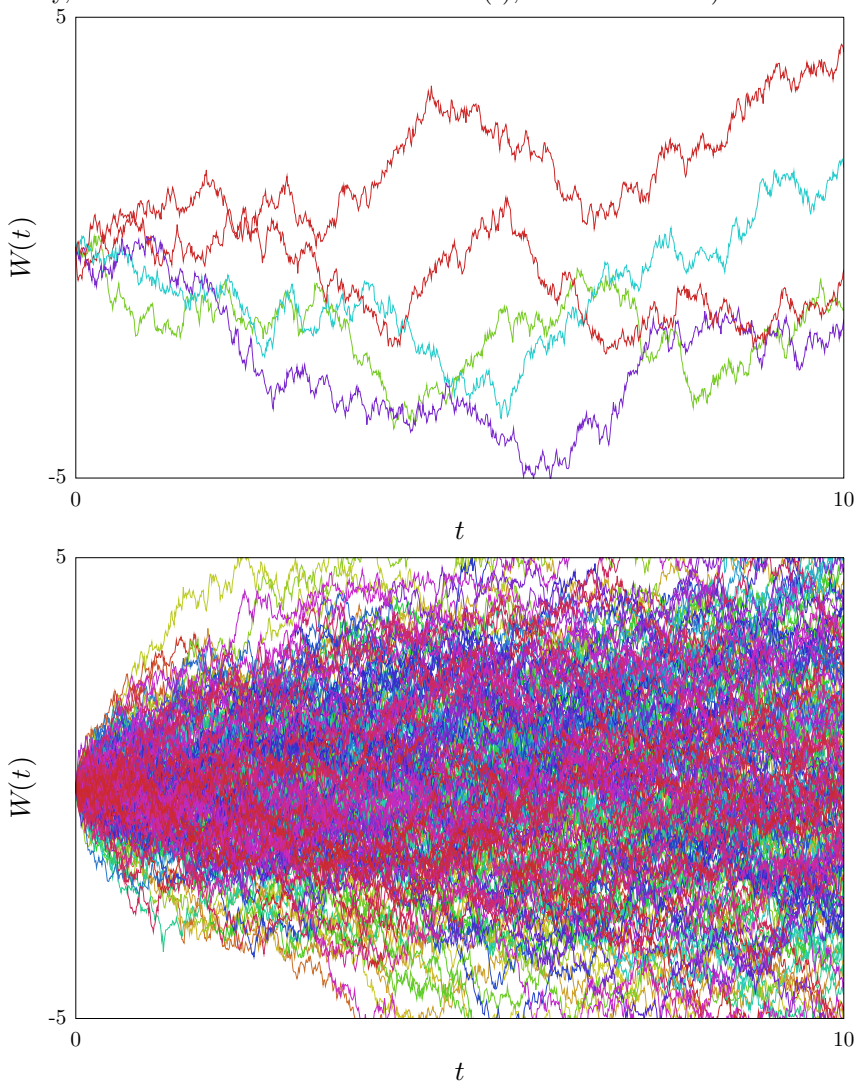
17.2 Continuous Random Walks: Wiener Process

Let's first define the **Wiener process** $W(t)$ as a sort of ''ideal'' random walk with arbitrarily small, independent steps taken arbitrarily often. That is, the usual random walk is usually taken to be a sequence of random steps of finite average (rms) size, taken after every finite time interval Δt. Recall from Section 17.1

that under fairly reasonable assumptions (such as the existence and finiteness of the one-step variance, and independence of the individual steps), that the central limit theorem guarantees that for long times, the probability density for the walker's location is Gaussian, *independent of the one-step distribution*, and the width (standard deviation) increases as $\sqrt{t}$. The Wiener process is essentially the idealized limit where $\Delta t \longrightarrow 0$, but where the size of each step *decreases* as appropriate to maintain the same asymptotic distribution. In this sense, the Wiener process is scale-free, since it has random steps on arbitrarily small time scales, and in fact is a fractal object: a Wiener process with appropriate but arbitrary rescaling (magnification) is still a Wiener process. We choose the Wiener process to correspond to a *symmetric* random walk, and so $W(t)$ is a normally distributed random variable with zero mean. To fix the scale of the random walk, we choose the variance of $W(t)$ to be simply t. That is, the (rms) width of the distribution is $\sqrt{t}$, as is characteristic of a diffusive process. In particular, $W(t)$ has the dimensions of $\sqrt{t}$. We can thus write the probability density for $W(t)$ as

$$P(W,t) = \frac{1}{\sqrt{2\pi t}} e^{-W^2/2t}. \tag{17.91}$$

Note that we have taken the convention that $W(0) = 0$, so that $P(W,0) = \delta(W)$. Again, it is important to emphasize that in view of the central-limit theorem, *any* simple random walk gives rise to a Wiener process in the continuous limit, independent of the one-step probability distribution (so long as the one-step variance is finite). To get an idea what these look like, 5 and 200 Wiener processes are respectively shown in the two plots below. (Actually, these are finite realizations of $\Delta W(t)$, with $\Delta t = 0.01$.)



Intuitively, $W(t)$ is a function that is continuous but everywhere nondifferentiable. (Of course, any such statement necessarily includes the proviso that the statement is true except for possibly a set of realizations of zero measure.) Naturally, the first thing we will want to do is to develop the analogue of the derivative for the Wiener process. We can start by defining the **Wiener increment**

$$\Delta W(t) := W(t + \Delta t) - W(t) \tag{17.92}$$

corresponding to a time interval Δt. Again, ΔW is a normally distributed random variable with zero mean and variance Δt. Note again that this implies that the rms amplitude of ΔW scales as $\sqrt{\Delta t}$. We can understand this intuitively since it is the *variances* that add for successive steps in a random walk, not the standard deviations. Mathematically, we can write the variance as

$$\langle\langle (\Delta W)^2 \rangle\rangle = \Delta t, \tag{17.93}$$

where the double angle brackets $\langle\langle\ \rangle\rangle$ denote an ensemble average over all possible realizations of the Wiener process. This relation suggests the notion that *second*-order terms in ΔW contribute at the same level as *first*-order terms in Δt, thinking about both of these variables in a small-time expansion of the evolution. In the infinitesimal limit of $\Delta t \longrightarrow 0$, we will write $\Delta t \longrightarrow dt$ and $\Delta W \longrightarrow dW$. Then $dW(t)$ is the **Wiener differential**, which is a fundamental object underlying stochastic calculus, and our analogue of the derivative of the Wiener process. Thought of as a ''signal,'' it is everywhere discontinuous. Notice that it is somewhat unusual: one sometimes writes a noise process as

$$\xi(t) := \frac{dW(t)}{dt}, \tag{17.94}$$

but this object is singular (i.e., has unbounded variance at any given time t), because as $\Delta t \longrightarrow 0$,

$$\frac{\Delta W}{\Delta t} \sim \frac{\sqrt{\Delta t}}{\Delta t} = \frac{1}{\sqrt{\Delta t}} \longrightarrow \infty. \tag{17.95}$$

It is possible to work with this singular fraction so long as you are careful with it, in the same sense that you can work with the singular delta function. We will tend to stick to the notation of differentials dt and $dW(t)$, but note that while dW is ''zero,'' it is not ''quite as small as'' dt.

There is, in fact, a deeper connection of dW with the delta function. If we think of dW as a temporal ''noisy'' signal, the reason dW/dt is singular is that it contains contributions from *all frequencies with equal weights*—it's **white noise**—and that's the reason why the Wiener process contains random steps on all time scales. However, the total power for such a system, if the power in any band is finite, must be infinite. This is consistent with the fact that dW/dt diverges on average. On the other hand, if we do anything that *limits the bandwidth* of this signal, such as convolution with a finite function, or using a bandpass filter, it makes the resulting signal finite and well-behaved. Of course, any physical calculation or physical system involves just such a procedure, say, via dissipation through friction. This is exactly analogous to the delta function. The difference is that in the delta function, the frequency components have well-defined relative phases, while for the Wiener process they have effectively random phases.

17.3 Itō Calculus

Now that we have introduced a white-noise process, we will explore the formalism for handling this, particularly for handling the evolution of systems that are *driven* by white noise. It turns out that adding a white-noise stochastic process changes the basic structure of the calculus for treating the evolution equations. In particular, the usual Riemann integral is undefined for stochastic processes. There is more than one formulation to treat stochastic processes, but we will start out with **Itō calculus**,[6] which is the one most commonly used in treating quantum systems. We will start by showing how to *use* this calculus, since the rules are a bit different than what you're probably used to, and then we will justify the rules of usage.

[6]The name Itō is also commonly transliterated as Ito or Itô.

17.3.1 Usage

First, let's review the usual calculus in a slightly different way. A differential equation

$$\frac{dy}{dt} = \alpha(y,t) \tag{17.96}$$

can be instead written in terms of differentials as

$$dy = \alpha\, dt. \tag{17.97}$$

The basic rule in the familiar *deterministic* calculus is that $(dt)^2 = 0$. To see what we mean by this, we can try calculating the differential dz for the variable $z = e^y$ in terms of the differential for dy as follows:

$$dz = e^{y+dy} - e^y = z\left(e^{\alpha\, dt} - 1\right). \tag{17.98}$$

Expanding the exponential and applying the rule $(dt)^2 = 0$, we find

$$dz = z\alpha\, dt. \tag{17.99}$$

This is, of course, the same result as that obtained by using the chain rule to calculate dz/dy and multiplying through by dy. The point here is that calculus breaks up functions and considers their values within short intervals Δt. In the infinitesimal limit, the quadratic and higher order terms in Δt end up being too small to contribute.

In Itō calculus, we have an additional differential element dW representing white noise. The basic rule of Itō calculus is that $dW^2 = dt$, while $dt^2 = dt\, dW = 0$. We will justify this later, but to use this calculus, we simply note that we ''count'' the increment dW as if it were equivalent to $\sqrt{dt}$ in deciding what orders to keep in series expansions of functions of dt and dW. As an example, consider the stochastic differential equation (SDE)

$$dy = \alpha(y,t)\, dt + \beta(y,t)\, dW. \tag{17.100}$$

We obtain the corresponding differential equation for $z = e^y$ by expanding to *second* order in dy:

$$dz = e^y\left(e^{dy} - 1\right) = z\left(dy + \frac{(dy)^2}{2}\right). \tag{17.101}$$

Only the dW component contributes to the quadratic term; the result is

$$dz = z\left(\alpha + \frac{\beta^2}{2}\right) dt + z\beta\, dW. \tag{17.102}$$

The extra β^2 term is crucial in understanding many phenomena that arise in continuous-measurement processes.

17.3.2 Itō Rule: Justification

We now want to show that the Wiener differential dW satisfies the Itō rule $dW^2 = dt$. We already noted above that by definition, the *ensemble average* of $(\Delta W)^2$ is equal to Δt. However, in the infinitesimal limit, we will show that $dW^2 = dt$ holds *without* the ensemble average. This is surprising, since dW is a stochastic quantity, while dt obviously is not. To show this, consider the probability density function for $(\Delta W)^2$, which we can obtain by a transforming the Gaussian probability density for ΔW:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-(\Delta W)^2/2\Delta t}. \tag{17.103}$$

We accomplish this for the coordinate transformation $y = f(x)$ by the transformation

$$P_y(y)\, dy = \sum_{x \in f^{-1}(y)} P_x(x)\, dx, \tag{17.104}$$

which is also equivalent to the **Frobenius–Peron equation**

$$P_y(y) = \int dx'\, P_x(x')\delta[y - f(x')]. \tag{17.105}$$

Then we may write

$$P\left[(\Delta W)^2\right] = \frac{e^{-(\Delta W)^2/2\Delta t}}{\sqrt{2\pi\,\Delta t\,(\Delta W)^2}}. \tag{17.106}$$

In particular, the mean and variance of this distribution for $(\Delta W)^2$ are

$$\left\langle\!\left\langle (\Delta W)^2 \right\rangle\!\right\rangle = \Delta t \tag{17.107}$$

and

$$\mathrm{Var}\left[(\Delta W)^2\right] = 2(\Delta t)^2, \tag{17.108}$$

respectively. To examine the continuum limit, we will sum the Wiener increments over N intervals of duration $\Delta t_N = t/N$ between 0 and t. The corresponding Wiener increments are

$$\Delta W_n := W[(n+1)\Delta t_N] - W(n\Delta t_N). \tag{17.109}$$

Now consider the sum of the squared increments

$$\sum_{n=0}^{N-1} (\Delta W_n)^2, \tag{17.110}$$

which corresponds to a random walk of N steps, where a single step has average value t/N and variance $2t^2/N^2$. According to the central limit theorem, for large N the sum (17.110) is a Gaussian random variable with mean t and variance $2t^2/N$. In the limit $N \longrightarrow \infty$, the variance of the sum vanishes, and the sum becomes t with certainty. Symbolically, we can write

$$\int_0^t [dW(t')]^2 := \lim_{N\to\infty} \sum_{n=0}^{N-1} (\Delta W_n)^2 = t = \int_0^t dt'. \tag{17.111}$$

For this to hold over any interval $(0, t)$, we must make the formal identification

$$dt = dW^2. \tag{17.112}$$
(Itō rule)

This means that even though dW is a random variable, dW^2 is not, since it has no variance when integrated over any finite interval. Incidentally, we can also write down a similar expression for dW with itself, but at *different* times. The basis of this relation is the observation that $\langle\!\langle \Delta W(t)\, \Delta W(t') \rangle\!\rangle = 0$ for time increments $\Delta t < |t - t'|$, since the Wiener increments are independent. By a similar argument to the one above, the variance vanishes in the continuum limit—the variance of $\Delta W(t)\, \Delta W(t')$ is bounded above by the variance of $[\Delta W(t)]^2$—and thus it also follows that $dW(t)\, dW(t') = 0$ with certainty for $t \neq t'$, and thus we need not have an explicit ensemble average when replacing this product by zero.

17.3.3 Ensemble Averages

Finally, we need to justify a relation useful for averaging over noise realizations, namely that

$$\left\langle\!\left\langle y\, dW \right\rangle\!\right\rangle = 0 \tag{17.113}$$
(Itō ensemble average)

for a solution $y(t)$ of Eq. (17.100). This makes it particularly easy to compute averages of functions of $y(t)$ over all possible realizations of a Wiener process, since we can simply set $dW = 0$, even when it is multiplied

by y. We can see this as follows. Clearly, $\langle\langle dW \rangle\rangle = 0$. Also, Eq. (17.100) is the continuum limit of the discrete relation

$$y(t + \Delta t) = y(t) + \alpha \Delta t + \beta \Delta W(t). \tag{17.114}$$

This discrete form here turns out to be the defining feature of Itō calculus, as we will see. Thus, $y(t)$ depends on $\Delta W(t - \Delta t)$, but is independent of $dW(t)$, which gives the desired result, Eq. (17.113). This gives the important feature of Itō calculus that makes it useful for computing ensemble averages: at a given time, the state of the noise process and the state of the system are independent. In particular, it is simple to write down an equation for the ensemble average of Eq. (17.100),

$$d\langle\langle y(t) \rangle\rangle = \langle\langle \alpha(y, t) \rangle\rangle \, dt, \tag{17.115}$$

which we obtain simply by setting $dW \longrightarrow 0$ in the SDE.

This leads us to some common terminology. Any process $y(t)$ satisfying an SDE of the form of Eq. (17.100) with no deterministic term ($\alpha = 0$) satisfies

$$d\langle\langle y(t) \rangle\rangle = \langle\langle y(t + dt) - y(t) \rangle\rangle = 0. \tag{17.116}$$

Any process satisfying this average condition is called a **martingale**, and is special in that each step in time is *unbiased* as a random walk.

17.3.4 Correlation Function

Now we are in a position to justify the ''whiteness'' of the noise. Recalling the singular noise signal

$$\xi(t) = \frac{dW(t)}{dt}, \tag{17.117}$$

let's compute the ensemble average

$$\langle\langle \xi(t) \, \xi(t') \rangle\rangle, \tag{17.118}$$

which is just the correlation function of the noise signal. Note that the *ensemble* average here can just as well be replaced by a *time* average, to get the correlation function in the time-averaged sense. If $t \neq t'$, we can simply write

$$\langle\langle \xi(t) \, \xi(t') \rangle\rangle = \left\langle\left\langle \frac{dW(t) \, dW(t')}{dt \, dt'} \right\rangle\right\rangle = \frac{\langle\langle dW(t) \rangle\rangle \langle\langle dW(t') \rangle\rangle}{dt \, dt'} = 0 \qquad (t \neq t'), \tag{17.119}$$

since in this case $dW(t)$ and $dW(t')$ are statistically independent. However, if $t = t'$, then

$$\langle\langle \xi(t) \, \xi(t) \rangle\rangle = \frac{\langle\langle (dW)^2 \rangle\rangle}{(dt)^2} = \frac{1}{dt} \longrightarrow \infty. \tag{17.120}$$

Thus we see the divergent behavior. In fact, we can get the normalization from

$$\int dt \, \langle\langle \xi(t) \, \xi(t') \rangle\rangle = 1, \tag{17.121}$$

since there is a contribution of $dt \cdot (1/dt) = 1$ from the integration point $t = t'$, and no contributions from any other point in the integration range. Thus, we can infer that $\xi(t)$ is **delta-correlated**:

$$\langle\langle \xi(t) \, \xi(t') \rangle\rangle = \delta(t - t'). \tag{17.122}$$

(white-noise correlation function)

This justifies the notion of dW [equivalently, $\xi(t)$] as representing *white* noise, since the power spectrum is the Fourier transform of the correlation function according to the Wiener–Khinchin theorem, which in this case turns out to be a constant function over all frequencies. Note also the peculiarity that everything in this derivation carries through *without* the ensemble average, so that in Itō calculus, $\xi(t)\xi(t') = \delta(t - t')$.

17.3.5 Diffusion

The noise term in the Itō stochastic differential equation (SDE)

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW \tag{17.123}$$

causes, as you might expect, *diffusion* of the trajectories $y(t)$. To see this, we need the evolution of the *width* of the ensemble. Using the Itō rule

$$d(y^2) = 2y\,dy + (dy)^2 = (2\alpha y + \beta^2)\,dt + 2\beta y\,dW, \tag{17.124}$$

we find the mean-square trajectory

$$d\langle\!\langle y^2\rangle\!\rangle = \big\langle\!\big\langle (2\alpha y + \beta^2)\,dt + 2\beta y\,dW\big\rangle\!\big\rangle = \big\langle\!\big\langle 2\alpha y + \beta^2\big\rangle\!\big\rangle dt. \tag{17.125}$$

Then defining the ensemble variance by

$$V_y := \big\langle\!\big\langle (y - \langle\!\langle y\rangle\!\rangle)^2\big\rangle\!\big\rangle = \langle\!\langle y^2\rangle\!\rangle - \langle\!\langle y\rangle\!\rangle^2, \tag{17.126}$$

we can use

$$\begin{aligned} d\langle\!\langle y\rangle\!\rangle &= \langle\!\langle \alpha\rangle\!\rangle dt \\ d\left[\langle\!\langle y\rangle\!\rangle^2\right] &= 2\langle\!\langle y\rangle\!\rangle d\langle\!\langle y\rangle\!\rangle = 2\langle\!\langle y\rangle\!\rangle\langle\!\langle \alpha\rangle\!\rangle dt, \end{aligned} \tag{17.127}$$

to write the variance evolution as

$$\begin{aligned} dV_y &= d\langle\!\langle y^2\rangle\!\rangle - d\big[\langle\!\langle y\rangle\!\rangle^2\big] \\ &= \Big[2\Big(\langle\!\langle \alpha y\rangle\!\rangle - \langle\!\langle \alpha\rangle\!\rangle\langle\!\langle y\rangle\!\rangle\Big) + \langle\!\langle \beta^2\rangle\!\rangle\Big]\,dt. \end{aligned} \tag{17.128}$$

(SDE variance evolution)

Thus, the variance is affected by gradients of α with y, or "spatial" dependence of the drift coefficient that can stretch or compact the distribution. This is the deterministic component of the variance. The noise part of the equation also contributes the β^2 term, so that the noise always tends to increase the ensemble variance, thus causing diffusion.

17.3.5.1 Fokker–Planck Equation

The evolution of the mean (17.127) and variance (17.128) are equivalent to the mean and variance according to the *deterministic* Fokker–Planck equation for the probability density $f(y,t)$ (Problem 5.18)

$$\partial_t f(y,t) = -\partial_y \alpha(y,t) f(y,t) + \frac{1}{2}\partial_y^2 \beta^2(y,t) f(y,t).$$

(equivalent Fokker–Planck equation) (17.129)

In fact, this Fokker–Planck equation turns out to be the correct one to evolve the ensemble density. Recall that the standard form for the Fokker–Planck equation in one dimension is [from Section (5.8.6.1)]

$$\partial_t P(y,t) = -\partial_y A(y,t) P(y,t) + \frac{1}{2}\partial_y^2 D(y,t) P(y,t),$$

(general Fokker–Planck equation) (17.130)

where $A(y,t)$ is the **drift coefficient**, and $D(y,t)$ is the **diffusion coefficient**. Thus, we identify the stochastic drift coefficient $\alpha(y,t)$ with the Fokker–Planck drift $A(y,t)$, while we identify the *squared* stochastic coefficient $\beta^2(y,t)$ with the diffusion coefficient $D(y,t)$. (For an alternate connection between stochastic trajectories and diffusion-type equations, see Section 17.11.)

To prove this, let's review a couple of concepts regarding probability theory. The **conditional probability density** $P(y,t|y_0,t_0)$, is a probability density in y, with $P(y,t|y_0,t_0)\,dy$ representing the probability density for finding the particle between y and $y+dy$ at time t, *given* the particle was at y_0 at time t_0.

This is distinct from the **joint density** $P(y,t;y_0,t_0)$, which is a probability density in both y and y_0, where $P(y,t;y_0,t_0)\,dy\,dy_0$ is the probability for finding the particle between y and $y+dy$ at time t *and* between y_0 and y_0+dy_0 at time t_0. The individual probability densities are given by integrating out the other variable,

$$P(y,t)=\int dy_0\,P(y,t;y_0,t_0),\qquad P(y_0,t)=\int dy\,P(y,t;y_0,t_0). \tag{17.131}$$

The joint and conditional densities are related by the conditional probability relation, which states that the probability for A *and* B to occur is the product of the probability for A *given* that B occured and the probability for B to occur:

$$P(y,t;y_0,t_0)=P(y,t|y_0,t_0)P(y_0,t_0). \tag{17.132}$$

For **Markovian** evolution, that is, evolution where the entire state of the system for *all future* times is determined by the state of the system at the *present* time, this means that $P(y,t)$ is determined by $P(y_0,t_0)$. In this case, the conditional density satisfies the **Chapman–Kolmogorov equation**,

$$P(y,t|y_0,t_0)=\int dy'\,P(y,t|y',t')P(y',t'|y_0,t_0)$$

(Chapman–Kolmogorov equation) (17.133)

which certainly seems a reasonable property of the conditional density: two steps of the evolution of the density may be composed into a single step by integrating over all possible intermediate values.

Now to derive the Fokker–Planck equation.[7] To do this, consider the evolution of the ensemble average of an arbitrary function $g[y(t)]$, where $y(t)$ is a solution to the SDE (17.123):

$$\begin{aligned} d\langle\!\langle g(y)\rangle\!\rangle &= \left\langle\!\!\left\langle g'(y)\,dy+\frac{1}{2}g''(y)\,(dy)^2\right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \alpha(y,t)g'(y)\,dt+\frac{\beta^2(y,t)}{2}\,g''(y)\,dt\right\rangle\!\!\right\rangle. \end{aligned} \tag{17.134}$$

We can obviously rewrite this as

$$\partial_t\langle\!\langle g(y)\rangle\!\rangle=\left\langle\!\!\left\langle \alpha(y,t)\,\partial_y g(y)+\frac{\beta^2(y,t)}{2}\,\partial_y^2 g(y)\right\rangle\!\!\right\rangle. \tag{17.135}$$

The operator acting on $g(y)$ on the right-hand side,

$$\mathcal{G}:=\alpha(y,t)\,\partial_y+\frac{\beta^2(y,t)}{2}\,\partial_y^2, \tag{17.136}$$

(generator of the SDE)

is often called the **generator** or **infinitesimal generator** corresponding to the SDE (17.129), because it "generates" the average change in a function of y over an infinitesimal time step dt.

Now let us write out the ensemble average explicitly, using the conditional density $P(y,t|y_0,t_0)$ for $y(t)$:

$$\int dy\,g(y)\,\partial_t P(y,t|y_0,t_0)=\int dy\,P(y,t|y_0,t_0)\left[\alpha(y,t)\,\partial_y g(y)+\frac{\beta^2(y,t)}{2}\,\partial_y^2 g(y)\right]. \tag{17.137}$$

Integrating by parts and discarding boundary terms,

$$\int dy\,g(y)\,\partial_t P(y,t|y_0,t_0)=\int dy\,g(y)\left[-\partial_y\alpha(y,t)P(y,t|y_0,t_0)+\frac{1}{2}\partial_y^2\beta^2(y,t)P(y,t|y_0,t_0)\right]. \tag{17.138}$$

Since $g(y)$ is arbitrary, we may equate the integrands, and thus $P(y,t|y_0,t_0)$ satisfies an equation with the form of the Fokker–Planck equation:

$$\partial_t P(y,t|y_0,t_0)=-\partial_y\alpha(y,t)P(y,t|y_0,t_0)+\frac{1}{2}\partial_y^2\beta^2(y,t)P(y,t|y_0,t_0).$$

(Kolmogorov forward equation) (17.139)

[7] As in C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer, 2004) (ISBN: 9783540707127), Section 4.3.5, p, 93.

This equation is called the **Kolmogorov forward equation**, from which the Fokker–Planck equation (17.130) follows by multiplying through by $P(y_0, t_0)$ and integrating over y_0.

The above argument may also be adapted to give an evolution equation in terms of the *initial* time t_0.[8] Then writing

$$\begin{aligned}\partial_{t_0} P(y,t|y_0,t_0) &= \lim_{\delta t\to 0} \frac{P(y,t|y_0,t_0+\delta t) - P(y,t|y_0,t_0)}{\delta t} \\ &= \lim_{\delta t\to 0} \frac{1}{\delta t}\int dy'\, P(y',t_0+\delta t|y_0,t_0)\big[P(y,t|y_0,t_0+\delta t) - P(y,t|y',t_0+\delta t)\big],\end{aligned} \tag{17.140}$$

where the first term takes advantage of $P(y', t_0 + \delta t|y_0, t_0)$ acting as a normalized distribution in y', and the second term uses the Chapman–Kolmogorov equation (17.133). Under the assumption that the distributions $P(y, t|y_0, t_0)$ are continuous functions, the two conditional densities in the difference above can be expanded to lowest order in δt, with first- and higher-order terms not contributing in the limit $\delta t \longrightarrow 0$:

$$\partial_{t_0} P(y,t|y_0,t_0) = \lim_{\delta t\to 0} \frac{1}{\delta t}\int dy'\, P(y',t_0+\delta t|y_0,t_0)\big[P(y,t|y_0,t_0) - P(y,t|y',t_0)\big]. \tag{17.141}$$

Now using the forward equation (17.139) to replace the first conditional distribution,

$$\begin{aligned}&\partial_{t_0} P(y,t|y_0,t_0) \\ &\quad = \int dy'\Big(-\partial_{y'}\alpha(y',t_0)P(y',t_0|y_0,t_0) + \frac{1}{2}\partial_{y'}^2\beta^2(y',t_0)P(y',t_0|y_0,t_0)\Big)\big[P(y,t|y_0,t_0) - P(y,t|y',t_0)\big].\end{aligned} \tag{17.142}$$

Integrating by parts, we have (after discarding surface terms)

$$-\partial_{t_0} P(y,t|y_0,t_0) = \int dy'\, P(y',t_0|y_0,t_0)\Big(\alpha(y',t)\,\partial_{y'}P(y,t|y',t_0) + \frac{1}{2}\beta^2(y',t)\,\partial_{y'}^2 P(y,t|y',t_0)\Big). \tag{17.143}$$

Finally using $P(y', t_0|y_0, t_0) = \delta(y' - y_0)$, we can carry out the remaining integral to obtain

$$\boxed{-\partial_{t_0} P(y,t|y_0,t_0) = \alpha(y_0,t_0)\,\partial_{y_0}P(y,t|y_0,t_0) + \frac{\beta^2(y_0,t_0)}{2}\partial_{y_0}^2 P(y,t|y_0,t_0).}$$

(Kolmogorov backward equation) (17.144)

This peculiar partial differential equation for the *initial* values y_0 and t_0 is called the **Kolmogorov backward equation**. It has a form similar to the Fokker–Planck equation, except for the order of the derivatives and the coefficients, and the minus sign on the time derivative.

17.3.5.2 Multidimensional Fokker–Planck Equation

In multiple dimensions, it is fairly straightforward to generalize the equivalent Fokker–Planck equation (17.129). We can start with a multidimensional generalization of the SDE (17.123),

$$dx_i = \alpha_i(\mathbf{x},t)\,dt + \beta_{ij}(\mathbf{x},t)\,dW_j, \tag{17.145}$$

where repeated indices are summed. In this case, the equivalent Fokker–Planck equation becomes (Problem 17.2)

$$\boxed{\partial_t f(\mathbf{x},t) = -\partial_i\alpha_i(\mathbf{x},t)f(\mathbf{x},t) + \frac{1}{2}\partial_i\partial_j D_{ij}(\mathbf{x},t)f(\mathbf{x},t),}$$

(equivalent Fokker–Planck equation) (17.146)

where the diffusion tensor is

$$\boxed{D_{ij} := \beta_{ik}\beta_{jk} = \beta_{ik}\beta^{\mathsf{T}}_{kj} = (\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}})_{ij}}$$

(equivalent Fokker–Planck equation) (17.147)

in terms of the noise-coupling matrix $\boldsymbol{\beta}$.

[8] The beginning of this argument is as in C. Gardiner, *op. cit.*, Section 3.6, p. 55.

17.3.6 Ornstein–Uhlenbeck Process

As an example of using Itō calculus, we will consider the **Ornstein–Uhlenbeck process**, which we can define as the solution of the damped equation driven by a Wiener process:

$$dy = -\gamma y\, dt + dW. \tag{17.148}$$
(Ornstein–Uhlenbeck process)

This equation is the **Langevin equation**. To solve this, we write the equation in the form

$$d\left(ye^{\gamma t}\right) = e^{\gamma t} dW, \tag{17.149}$$

which we can integrate to obtain

$$y(t)e^{\gamma t} - y_0 = \int_0^t e^{\gamma t'} dW(t'), \tag{17.150}$$

or simply

$$y(t) = y_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} dW(t'). \tag{17.151}$$

The first term is clearly a decaying transient due to the initial condition, while the second is a convolution of the Wiener process with an exponential kernel, effectively smoothing the white noise. Note that since the Wiener differentials are Gaussian random variables, we see from this that the Ornstein–Uhlenbeck process is the sum over Gaussian random variables and is thus itself Gaussian. It is thus sufficient to completely characterize it by computing the mean and autocorrelation function. The mean is simply given by a decaying transient induced by the initial condition,

$$\left\langle\!\left\langle y(t) \right\rangle\!\right\rangle = y_0 e^{-\gamma t}, \tag{17.152}$$

since in Itō calculus we compute ensemble averages by setting $dW = 0$. The correlation function is given by (taking $t' > t$)

$$\begin{aligned}
\left\langle\!\left\langle y(t)\, y(t') \right\rangle\!\right\rangle &= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds \int_0^{t'} ds'\, e^{-\gamma(t-s)} e^{-\gamma(t'-s')} \left\langle\!\left\langle \frac{dW(s)}{ds} \frac{dW(s')}{ds'} \right\rangle\!\right\rangle \\
&= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds \int_0^{t'} ds'\, e^{-\gamma(t-s)} e^{-\gamma(t'-s')} \delta(s-s') \\
&= y_0^2 e^{-\gamma(t+t')} + \int_0^t ds\, e^{-\gamma(t-s)} e^{-\gamma(t'-s)} \\
&= y_0^2 e^{-\gamma(t+t')} + e^{-\gamma(t+t')} \int_0^t ds\, e^{-2\gamma s} \\
&= \left(y_0^2 - \frac{1}{2\gamma}\right) e^{-\gamma(t+t')} + \frac{1}{2\gamma} e^{-\gamma(t'-t)}.
\end{aligned} \tag{17.153}$$

If we regard y_0 as an belonging to an ensemble of variance $1/2\gamma$, or if we consider the limit $t \longrightarrow \infty$ (with $t - t'$ fixed), we can ignore the transient part and take the correlation function to be

$$\left\langle\!\left\langle y(t)\, y(t') \right\rangle\!\right\rangle = \frac{1}{2\gamma} e^{-\gamma|t-t'|}, \tag{17.154}$$
(Ornstein–Uhlenbeck correlation)

where we have use the fact that for the real correlation function, $\left\langle\!\left\langle y(t)\, y(t') \right\rangle\!\right\rangle = \left\langle\!\left\langle y(t')\, y(t) \right\rangle\!\right\rangle$. The Ornstein–Uhlenbeck process, although corresponding to Gaussian noise, does thus not have independent increments at different times. We see that the damping introduces a ‘‘memory’’ in the dynamics. Further, since the correlation function is exponential, we immediately see that the power spectral density for the Ornstein–Uhlenbeck process is Lorentzian, and thus decays asymptotically as ω^{-2}, and thus corresponds to ‘‘$1/f^2$’’ noise. But also notice that the correlation function is finite for all times, and thus the damping, which has cut off the high frequencies, has made the noise bounded and well-behaved.

17.3.6.1 Brownian Motion

The Ornstein–Uhlenbeck process is a model for Brownian motion, if we use it to model the velocity of a particle subject to friction and to frequent "kicks," as from collisions with many background-gas atoms:

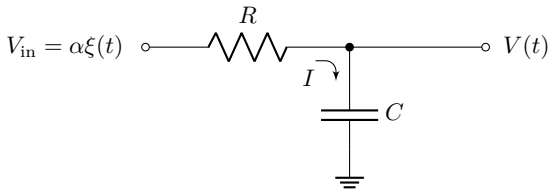
$$dv = -\gamma v\, dt + dW. \tag{17.155}$$

We are assuming that we are dealing with a scaled velocity such that the units come out right. Again, we can equivalently write this in the (possibly) more familiar form

$$\partial_t v = -\gamma v + \xi(t), \tag{17.156}$$

so that we have the usual equation of motion for the velocity, but driven by a white noise "force" $\xi(t)$.

The Ornstein–Uhlenbeck process also corresponds to a white-noise voltage signal $\alpha\xi(t)$ passing through a low-pass filter, with no connection at the output $V(t)$.



To see this, note that the current I flowing through the resistor is

$$I(t) = \frac{\alpha\xi(t) - V(t)}{R}, \tag{17.157}$$

and the voltage across the capacitor is related to the current by

$$\partial_t V = \frac{I}{C} = \frac{\alpha\xi(t) - V(t)}{RC}, \tag{17.158}$$

which we can write as

$$\partial_t V = -\gamma V + \alpha\gamma\xi(t), \tag{17.159}$$

where $\gamma = 1/RC$, as we expect for the low-pass filter. Thus, the output of the low-pass filter corresponds to a scaled Ornstein–Uhlenbeck process, and thus a physical signal despite the idealized input.

Note that from Eq. (17.154), an Ornstein–Uhlenbeck process of the form

$$dy = -\gamma y\, dt + dW \tag{17.160}$$

has $\langle\langle y^2 \rangle\rangle = 1/2\gamma$, and thus an rms fluctuation

$$y_{\text{rms}} = \frac{1}{\sqrt{2\gamma}}. \tag{17.161}$$

If we write Eq. (17.159) in the form

$$dV = -\gamma V\, dt + \alpha\gamma\, dW, \tag{17.162}$$

and then we let $t \longrightarrow t/(\alpha\gamma)^2$ and $W \longrightarrow W/\alpha\gamma$, we have

$$dV = -\frac{1}{\alpha^2\gamma} V\, dt + dW, \tag{17.163}$$

which is in the form of (17.160). Thus, the rms output voltage of the low-pass filter is

$$V_{\text{rms}} = \alpha\sqrt{\frac{\gamma}{2}}, \tag{17.164}$$

(amplitude of filtered white noise)

for an input signal of $\alpha\xi(t)$. Note that the rms output voltage *increases* as the square root of the filter bandwidth ($\gamma/2\pi = 1/2\pi RC$ is the corner frequency of the low-pass filter), and so this result shows that the transmitted *power* is proportional to the width of frequency band passed by the filter. Evidently, α has the dimensions of $\text{V}/\sqrt{\text{Hz}}$.

17.4 Stratonovich Calculus

The main alternative to Itō calculus is **Stratonovich calculus**, which we will introduce primarily to gain more insight into Itō calculus. Although Stratonovich calculus has some aesthetically nice features, it is often easier to perform calculations in Itō form, and we will mainly stick to Itō equations in our discussion of quantum measurement.

Consider the deterministic ODE

$$dy = \alpha(y(t), t)\, dt. \tag{17.165}$$

This is the continuous limit of the discrete relation

$$y(t + \Delta t) = y(t) + \alpha(y(\tau), \tau)\, \Delta t, \tag{17.166}$$

where τ is an *arbitrary* time in the range $[t, t + \Delta t]$. This is because the formal (implicit) solution of (17.165) is given by the Riemann integral

$$y(t) = y_0 + \int_0^t \alpha(y(t'), t')\, dt'. \tag{17.167}$$

The Riemann integral is approximated by successively finer refinements of discrete ''rectangle'' areas of width Δt, where the area of each rectangle is determined by the value of the integrand at *any* point within the interval Δt.

Returning to the SDE

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t)\, dW(t), \tag{17.168}$$

we must also interpret the solution of this equation in terms of the implicit integral

$$y(t) = y_0 + \int_0^t \alpha(y(t'), t')\, dt' + \int_0^t \beta(y(t'), t')\, dW(t'). \tag{17.169}$$

The first integral is an ordinary Riemann integral, but the second is of the form

$$\int_0^t \beta(y(t'), t')\, dW(t') = \int_0^t \beta(y(t'), t')\, \frac{dW}{dt'}\, dt'. \tag{17.170}$$

Due to the highly singular nature of $\xi(t) = dW/dt$, the Riemann integral does not in general exist. The way to save this is that in the successive finite approximations, if you consistently pick the *same* point within each interval of equal length Δt, the integral is defined. However, the *result* that you get by evaluating the integral will depend on your choice. Of course, if β is constant or even a smooth function of time—in the case of **additive noise**—then this won't be a problem, since the result amounts to integrating dW to get $W(t)$. The problem arises in the case of **multiplicative noise** when β is a function of y. Thus, when we regard the SDE (17.168) as the continuum limit of the finite-difference equation

$$y(t + \Delta t) = y(t) + \alpha(y(t), t)\, \Delta t + \beta(y(\tau), \tau)\, \Delta W(t), \tag{17.171}$$

where $\tau \in [t, t + \Delta t]$, we obtain a different limit depending on where in the interval we choose τ. In view of Eq. (17.114), Itō calculus takes the choice $\tau = t$, while Stratonovich calculus takes the choice $\tau = t + \Delta t/2$. Since we expect different results according to what calculus we intend to use to solve the equation, we must use a notation to distinguish the calculus that goes with the SDE. Thus, we will write as usual

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t)\, dW(t) \tag{17.172}$$

(notation: Itō SDE)

to denote an Itō SDE, while we will use the special notation

$$dy = \alpha(y(t), t)\, dt + \beta(y(t), t) \circ dW(t) \tag{17.173}$$

(notation: Stratonovich SDE)

to refer to a Stratonovich equation.

17.4.1 Example: Stochastic Integration

To illustrate the consequences of this choice, we will compute the sample Itō integral

$$I = \int_{t_0}^{t} W(t')\, dW(t'), \tag{17.174}$$

and compare it to the Stratonovich integral

$$J = \int_{t_0}^{t} W(t') \circ dW(t'), \tag{17.175}$$

of the same form, to see that they indeed give different results. Note that if these were ordinary Riemann integrals, we would simply have

$$\int_{t_0}^{t} f(t')\, df(t') = \frac{1}{2}\left[f^2(t) - f^2(t_0)\right] \tag{17.176}$$

for a sufficiently well-behaved function $f(t)$. The Itō integral follows from the continuum limit of the N-step approximation (with $t_j := t_0 + j(t-t_0)/N$)

$$\begin{aligned}
I &= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\Delta W(t_j) \\
&= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\left[W(t_{j+1}) - W(t_j)\right] \\
&= \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W(t_j) + W(t_{j+1}) + W(t_j) - W(t_{j+1})\right]\left[W(t_{j+1}) - W(t_j)\right] \\
&= \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W^2(t_{j+1}) - W^2(t_j)\right] - \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[W(t_{j+1}) - W(t_j)\right]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \lim_{N\to\infty} \frac{1}{2}\sum_{j=0}^{N-1} \left[\Delta W(t_j)\right]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}\int_{t_0}^{t} [dW(t')]^2 \\
&= \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}(t - t_0).
\end{aligned} \tag{17.177}$$

The shortcut for this calculation is to notice that by the Itō rule, $d(W^2) = 2W\, dW + (dW)^2 = 2W\, dW + dt$, so that

$$I = \int_{t_0}^{t} W(t')\, dW(t') = \frac{1}{2}\int_{t_0}^{t} \left[d(W^2) - dt\right] = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] - \frac{1}{2}(t - t_0). \tag{17.178}$$

We thus see how the Itō rule enforces the choice of approximating integration intervals by the *beginning* point of the interval.

In the Stratonovich case, the integrand is evaluated at the intermediate time $t_{j+1/2} := (t_{j+1} + t_j)/2$. However, let us evaluate the most general case $t_{j+s} := (1-s)t_j + s t_{j+1}$, with $s \in [0,1]$, where Itō corresponds

to $s = 0$ and Stratonovich to $s = 1/2$:

$$\begin{aligned} J_s &:= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_{j+s})\Delta W(t_j) \\ &= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_{j+s})\left[W(t_{j+1}) - W(t_j)\right] \\ &= \lim_{N\to\infty} \sum_{j=0}^{N-1} W(t_j)\left[W(t_{j+1}) - W(t_j)\right] + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_j)\right] \\ &= I + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_j)\right] \\ &= I + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]\left[W(t_{j+1}) - W(t_{j+s})\right] + \lim_{N\to\infty} \sum_{j=0}^{N-1} \left[W(t_{j+s}) - W(t_j)\right]^2 \end{aligned} \tag{17.179}$$

The second term corresponds to the continuum limit of a product of independent Wiener increments, which vanishes according to the last argument of Section (17.3.2). The last term is the sum of squared, independent Wiener increments corresponding to time intervals $s\Delta t$, and is thus given by $s(t - t_0)$. Thus,

$$J_s = I + s(t - t_0) = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right] + \left(s - \frac{1}{2}\right)(t - t_0). \tag{17.180}$$

In particular, the Stratonovich integral is

$$J = J_{1/2} = \frac{1}{2}\left[W^2(t) - W^2(t_0)\right]. \tag{17.181}$$

Note that this is exactly the same result as if we had just used *ordinary* calculus, so that in Stratonovich calculus it is appropriate to take $d(W^2) = 2W\,dW$. That is, the usual chain rule applies in Stratonovich calculus, so that $(dW)^2 = 0$. We will prove this after we see how to convert between Itō calculus and Stratonovich calculus.

17.4.2 Itō–Stratonovich Conversion

Itō and Stratonovich SDEs in general give different results for the *same* coefficients, but in what sense are they equivalent? That is, how do we convert between Itō and Stratonovich SDEs? Suppose we have the Itō SDE

$$dy(t) = \alpha(y,t)\,dt + \beta(y,t)\,dW(t), \tag{17.182}$$

and the Stratonovich SDE

$$dy(t) = \bar{\alpha}(y,t)\,dt + \beta(y,t) \circ dW(t). \tag{17.183}$$

Then what we will show is that these two SDEs are equivalent if and only if

$$\bar{\alpha} = \alpha - \frac{1}{2}\beta\partial_y\beta. \tag{17.184}$$

(Itō–Stratonovich conversion)

Clearly, this distinction only matters in the case of multiplicative noise. To show this, recall that the Itō SDE is the continuum limit of the discrete relation

$$y(t + \Delta t) = y(t) + \alpha(y(t),t)\Delta t + \beta(y(t),t)\Delta W(t), \tag{17.185}$$

while the Stratonovich SDE is the continuum limit of

$$y(t + \Delta t) = y(t) + \bar{\alpha}(y(t),t)\,\Delta t + \beta(y(t + \Delta/2), t + \Delta t/2)\,\Delta W(t). \tag{17.186}$$

Now noting that

$$\beta(y(t+\Delta t/2), t+\Delta t/2) = \beta(y(t),t) + \partial_y\beta(y(t),t)\left[\beta(y(t),t)\,\Delta W^{(1/2)}(t)\right] + O(\Delta t), \tag{17.187}$$

where the factor in square brackets is Δy, $\Delta W^{(1/2)}(t) := W(t+\Delta t/2) - W(t)$ is a Wiener "half increment," and we are dropping terms of order Δt and higher since this expression will be multiplied by ΔW. We can thus use this result to write

$$y(t+\Delta t) = y(t) + \bar{\alpha}(y(t),t)\,\Delta t + \beta(y(t),t)\,\Delta W(t) + \beta(y(t),t)\,\partial_y\beta(y(t),t)\,\Delta W(t)\,\Delta W^{(1/2)}(t). \tag{17.188}$$

In the continuous limit, we can write

$$\Delta W(t)\,\Delta W^{(1/2)}(t) \longrightarrow \frac{dt}{2}, \tag{17.189}$$

since $[\Delta W^{(1/2)}]^2 \longrightarrow dt/2$ with certainty, and the product of $\Delta W^{(1/2)}$ from two different time intervals will converge to zero with certainty. Thus, the continuum limit of (17.186) is the Itō-form SDE

$$dy = \left[\bar{\alpha} + \frac{1}{2}\beta\partial_y\beta\right]\,dt + \beta\,dW. \tag{17.190}$$

Since this is the continuum limit of the same equation as the Stratonovich form (17.183), we can thus conclude that the Itō and Stratonovich forms are equivalent in the case

$$\bar{\alpha} = \alpha - \frac{1}{2}\beta\partial_y\beta. \tag{17.191}$$

Thus, when writing down an SDE, we again see that it is crucial to specify which calculus it assumes, since the solution would otherwise be ambiguous.

17.4.3 Stratonovich Calculus and the Chain Rule

Recall that the Itō equation

$$dy = \alpha\,dt + \beta\,dW \tag{17.192}$$

transforms under the coordinate change $z = f(y)$ via an "extended chain rule"

$$\begin{aligned} dz = df(y) &= f'(y)\,dy + \frac{1}{2}f''(y)(dy)^2 \\ &= \left[f'(y)\,\alpha + \frac{1}{2}f''(y)\,\beta^2\right]dt + f'(y)\,\beta\,dW. \end{aligned} \tag{17.193}$$

(Itō chain rule, $z = f(y)$)

We will now show that in Stratonovich calculus, the "extra" β^2 term does not appear, and thus the usual chain rule applies (and thus $dW^2 = 0$).

Thus, consider the *Stratonovich* equation

$$dy = \alpha\,dt + \beta \circ dW \tag{17.194}$$

under the same transformation $z = f(y)$. The equivalent Itō equation is

$$dy = \left[\alpha + \frac{1}{2}\beta\partial_y\beta\right]dt + \beta\,dW, \tag{17.195}$$

and now we can use Itō rules to accomplish the transformation:

$$\begin{aligned} dz &= f'(y)\,dy + \frac{1}{2}f''(y)(dy)^2 \\ &= \left[f'(y)\left(\alpha + \frac{1}{2}\beta\partial_y\beta\right) + \frac{1}{2}f''(y)\beta^2\right]dt + f'(y)\beta\,dW. \end{aligned} \tag{17.196}$$

Now transform this back into Stratonovich form:

$$dz = \left[f'(y) \left(\alpha + \frac{1}{2} \beta \partial_y \beta \right) + \frac{1}{2} f''(y) \beta^2 - \frac{1}{2} f'(y) \beta \partial_z [f'(y) \beta] \right] dt + f'(y) \beta \circ dW. \tag{17.197}$$

Noting that

$$\partial_z = \frac{1}{f'(y)} \partial_y, \tag{17.198}$$

we can write the last dt term as

$$\begin{aligned} -\frac{1}{2} f'(y) \beta \partial_z [f'(y) \beta] &= -\frac{1}{2} \beta \partial_y [f'(y) \beta] \\ &= -\frac{1}{2} f''(y) \beta^2 - \frac{1}{2} f'(y) \beta \partial_y \beta. \end{aligned} \tag{17.199}$$

Thus, this term, which we obtained from switching from Itō to Stratonovich form, cancels the other two dt terms that involve β, leaving

$$dz = f'(y)\, \alpha\, dt + f'(y)\, \beta \circ dW. \tag{17.200}$$

(Stratonovich chain rule, $z = f(y)$)

Thus, Stratonovich calculus obeys the usual chain rule, and we have no need for terms of order dW^2.

17.4.4 Comparison

We will now summarize the differences between Itō and Stratonovich SDEs, and then explain why we tend to favor Itō calculus. For Itō calculus:

- The rules of stochastic integration are slightly more complicated than for ordinary integration, since the ordinary chain rule does not apply (i.e., $dW^2 = dt$).
- The solution $y(t)$ of an Itō SDE and the driving Wiener process $dW(t)$ are statistically independent at equal times, so that ensemble averages are simply computed by setting $dW = 0$.
- An Itō SDE is ''natural'' as the continuum limit of an evolution constructed by a discrete-step process, since

  $$dy = \alpha(y)\, dt + \beta(y)\, dW \tag{17.201}$$

  is the continuous limit of

  $$y(t + \Delta t) = y(t) + \alpha(y(t))\, \Delta t + \beta(y(t))\, \Delta W(t). \tag{17.202}$$

On the other hand, for Stratonovich calculus:

- The rules of stochastic integration are those of ordinary Riemann integration, since the usual chain rule applies (i.e., $dW^2 = 0$).
- The solution $y(t)$ of a Stratonovich SDE and the driving Wiener process $dW(t)$ are *not* statistically independent at equal times. This is clear from the above Itō–Stratonovich conversion, since setting $dW = 0$ in a Stratonovich SDE does not give the same result as setting $dW = 0$ in the equivalent Itō SDE. In fact, the easiest rule for computing an ensemble average is to convert the SDE to Itō form and then set $dW = 0$.
- A Stratonovich SDE is ''natural'' as the idealization of a *physical* noise process in the following sense. If one models a stochastic system as being driven by a *physical* noise of finite bandwidth and bounded variance, then the normal rules of calculus apply. For example if dO represents an Ornstein–Uhlenbeck process, then we could model a system by the SDE

  $$dy = \alpha\, dt + \beta\, dO. \tag{17.203}$$

Then if you take the white-noise limit for the driving process ($\gamma \longrightarrow 0$), then dO goes over to dW, but because the ODE always obeyed the rules of ordinary calculus, the white-noise limit

$$dy = \alpha\, dt + \beta \circ dW. \tag{17.204}$$

should be interpreted as a Stratonovich SDE.

Also note that most proofs, including the construction of Stratonovich calculus, are usually proved in Itō calculus, so its advantages tend to outweigh its peculiarities. For handling quantum measurements, we will often want to compute ensemble averages to obtain unconditioned master equations, and we will also in general construct continuous measurements as limits of discrete processes of the form (17.202). Thus, we will virtually always use Itō-form SDEs to handle continuous quantum measurements.

17.5 Poisson Process

Recall that the Poisson probability distribution of mean λ is

$$P(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{17.205}$$
(Poisson distribution)

where n is a nonnegative integer. The variance of the Poisson distribution is equal to the mean λ. The Poisson distribution models the number of *independent* random events that occur in a given interval of time, such as the number of cars that arrive at an intersection or the number of atoms that decay in a large ensemble. Poisson random variation is also responsible for **shot noise**, which occurs as noise in electrical current due to random fluctuations in the rate at which electrons flow through a device, or as noise in the detected intensity of classical light due to the random detection times of individual photons. In general, we can speak of a *rate* at which events occur by setting $\lambda = \Gamma\,\Delta t$ for finite time interval Δt, where Γ is the mean rate of occurence (also called the **intensity** of the Poisson process). Then

$$P(n) = \frac{e^{-\Gamma\,\Delta t}(\Gamma\,\Delta t)^n}{n!}. \tag{17.206}$$

Note that the Poisson distribution implies an exponential waiting time for the first event, because the probability for the event to occur *after* waiting a time Δt is given by setting $n = 0$ in the above probability function:

$$P(0) = e^{-\Gamma\,\Delta t}. \tag{17.207}$$

Then according to our interpretation, this probability is related to the probability density $P'(t)$ for the time of first occurence by

$$e^{-\Gamma\,\Delta t} = \int_{\Delta t}^{\infty} P'(t)\, dt, \tag{17.208}$$

so that

$$P'(t) = \Gamma e^{-\Gamma t}. \tag{17.209}$$

Thus, Poisson random variables are intimately connected with exponential-decay processes, such as spontaneous emission from an atom prepared in the excited state.

In the infinitesimal limit $\Delta t \longrightarrow dt$, all the probabilities for $n \geq 2$ becomes negligible (of order dt^2 or higher). The probability for a single event occuring during an interval of duration dt thus becomes $\Gamma\, dt$, with no events occuring otherwise. We can denote this by the infinitesimal random variable $dN(t)$—the **Poisson process**—which has an ensemble mean

$$\langle\!\langle dN(t) \rangle\!\rangle = \Gamma\, dt. \tag{17.210}$$
(Poisson process: ensemble mean)

In the standard Poisson process $N(t)$, the intensity Γ is a constant that characterizes the process. Thus, in general, when writing down a Poisson process, you must always also specify the intensity, which is

not specified in the notation dN in contrast to the Wiener process dW (where there is no freedom to specify the moments). In generalizations of the Poisson process to time- or state-dependent intensities (see Section 17.5.2), an explicit specification of the intensity is even more critical.

Again, as an integer-valued differential random variable, dN can take on only the values 0 and 1, where the value of 1 occurs with probability equal to the mean. Because $dN(t) \in \{0, 1\}$, it immediately follows that

$$dN^2 = dN, \tag{17.211}$$
(Poisson-process property)

so that the ensemble-averaged variance is equal to the ensemble mean,

$$\langle\langle dN^2 \rangle\rangle = \langle\langle dN \rangle\rangle = \Gamma\, dt, \tag{17.212}$$

as we expect for a Poisson-distributed random variable. Note that the infinitesimal variance here is just the second moment, since the square of the mean is $O(dt^2)$.

In another view, note that $dN(t)/dt$ is zero except in isolated intervals of length dt, where the value is $1/dt$. Thus, we can write this form of the Poisson process as the sum of delta functions,

$$\frac{dN(t)}{dt} = \sum_j \delta(t - t_j), \tag{17.213}$$

if the events occur at times t_j. In view of our discussion above, $\Delta t_j := t_{j+1} - t_j$ is a random variable with probability density

$$P(\Delta t_j) = \Gamma e^{-\Gamma\, \Delta t_j}, \tag{17.214}$$

since the waiting time until the next event is given by the exponential distribution.

17.5.1 The Poisson Process Implies the Poisson Distribution

If we take the Poisson process $dN(t)$ of intensity Γ (i.e., of mean $\Gamma\, dt$) as the fundamental object, we should also be able to derive the Poisson distribution for the frequency of events in finite time intervals. That is, we can show that the Poisson distribution arises if there is a constant probability per unit time of a *single* event occuring during an arbitrarily short time interval. Specifically, defining the time integral of the Poisson process,

$$\Delta N := \int_t^{t+\Delta t} dN(t') \tag{17.215}$$

we can ask, what is the probability distribution for ΔN? For a given value n of ΔN, this means that during exactly n infinitesimal intervals, $dN(t)$ took on the value unity, while it took on the value of zero during the remaining intervals. The probability of doing so is the product of three factors:

1. The probability for having exactly n such events, one in each of n particular time intervals: $(\Gamma\, dt)^n$.
2. The number of ways to distribute the n events among all such intervals. In the time interval $[t, t + \Delta t)$, there are $\Delta t/dt$ such time intervals, and so there are $(\Delta t/dt)^n$ ways to distribute n events among all possible time intervals. But we divide by $n!$ since we take the n events to be indistinguishable, so we don't overcount, so the total factor is $(\Delta t/dt)^n/n!$.
3. The probability of having zero events in *all other* time intervals. Again, there are $(\Delta t/dt)$ total time intervals, and the probability of zero events in a given time interval is $(1 - \Gamma\, dt)$, so the total probability is

$$(1 - \Gamma\, dt)^{\Delta t/dt} = \lim_{M\to\infty} \left(1 - \frac{\Gamma\, \Delta t}{M}\right)^M = e^{-\Gamma\, \Delta t}. \tag{17.216}$$

We are being cavalier in slinging around factors of dt, but our manipulations here are equivalent to the "correct" approach of using finite, small subintervals δt, where we neglect n compared to $\Delta t/\delta t$, and we

neglect the probability that two events end up in the same subinterval. Both of these approximations are appropriate (and exact) in the continuum limit. The total probability is thus

$$\begin{aligned} P(\Delta N = n) &= (\Gamma\, dt)^n \left(\frac{\Delta t}{dt}\right)^n \frac{1}{n!} e^{-\Gamma\,\Delta t} \\ &= \frac{e^{-\Gamma\,\Delta t}(\Gamma\,\Delta t)^n}{n!} \\ &= \frac{e^{-\lambda}\lambda^n}{n!}, \end{aligned} \tag{17.217}$$

which is the Poisson distribution, where again the mean is $\lambda = \Gamma\,\Delta t$. Thus, the Poisson distribution results when there are many ''trials'' (short time intervals), where there is a vanishingly small probability of ''success'' (an event occurrence) in each trial.

More mathematically, the Poisson process $N(t)$ is typically defined[9] for $t \geq 0$ in such a way that $N(t)$ is piecewise constant, taking on only constant, nonnegative integer values in intervals of the form $[t_j, t_{j+1})$, and with ''jumps'' of unit size occurring at the times t_j, so that $N(t)$ is also nondecreasing. Further, $N(t)$ must follow a Poisson distribution (with mean Γt in the homogeneous case, where Γ is the intensity). A more general process that relaxes the requirement of the Poisson distribution is called a **counting process**, and a **jump process** additionally relaxes the requirement of unit jumps (and therefore of monotonicity, if negative jumps are allowed).

17.5.2 Inhomogeneous Poisson Process and State Dependence

In general, the rate Γ of event occurence may depend on time, either explicitly or via dependence on the state $y(t)$ of the system. We can handle this by noting that according to our construction for the Poisson distribution above, then if X_1 and X_2 are Poisson-distributed random variables with means λ_1 and λ_2, then $X_1 + X_2$ is a Poisson random variable of mean $\lambda_1 + \lambda_2$. This statement amounts to agglomerating two adjacent time intervals of different duration in the above derivation of the Poisson distribution. Then if Γ is time-dependent, we can subdivide the interval $[t, t + \Delta t)$ into sufficiently fine increments such that Γ is constant over each increment, and sum them to find that the number of events occuring in the interval $[t, t + \Delta t)$ is still a Poisson variable with mean

$$\bar{\lambda} = \int_t^{t+\Delta t} \Gamma(t')\, dt'. \tag{17.218}$$
(inhomogeneous Poisson process: mean)

The variance is also of course just $\bar{\lambda}$.

However, while we can define the inhomogeneous Poisson process as above, a generalization to a process with state-dependent intensity $\Gamma(y)$, where $y(t)$ is some process driven by $dN(t)$, is *not* a Poisson process: the argument above for the inhomogeneous process does not apply, because $dN(t)$ is no longer statistically independent at different times.[10] Since in this case $N(t)$ is no longer Poisson-distributed, it is more proper to refer to it as a *counting process*, as we defined in Section 17.5.1.

17.5.3 White-Noise Limit

Consider the scaled Poisson process

$$dy = \frac{dN}{\sqrt{\Gamma}}, \tag{17.219}$$

where

$$\langle\langle dN \rangle\rangle = \Gamma\, dt. \tag{17.220}$$

[9] see, e.g., Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes* (Chapman & Hall/CRC, 2004), Section 2.5.3, p. 48.

[10] Note, however, that this is *still* sometimes referred to as a ''state-dependent Poisson process.'' See, e.g., Edoardo Daly and Amilcare Porporato, ''Intertime jump statistics of state-dependent Poisson processes,'' *Physical Review E* **75**, 011119 (2007) (doi: 10.1103/PhysRevE.75.011119).

It may be that in a given system, the rate Γ of Poisson events is much faster than the processes of physical interest. In such a case, we can ignore the discreteness of the events, and coarse-grain the dynamics to approximate the Poisson events by white noise. Note in particular that the mean of dy is

$$\langle\langle dy \rangle\rangle = \frac{\langle\langle dN \rangle\rangle}{\sqrt{\Gamma}} = \sqrt{\Gamma}\, dt, \tag{17.221}$$

while the variance is

$$\langle\langle (dy)^2 \rangle\rangle = \frac{\langle\langle dN^2 \rangle\rangle}{\Gamma} = \frac{\langle\langle dN \rangle\rangle}{\Gamma} = dt. \tag{17.222}$$

Thus, if events occur rapidly on time scales of interest—that is, we only measure Δy over time intervals $\Delta t \gg 1/\Gamma$, by the central limit theorem, we may effectively regard dy as a *Gaussian* random variable of mean $\sqrt{\Gamma}\, dt$ and variance dt. In particular, the Poisson process corresponds to a random walk of steps of length $\sqrt{\Gamma}$, in one direction only, taken at random times, as we can see by writing

$$y(t) = \frac{1}{\sqrt{\Gamma}} \int_0^t dN(t') = \frac{1}{\sqrt{\Gamma}} \int_0^t \frac{dN(t')}{dt'}\, dt', \tag{17.223}$$

and recalling from Eq. (17.213) that dN/dt is a sum of delta functions. After many events, the result is the same as a biased random walk, and the central limit theorem again guarantees an asymptotic Gaussian probability density. In this limit, it thus is a good approximation to write

$$dy = \sqrt{\Gamma}\, dt + dW. \tag{17.224}$$

Thus, in the limit where Poisson events occur at a very large rate Γ, we can make the formal replacement

$$dN \longrightarrow \Gamma\, dt + \sqrt{\Gamma}\, dW, \tag{17.225}$$

(white-noise limit of Poisson process)

to approximate the Poisson process with a mean drift plus white noise.

17.5.3.1 Shot Noise

As an example, let's consider shot noise of an electrical current. Let Q denote the total charge that has crossed a certain point along a wire. The current is given by

$$I = \frac{dQ}{dt}, \tag{17.226}$$

so that

$$dQ = I\, dt. \tag{17.227}$$

We will model the current as a stream of *independent* electrons of charge $-e$ with Poisson arrival times, so that

$$dQ = -e\, dN, \tag{17.228}$$

with $\langle\langle dN \rangle\rangle = \Gamma dt$ as usual. Then equating the two expression for dQ, we find

$$I\, dt = -e\, dN. \tag{17.229}$$

The mean current is then given by taking the ensemble average of this relation, so that

$$\langle\langle I \rangle\rangle = -e\Gamma. \tag{17.230}$$

Frequencies of interest for measuring electrical currents generally range from ns to s, whereas $\Gamma \sim 10^{19}\ \mathrm{s}^{-1}$. Thus, the white-noise approximation is quite appropriate, and thus

$$\begin{aligned} dQ &\approx -e\Gamma\, dt - e\sqrt{\Gamma}\, dW \\ &= \big\langle\big\langle I \big\rangle\big\rangle dt + \frac{\langle\langle I \rangle\rangle}{\sqrt{\Gamma}}\, dW \\ &= \big\langle\big\langle I \big\rangle\big\rangle dt - \sqrt{|e\langle\langle I \rangle\rangle|}\, dW. \end{aligned} \tag{17.231}$$

Thus, we see that due simply to the discreteness of charge, a mean current $\langle\langle I \rangle\rangle$ is accompanied by white noise of amplitude

$$\frac{|\langle\langle I \rangle\rangle|}{\sqrt{\Gamma}} = \sqrt{|e\langle\langle I \rangle\rangle|}. \tag{17.232}$$

Note that the SI units of the current noise amplitude are in $\mathrm{A}/\sqrt{\mathrm{Hz}}$, since when multiplied by dW/dt, which has dimensions $1/\sqrt{\mathrm{s}}$, the noise amplitude takes the dimensions of current. The alternate way to view this is that physically, the noise is always bandwidth-limited (as in the RC model of the Ornstein–Uhlenbeck process), and so the filtered noise amplitude is given by multiplying the above noise amplitude by the square root of the circuit bandwidth. More explicitly, the above white noise corresponds to a uniform spectral density of signal power. According to Eq. (17.164), an input signal $\alpha\xi(t)$ corresponds to rms fluctutations of $\alpha\sqrt{\gamma/2}$ at the output of a low-pass filter. Thus, the rms current fluctuation through a low-pass filter due to shot noise is given by

$$\delta I_{\mathrm{rms}} = \sqrt{|e\langle\langle I \rangle\rangle|}\sqrt{\frac{\gamma}{2}}, \tag{17.233}$$

where $\gamma = 1/RC$ is the angular cutoff frequency for the low-pass filter. The **equivalent-power bandwidth** $\Delta\nu$ is defined as the bandwidth of the ''brick wall'' filter (with flat response up to a sudden cutoff at frequency $\Delta\nu$, where $\Delta\nu$ is a frequency in Hz, *not* an angular frequency). The low-pass filter function for power transmission is (up to a constant factor, set by requiring the dc transmission to unity) the Fourier transform of the Ornstein–Uhlenbeck correlation function $e^{-\gamma\tau}$, so that we may write the transmission function as

$$T(\omega) = \frac{\gamma^2}{\gamma^2 + \omega^2}. \tag{17.234}$$

Obviously, γ is the (angular) ''3 dB'' frequency, or the frequency where the transmission drops to 1/2 the dc value:

$$f_{3\,\mathrm{dB}} = \frac{\gamma}{2\pi}. \tag{17.235}$$

Since

$$\int_0^\infty T(\omega)\,d\omega = \frac{\pi}{2}\gamma = 2\pi\Delta\nu, \tag{17.236}$$

where the second result applies to the brick-wall filter, we can write

$$\Delta\nu = \frac{\gamma}{4} = \frac{\pi}{2} f_{3\,\mathrm{dB}}, \tag{17.237}$$

and thus the shot-noise magnitude is

$$\delta I_{\mathrm{rms}} = \sqrt{2\Delta\nu|e\langle\langle I \rangle\rangle|}. \tag{17.238}$$

(shot-noise amplitude)

This expression applies to filters beyond the low-pass filter, so long as the appropriate equivalent-power bandwidth is used in this relation. Thus, a 1 A average current detected in a 1 MHz bandwidth has an rms noise current of 0.57 μA, a fluctuation at under the ppm level. Shot noise clearly gets much worse for smaller currents: for the same bandwidth, an average current of 1 μA has fluctuations of 0.57 nA rms, or 0.057% relative noise, and an average current of 1 pA has fluctuations of 0.57 pA rms, or 57% relative noise. Note that this model assumes the independence of electrons, and gives an appropriate result, e.g., for semiconductor junctions, but not in metallic-wire circuits, where long-range correlations between electrons suppress shot noise.[11] Essentially, this is just because of Coulomb interactions between electrons, which causes them to antibunch: in a conducting, crystalline lattice, it is energetically favorable to have two electrons in *different* lattice sites, as compared to having them occupy the *same* lattice site. Probabilities of seeing more than one electron pass in a metallic wire in a short time interval are thus suppressed compared to the Poissonian expectation.

Of course, we can adapt this result to the case of optical shot noise. Instead of an electrical current, we have a detected optical power P. We can treat the photon arrival times as independent in the case of coherent

[11] Paul Horowitz and Winfield Hill, *The Art of Electronics*, 2nd ed. (Cambridge, 1989), pp. 431-2.

light (recall that in a coherent state the photon-number occupation probabilities are Poisson-distributed). The rms fluctuations for a mean optical power $\langle\langle P \rangle\rangle$ detected in an equivalent-power bandwidth $\Delta\nu$ are then given by

$$\delta P_{\rm rms} = \sqrt{2\Delta\nu\,\hbar\omega\langle\langle P\rangle\rangle|}, \qquad \text{(optical shot-noise amplitude)} \tag{17.239}$$

where the photon energy $\hbar\omega$ plays the role of the electron charge. For 780 nm light in a 1 MHz detection bandwidth, a 1 W power has fluctutations of 0.71 μW, less than the ppm level. For the same bandwidth, a 1 μW mean power has fluctuations of 0.71 nW (0.071%), and a 1 pW mean power has fluctuations of 0.71 pW (71%). Of course, the relative fluctuations will be even larger for thermal (bunched) light, and smaller for antibunched light (nonclassical light, as for resonance fluorenscence of a two-level atom).

17.6 Particle Subject to a Stochastic Force

As an important example to tie together the stochastic-process concepts that we have developed so far, we will treat in some depth the dynamics of a particle subject to a stochastic force, together with damping and other external forces. In particular consider the equations of motion

$$\begin{aligned} dx &= \frac{p}{m}\,dt \\ dp &= \big[F(x,p;t) - \gamma p\big]dt + \sigma\,dW, \end{aligned} \qquad \text{(damped, stochastically forced particle)} \tag{17.240}$$

with F defining an external force, γ the damping coefficient, and σ defining the magnitude of the stochastic force. This is for a particle in one spatial dimension, but the generalization to multiple degrees of freedom is straightforward.

Also, note that γ could depend on x and p (and even t), as could σ. Note that if σ depends on p, then the particle is coupled to multiplicative noise (Section 17.4), and it is important to specify which calculus to use in treating the SDE system. As discussed in Section 17.4.4, if the noise arises as the continuum limit of a physical noise process of finite bandwidth, then Eqs. (17.240) should be interpreted in the Stratonovich sense. On the other hand, if the noise arises due to a process like spontaneous emission, where the emission probability depends on the present state (or arises in terms of a discrete Poisson process, which goes over to white noise in the limit of frequent emission events, as in Section 17.5.3), then the equations of motion should be interpreted in the Itō sense. For concreteness we will treat this system in the Itō sense, although it is useful to remember that as we make assumptions on σ in the discussion below, the importance of this distinction drops away.

From Eqs. (17.145) and (17.146), we can write out the equivalent Fokker–Planck equation for the probability density $f(x,p;t)$ of the particle state as

$$\partial_t f(x,p;t) = -\frac{p}{m}\partial_x f + \partial_p\big[F(x,p;t) - \gamma p\big]f + \frac{1}{2}\partial_p^2\sigma^2 f. \qquad \text{(equivalent Fokker–Planck equation)} \tag{17.241}$$

Note that any dependence of F, γ, and σ on p enforces the ordering of the derivative operators as shown; in the case where there is no such momentum dependence, the derivative operators ∂_p can commute with these functions and operate only on the probability density. But again notice that dependence on x causes no problem; for example, the diffusion term can have the form $\sigma^2(x)\,\partial_p^2 f$ in the case that σ depends only on position.

17.6.1 Free-Particle Limit

One of the simplest limits of the above equations of motion is to assume vanishingly small damping ($\gamma \longrightarrow 0$) and external force ($F \longrightarrow 0$). In this case we can integrate the momentum equation to obtain

$$p(t) = p(0) + \int_0^t dt'\,\sigma(x,p;t')\,dW(t'). \tag{17.242}$$

Then putting this into the position equation and integrating, we find

$$x(t) = x(0) + \frac{p(0)\,t}{m} + \frac{1}{m}\int_0^t dt' \int_0^{t'} dt''\,\sigma(x,p;t'')\,dW(t''). \tag{17.243}$$

It is difficult to carry this further in the general case. However, in the special case where σ is constant, we can characterize the transport of x quite precisely. In this case, we obtain

$$x(t) = \frac{\sigma}{m}\int_0^t dt'\,W(t'), \tag{17.244}$$

where we are also ignoring transients by setting $x(0) = p(0) = 0$. The integral factor here is important, and for obvious reasons it goes by the names **integrated Brownian motion** and **Brownian area** (this is also the iterated integral I_{10} that we discuss in Section 27.3), and here we will write it as

$$I(t) := \int_0^t dt'\,W(t'). \tag{17.245}$$
(integrated Brownian motion)

Since $I(t)$ is a linear combination of Gaussian increments, it is itself Gaussian, so we can characterize it fully by computing its mean and correlation. The mean is simple, since in Itō calculus we just set $dW = 0$:

$$\langle\!\langle I(t) \rangle\!\rangle = 0. \tag{17.246}$$

To compute the variance (see also Problem 27.1), we can write

$$\left\langle\!\!\left\langle I^2(t) \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \left[\int_0^t dt' \int_0^{t'} dW(t'')\right]^2 \right\rangle\!\!\right\rangle. \tag{17.247}$$

Then using the integral identity $\int_0^t dt' \int_0^{t'} dt'' g(t',t'') = \int_0^t dt'' \int_{t''}^t dt' g(t',t'')$ to interchange the order of integration, we have

$$\begin{aligned} \left\langle\!\!\left\langle I^2(t) \right\rangle\!\!\right\rangle &= \left\langle\!\!\left\langle \left[\int_0^t dW(t'') \int_{t''}^t dt'\right]^2 \right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \left[\int_0^t (t-t'')\,dW(t'')\right]^2 \right\rangle\!\!\right\rangle. \end{aligned} \tag{17.248}$$

Then expanding out the product, using $dW(t)\,dW(t') = \delta(t-t')\,dt^2$,

$$\left\langle\!\!\left\langle I^2(t) \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \int_0^t (t-t'')^2\,dt'' \right\rangle\!\!\right\rangle = \frac{t^3}{3}. \tag{17.249}$$

This implies a covariance

$$\left\langle\!\!\left\langle I(t)I(t') \right\rangle\!\!\right\rangle = \frac{1}{3}\big[\min\{t,t'\}\big]^3, \tag{17.250}$$
(covariance of integrated Brownian motion)

because for example if $t' > t$, then the Wiener increments associated with the time interval (t,t') are uncorrelated with any increments associated with the interval $(0,t)$, so the time interval (t,t') does not contribute to the correlation function.

Thus, the stochastically driven, undamped particle (17.244) tends to a Gaussian distribution in x, whose width grows as $(\sigma/m)t^{3/2}$, which is one power of t higher than the $t^{1/2}$ dependence we expect for diffusive growth in momentum.

17.6.2 Thermal Equilibrium

Now consider the particle equations (17.240) *with* damping at rate γ but no external force. This could correspond, for example, to a particle moving (in one dimension) in a static fluid at some temperature T. For simplicity we will assume constant temperature, noise coefficient σ, and damping rate γ.

First, note that the momentum equation in this case is basically a scaled Ornstein–Uhlenbeck process:

$$dp = -\gamma p\, dt + \sigma\, dW. \tag{17.251}$$

In fact, rescaling time via $\tilde{t} = \sigma^2 t$ and $d\tilde{W}(\tilde{t}) = \sigma\, dW(t)$, so that $d\tilde{W}^2 = [\sigma\, dW(t)]^2 = \sigma^2\, dt = d\tilde{t}$, gives

$$dp = -\frac{\gamma}{\sigma^2} p\, d\tilde{t} + d\tilde{W}, \tag{17.252}$$

which has the form of the Ornstein–Uhlenbeck process in Eq. (17.148). Then the results following from that analysis carry through here with the replacements $\gamma \longrightarrow \gamma/\sigma^2$ and $t \longrightarrow \sigma^2 t$. For example, the mean from Eq. (17.152) is only transient,

$$\langle\!\langle p(t) \rangle\!\rangle = p_0 e^{-\gamma t}, \tag{17.253}$$
(momentum mean)

and the correlation function (17.153) becomes

$$\langle\!\langle p(t)\, p(t') \rangle\!\rangle = \left(p_0^2 - \frac{\sigma^2}{2\gamma} \right) e^{-\gamma(t+t')} + \frac{\sigma^2}{2\gamma} e^{-\gamma|t'-t|}, \tag{17.254}$$
(momentum correlation)

where the first term is again transient and the second persists in steady state, and $t, t' \geq 0$.

In thermal equilibrium, we have

$$\frac{1}{2} m \langle\!\langle v^2 \rangle\!\rangle = \frac{\langle\!\langle p^2 \rangle\!\rangle}{2m} = \frac{1}{2} k_{\text{B}} T. \tag{17.255}$$

We can obtain $\langle\!\langle p^2 \rangle\!\rangle = \sigma^2/2\gamma$ in steady state from the last term in Eq. (17.254), which gives

$$\sigma = \sqrt{2m\gamma k_{\text{B}} T}, \tag{17.256}$$
(temperature related to diffusion)

which fixes the noise coefficient in terms of the temperature and damping. Thus, for example, Eq. (17.254) becomes

$$\langle\!\langle p(t)\, p(t') \rangle\!\rangle = m k_{\text{B}} T\, e^{-\gamma|t'-t|} \tag{17.257}$$

in terms of temperature, after ignoring any transients.

The position equation (17.240) then says that the position is the (scaled, shifted) integrated Ornstein–Uhlenbeck process:

$$x(t) = x(0) + \int_0^t dt'\, p(t'). \tag{17.258}$$

For the ‘‘pure’’ Ornstein–Uhlenbeck process (17.148),

$$dy = -\gamma y\, dt + dW, \tag{17.259}$$

we can define the integrated process by

$$G(t) := \int_0^t dt'\, y(t'). \tag{17.260}$$

Then clearly $G(t)$ is again a Gaussian process, with

$$\langle\!\langle G(t) \rangle\!\rangle = 0. \tag{17.261}$$

Computing the correlation function, we start with

$$\left\langle\!\left\langle G(t)\,G(t')\right\rangle\!\right\rangle = \int_0^t ds \int_0^{t'} ds' \left\langle\!\left\langle y(s)\,y(s')\right\rangle\!\right\rangle = \frac{1}{2\gamma}\int_0^t ds \int_0^{t'} ds' \left(e^{-\gamma|s-s'|} - e^{-\gamma(s+s')}\right). \tag{17.262}$$

where we used Eq. (17.153) for the Ornstein–Uhlenbeck correlation function, setting $y_0 = 0$ for simplicity but keeping the transient term, which is important when integrating the correlation function. Then if $t' \geq t$ (so that $t' \geq s$), we can split the second integral to simplify the absolute value:

$$\left\langle\!\left\langle G(t)\,G(t')\right\rangle\!\right\rangle = \frac{1}{2\gamma}\int_0^t ds \left[\int_0^s ds'\, e^{-\gamma(s-s')} + \int_s^{t'} ds'\, e^{\gamma(s-s')}\right] - \frac{1}{2\gamma}\int_0^t ds \int_0^{t'} ds'\, e^{-\gamma(s+s')}. \tag{17.263}$$

Carrying out the integration and simplifying gives

$$\left\langle\!\left\langle G(t)\,G(t')\right\rangle\!\right\rangle = \frac{1}{2\gamma^3}\left[2\left(e^{-\gamma t} + e^{-\gamma t'} - 1\right) - e^{-\gamma(t+t')} - e^{-\gamma(t'-t)}\right] + \frac{t}{\gamma^2}. \tag{17.264}$$

Again, using the requirement that the correlation function should be symmetric in t and t', we find

$$\left\langle\!\left\langle G(t)\,G(t')\right\rangle\!\right\rangle = \frac{1}{2\gamma^3}\left[2\left(e^{-\gamma t} + e^{-\gamma t'} - 1\right) - e^{-\gamma(t+t')} - e^{-\gamma|t'-t|}\right] + \frac{\min\{t,t'\}}{\gamma^2}.$$

(integrated-Ornstein–Uhlenbeck correlation) (17.265)

Note that $\langle\!\langle G(0)\,G(0)\rangle\!\rangle = 0$, as expected. Note also that $\langle\!\langle G(t)\,G(t')\rangle\!\rangle \geq 0$, because we derived it as the integral of $\langle\!\langle y(t)\,y(t')\rangle\!\rangle \geq 0$.

Returning to the transport of the particle in the fluid, we are interested in the asymptotic growth of the variance of $G(t)$. For large times, Eq. (17.265) gives an asymptotic growth

$$\left\langle\!\left\langle G^2(t)\right\rangle\!\right\rangle \sim \frac{t}{\gamma^2} \tag{17.266}$$

to leading order in t. The same procedure with $x(t) = \int_0^t dt'\, p(t')/m$ and Eq. (17.254) for the momentum correlation function gives

$$\left\langle\!\left\langle x^2(t)\right\rangle\!\right\rangle \sim \left(\frac{\sigma}{m\gamma}\right)^2 t = \frac{2k_{\mathrm{B}}T}{m\gamma}t. \tag{17.267}$$

(position diffusion)

That is, $x(t)$ grows diffusively with diffusion coefficient $(\sigma/m\gamma)^2$. Thus, the damping "transfers" the diffusive behavior from momentum (which is driven by the Wiener process) to position. We will see this more directly in the following section.

17.6.3 Strong-Damping Limit: Brownian Motion Revisited

Back in Section 17.3.6.1, we commented that the Ornstein–Uhlenbeck process is a model for Brownian motion—as we just showed, a damped and stochastically driven momentum leads to diffusion in position. At the same time, the Wiener process $W(t)$ itself is often referred to as "Brownian motion," although in the physical model it enters via the force, not directly in the particle's position. Here we will make the connection more directly of the stochastic force leading to diffusion in position, in the case of strong damping.

Starting again with the equations of motion (17.240),

$$\begin{aligned} dx &= \frac{p}{m}\,dt \\ dp &= \big[F(x,p;t) - \gamma p\big]dt + \sigma\,dW, \end{aligned} \tag{17.268}$$

with the same comments about the (x,p)-dependence of γ and σ, we will begin by assuming a large damping coefficient γ. In the limit of large damping, we will assume that the momentum is always approximately in

equilibrium with respect to the position state, so that we may perform an adiabatic elimination of momentum (see Section 5.8.3 for the adiabatic approximation in the two-level atom). Setting $dp \approx 0$ then leads to the adiabatic approximation

$$p \approx \frac{1}{\gamma}\left[F + \sigma\,\frac{dW}{dt}\right]. \tag{17.269}$$

In assuming large damping, we are essentially assuming that the damping term ''absorbs'' the entire effect of the external and fluctuating forces, provided we coarse-grain on time scales much longer than γ^{-1}. The difficulty here is that dW/dt is divergent and fluctuates on all time scales. We will proceed for now, but we will soon see that we need to be somewhat more careful with this approximation.

Putting Eq. (17.269) into the position equation $\dot{x} = p/m$, we find

$$dx \approx \frac{F}{m\gamma}\,dt + \frac{\sigma}{m\gamma}\,dW.$$

(particle motion, adiabatic approximation) (17.270)

In the absence of external forces ($F = 0$), then we see directly that $x(t) \propto W(t)$, which justifies the name Brownian motion for $W(t)$. And again appealing to Eqs. (17.145) and (17.146), the equivalent Fokker–Planck equation for this approximate SDE is

$$\partial_t f(x;t) = -\partial_x \frac{F}{m\gamma} f + \frac{1}{2}\partial_x^2 \frac{\sigma^2}{m^2\gamma^2} f.$$

(Smoluchowski equation) (17.271)

Here, F, γ, and σ can still depend on position and momentum; however, the dependence on momentum should be in the form of $\langle\!\langle p(x) \rangle\!\rangle = F(x,p;t)/\gamma$ from the adiabatic relation (17.269), because the large damping ''concentrates'' the momentum around the mean value. This diffusion equation is the counterpart of the original Fokker–Planck equation (17.241) after momentum is adiabatically eliminated. In the case of coupling to a uniform bath in thermal equilibrium, σ is constant and given by Eq. (17.256), such that

$$\partial_t f(x,t) = -\frac{1}{m\gamma}\partial_x F f + \frac{k_{\mathrm{B}}T}{m\gamma}\partial_x^2 f, \tag{17.272}$$

(Smoluchowski equation)

after also assuming constant damping γ. This is called the **Smoluchowski equation**.[12]

Note that in the form (17.272), at any given temperature, the steady-state solution is independent of m and γ. In particular, assuming that the force may be derived from a potential, $F(x) = -V'(x)$, we can write

$$\partial_x V' f + k_{\mathrm{B}}T\partial_x^2 f = 0 \tag{17.273}$$

in steady state. Removing one derivative gives

$$\partial_x f = -\frac{V'}{k_{\mathrm{B}}T} f, \tag{17.274}$$

where we can see that any constant of integration vanishes by comparing this equation with its $x \longrightarrow -x$ counterpart. Integrating this equation from 0 to x gives

$$f(x) = f(0)\, e^{-[V(x)-V(0)]/k_{\mathrm{B}}T}, \tag{17.275}$$

(Boltzmass distribution)

which is just the Boltzmann distribution.

[12] M. San Miguel and J. M. Sancho, ''A Colored-Noise Approach to Brownian Motion in Position Space. Corrections to the Smoluchowski Equation,'' *Journal of Statistical Physics* **22**, 605 (1980) (doi: 10.1007/BF01011341), Eq. (2.39).

17.6.3.1 Momentum Distribution

Given the reduced distribution $f(x,t)$ to the Smoluchowski equation (17.272), it is useful to see how to reconstruct the information about momentum that is implicit in the solution. First, as we already mentioned, the adiabatic relation (17.269) yields directly the mean momentum at each position:

$$\langle\!\langle p(x) \rangle\!\rangle \approx \frac{F(x, \langle\!\langle p(x) \rangle\!\rangle; t)}{\gamma}. \tag{17.276}$$

Here, to be precise, the ensemble average is over all dW, but at fixed position x. The mean position corresponding to the entire ensemble of particles must be averaged over the distribution $f(x,t)$:

$$\langle\!\langle p(t) \rangle\!\rangle_x \approx \int_{-\infty}^{\infty} dx\, f(x,t)\, \frac{F(x, \langle\!\langle p(x) \rangle\!\rangle; t)}{\gamma}. \tag{17.277}$$

(ensemble-mean momentum)

Here, the subscript on the ensemble average emphasizes the average over the spatial distribution.

However, the same approach applied to the second moment $\langle\!\langle p^2 \rangle\!\rangle$ is problematic, because the square of Eq. (17.269) contains a term with the divergent factor $dW^2/dt^2 = 1/dt$. This is a result of the infinite power associated with the white-noise force dW/dt. This is a sign that we need to treat the adiabatic approximation with more care. To do this we return to the momentum equation of motion in Eqs. (17.240),

$$dp = \big[F(x,p;t) - \gamma p\big]dt + \sigma\, dW, \tag{17.278}$$

and note that the same procedure that we used for the Ornstein–Uhlenbeck process in Section 17.3.6 applies here. Thus, we may rewrite the solution to Eq. (17.278) as

$$p(t) = p(0)\, e^{-\gamma t} + \int_0^t dt'\, e^{-\gamma(t-t')} \left[F + \sigma\, \frac{dW(t')}{dt'} \right]. \tag{17.279}$$

The essential point here is that the exponential convolution kernel here smooths the white-noise force, endowing it with finite power and removing the divergence in the second momentum moment. What we did before was to assume that the bracketed factor in the integrand changed slowly over the time scale γ^{-1}, and that $t \gg \gamma^{-1}$, so that we could make the replacement

$$e^{-\gamma(t-t')}\Theta(t-t') \longrightarrow \frac{1}{\gamma}\, \delta(t - t' - 0^+). \tag{17.280}$$

Making this replacement and carrying out the integral leads directly to the previous adiabatic relation (17.269). However, we are only justified in doing this for the external-force term; the white-noise term is *not* constant over *any* time scale. Thus, we may write Eq. (17.279) as

$$p(t) = p(0)\, e^{-\gamma t} + \frac{F}{\gamma} + \sigma \int_0^t e^{-\gamma(t-t')}\, dW(t'). \tag{17.281}$$

Under the assumption that F, γ, and σ vary slowly over the damping time γ^{-1}, we may regard these as constants. Coarse graining over the damping time scale, we can take the long-time limit of the above equation to remove transient behavior:

$$p(x) = \frac{F}{\gamma} + \sigma \int_0^t e^{-\gamma(t-t')}\, dW(t'). \tag{17.282}$$

Now we regard the momentum as a function of the particle's position, which is approximately constant. Under these conditions, the first term is the same mean position that we found before in Eq. (17.276), while the second represents (small) Gaussian fluctuations about the mean. Thus, the momentum distribution is a Gaussian tightly localized about the mean, and all that remains is to characterize the second moment. For this, note that $p(x) - F/\gamma = p(x) - \langle\!\langle p(x) \rangle\!\rangle$ is proportional to an Ornstein–Uhlenbeck process. Thus, we may adapt Eq. (17.154) to give the variance

$$V_p(x) = \Big\langle\!\!\Big\langle \big(p(x) - \langle\!\langle p(x) \rangle\!\rangle\big)^2 \Big\rangle\!\!\Big\rangle = \frac{\sigma^2}{2\gamma} = m k_{\text{B}} T, \tag{17.283}$$

where the latter expression follows from Eq. (17.256), and in the adiabatic regime can be regarded as correct even with spatial variation in temperature.

Now when measuring the momentum distribution of the entire ensemble, we should average the momentum variance over the spatial distribution. However, it is more appropriate to average the second *moment* of momentum rather than the variance, because local variations in the mean momentum are reflected as variance in the *global* distribution. Thus, we require

$$\left\langle\!\left\langle p^2(x) \right\rangle\!\right\rangle = V_p(x) + \left\langle\!\left\langle p(x) \right\rangle\!\right\rangle^2 = \frac{\sigma^2}{2\gamma} + \frac{F^2}{\gamma^2}. \tag{17.284}$$

Then averaging this over the spatial distribution from the Smoluchowski equation (17.272), we find

$$\left\langle\!\left\langle p^2(t) \right\rangle\!\right\rangle_x = \int_{-\infty}^{\infty} dx\, f(x,t) \left[\frac{\sigma^2}{2\gamma} + \frac{F^2}{\gamma^2}\right].$$

(distribution-average of second momentum moment) (17.285)

At this point one can subtract the square of the distribution-averaged mean (17.277) to obtain the momentum variance of the whole distribution. However, Eq. (17.284) is already sufficient to characterize the momentum distribution in important cases such as the steady-state distribution in a potential well, where the mean momentum vanishes.

17.7 Stochastic Boundary-Value Problems: Brownian Bridges

We have already studied the formalism to handle the simple stochastic differential equation

$$dB = dW, \tag{17.286}$$

for which the solution is $B(t) = W(t)$, up to an arbitrary additive constant. (Recall that this is the same in either Itō or Stratonovich calculus, since the noise is additive.) However, suppose we add the additional constraints $B(0) = B(1) = 0$, and we want the solution $B(t)$ for $t \in [0, 1]$. The $B(0) = 0$ constraint is not a problem, as it defines the initial condition of the problem. But the *final* condition $B(1) = 0$ is considerably more difficult, as $W(t)$ in general tends to wander *away* from zero. However, a (vanishingly) small subset of solutions obey this final condition, so in principle we could simulate many possible realizations of $W(t)$, and discard them until we find one that returns sufficiently close to zero at $t = 1$. This kind of constrained random walk, or "stochastic loop", comes up, for example, in quantum field theory.[13] This problem is also a nice example, showing alternate approaches to solving stochastic equations, and providing more insight into regular diffusion $W(t)$.

One simple guess at a solution is simply to *force* a regular Wiener path $W(t)$ back to its initial point by subtracting off its final value $W(t = 1)$, where the subtraction is pro-rated over the path:

$$B(t) := W(t) - tW(1). \tag{17.287}$$

(Brownian bridge)

This is called a **Brownian bridge**, and somewhat surprisingly, this solution satisfies the conditions above for our constrained Wiener path (with some cautions). This is something like viewing a Wiener path $W(t)$ as composed of a linear drift to a final destinations plus fluctuations about zero, and then subtracting off the drift.

To see that this is the case, first we note that $B(t)$ is still a Gaussian random variable, since it is a linear combination of $W(t)$ and $W(1)$, both of which are random variables. Because $\langle W(t) \rangle = 0$, we must also have $\langle y(t) \rangle = 0$. Thus, we must only check the variance of the increments to establish that $B(t)$ is a proper Wiener process. Dividing the unit time interval into N increments, with time steps $\Delta t = 1/N$, with points $B_n := y(n\Delta t)$ and increments $\Delta B_n := B_{n+1} - V_n$, then we will work with

$$\Delta B_n = \Delta W_n - \Delta t\, W(1). \tag{17.288}$$

[13] Holger Gies, Kurt Langfeld, and Laurent Moyaerts, "Casimir effect on the worldline," *Journal of High Energy Physics* **06**, 018 (2003) (doi: 10.1088/1126-6708/2003/06/018).

Using

$$\begin{aligned}\mathrm{Var}[X+Y] &= \big\langle (X+Y)^2 \big\rangle - \langle X+Y\rangle^2 \\ &= \big\langle X^2 \big\rangle + \big\langle X^2 \big\rangle + 2\langle XY\rangle - \langle X\rangle^2 + \langle Y\rangle^2 + 2\langle X\rangle\langle Y\rangle \\ &= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y],\end{aligned} \tag{17.289}$$

we can compute the variance as

$$\begin{aligned}\mathrm{Var}[\Delta B_n] &= \mathrm{Var}[\Delta W_n] + (\Delta t)^2\,\mathrm{Var}[W(1)] - 2\Delta t\,\mathrm{Cov}[\Delta W_n, W(1)] \\ &= \Delta t + (\Delta t)^2 - 2(\Delta t)^2 \\ &= \Delta t - (\Delta t)^2,\end{aligned} \tag{17.290}$$

where $\mathrm{Var}[W(1)] = 1$ and $\mathrm{Var}[\Delta W_n] = \Delta t$. This means that $\mathrm{Var}[dB] = dt$, since $dt^2 = 0$, and so we have the statistics of Wiener noise. Notice that the subtraction of the drift *did* skew the statistics, but by a negligible amount in the continuum limit because the fluctuations become comparatively large. Therefore one should be careful with a naïve discretization of Eq. (17.287),

$$B_n = \sum_{j=1}^{n} \Delta W_j - \frac{n}{N}\sum_{j=1}^{N} \Delta W_j = \frac{1}{\sqrt{N}}\sum_{j=1}^{n} z_j - \frac{n}{N^{3/2}}\sum_{j=1}^{N} z_j, \tag{17.291}$$

where z_j are independent, unit-normal random deviates. This "algorithm" generates paths with increments that have variance smaller than $1/N$ by a factor $(N-1)/N$, so really the entire path should then be rescaled by $N/(N-1)$. As a coordinate transformation, this discretization is also somewhat pathological in mapping N independent coordinates z_j to $N-1$ independent coordinates y_n (see Problem 17.6). Below we will consider another algorithm that generates a Brownian bridge without these issues.

Note that by the above argument, we can add *any* finite, deterministic function to a Wiener process and still obtain a Wiener process. Thus, for example, we can define

$$B_{a\to b}(t) := a + t(b-a) + W(t) - tW(1) \tag{17.292}$$
(Brownian bridge from a to b)

to be a Brownian bridge[14] that connects a to b over the time interval from 0 to 1. (Other time intervals are possible by shifting and scaling the time variable.) A closely related property is that the Brownian bridge $W(t) - tW(1)$ is independent of $W(1)$, as we can see by computing the correlation function

$$\Big\langle\!\Big\langle \big[W(t) - tW(1)\big]\, W(1) \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle W(t)W(1) - tW^2(1) \Big\rangle\!\Big\rangle = t - t = 0. \tag{17.293}$$

This is the continuous version of the coordinate-transform pathology in the discrete case that we noted above. This independence—along with our ability to stretch Wiener paths to create other Wiener paths—has an important meaning. A particular Wiener path $W(t)$ that wanders to $W(t=1)$ is possible with a Gaussian probability density in $W(1)$, so that in particular, large excursions are unlikely. However, once a *particular* value of $W(1)$ is stipulated, the possible paths are essentially Brownian bridges $B(t)$ that return to the initial point at $t = 1$, but with a uniform-velocity drift to $W(1)$. Even if a peculiarly large value of the overall drift $W(1)$ is stipulated, by far the most likely way to accomodate this is to distribute the drift uniformly over the whole time interval.

[14] While the name "Brownian bridge" may more sensibly apply to $B_{a\to b}(t)$, which "bridges" a to b, $B(t) = B_{0\to 0}(t)$ is a special case that bridges the origin to itself, and goes the name of a **standard Brownian bridge**. It is useful to think of $B(t)$ as a bridge in the sense of having a "pinned" solution at $t = 1$. Incidentally, other variations on conditioned Wiener paths are commonly defined, such as the **Brownian meander** $W^+(t)$ (a Wiener path, or Brownian motion $W(t)$, conditioned to be everywhere positive), and the **Brownian excursion**, which is a Brownian meander conditioned on $W^+(1) = 0$, or alternately a standard Brownian bridge $B(t)$ conditioned to be everywhere positive. See., e.g., Jim Pitman, "Brownian Motion, Bridge, Excursion, and Meander Characterized by Sampling at Independent Uniform Times," *Electronic Journal of Probability* **4**, 11 (1999) (doi: 10.1214/EJP.v4-48).

Finally, we can examine the fluctuations of the (loop-style) Brownian bridge,

$$\begin{aligned}\mathrm{Var}[B(t)] &= \mathrm{Var}[W(t) - tW(1)] \\ &= \mathrm{Var}[W(t)] + t^2\mathrm{Var}[W(1)] - 2t\mathrm{Cov}[W(t), W(1)] \\ &= t + t^2 - 2t^2,\end{aligned} \tag{17.294}$$

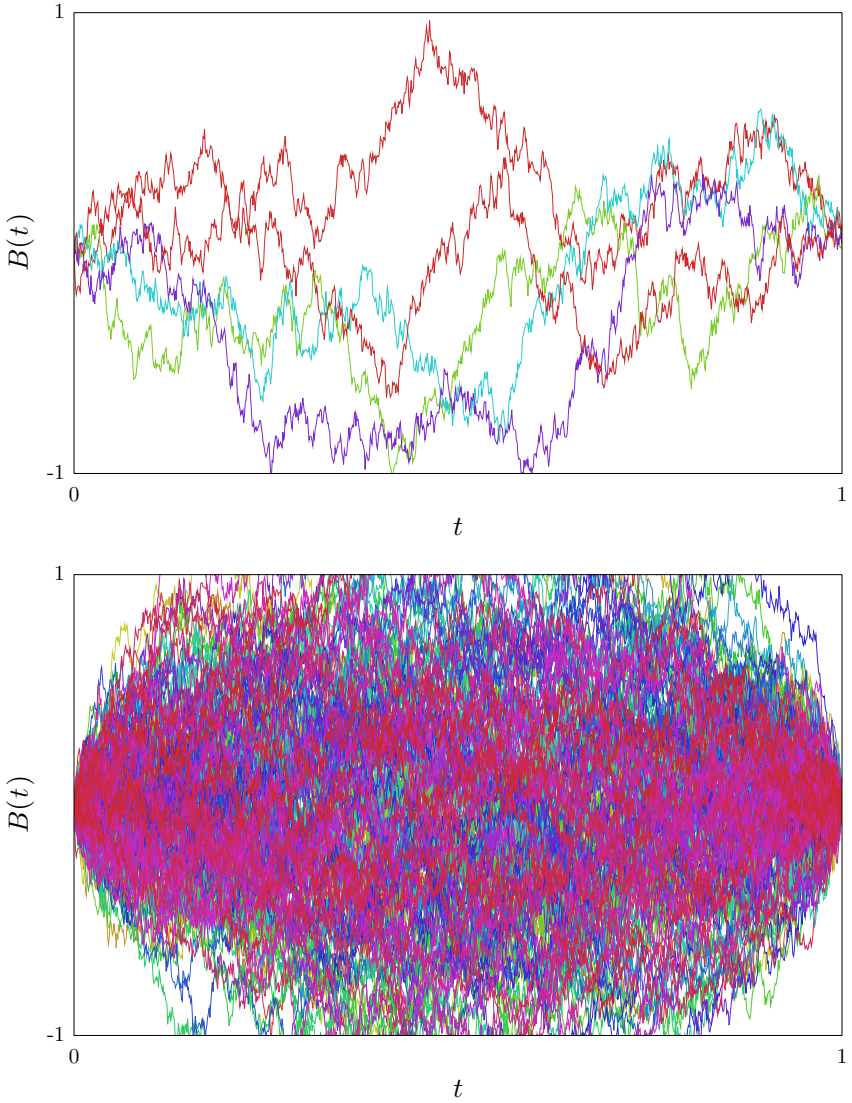
so that we find

$$\mathrm{Var}[B(t)] = t(1-t). \tag{17.295}$$

(variance of Brownian bridge)

Thus, the bridge fluctuates most when it is farthest away from either fixed endpoint, which is sensible.

Again, to get a better idea of what these look like, 5 and 200 Brownian bridges are respectively shown in the two plots below. (Again, these are actually finite realizations of $B(t)$, with $\Delta t = 0.001$.)



Both the symmetry about $t = 0.5$ and the character of diffusing away and returning are readily apparent here.

17.7.1 Finite Bridge Generation: Homogeneous Case

Here we will consider generating a finite numerical approximation[15] to a closed Wiener path in a more direct way than before. If we again divide the path into N increments, then we have time steps

$$\Delta t = \frac{1}{N}, \tag{17.296}$$

and the points $B_n := B(n\Delta t)$, for consistency with Eq. (17.286), must be such that $B_{n+1} - B_n$ is a normally distributed random variable of zero mean and variance $\Delta t = 1/N$. Thus we have the multidimensional probability density

$$P(B_1, \ldots, B_{N-1}) \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N}(B_j - B_{j-1})^2\right)\right], \tag{17.297}$$

where by construction $B_0 = 0$ and $B_N = 0$ are not dependent variables. We will proceed by a coordinate transformation, changing variables to obtain an standard normal Gaussian distribution in every dimension.

First, consider the sum in the exponent, which we may write as

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= \sum_{j=2}^{N-1}(B_j - B_{j-1})^2 + B_1^2 + B_{N-1}^2 \\ &= 2B_1^2 - 2B_1B_2 + 2B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2.\end{aligned} \tag{17.298}$$

Now separating out the B_1 dependence of the exponent,

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= 2\left(B_1 - \frac{1}{2}B_2\right)^2 + \frac{3}{2}B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \frac{3}{2}B_2^2 - 2B_2B_3 + 2B_3^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2.\end{aligned} \tag{17.299}$$

where we completed the square of B_1, and we defined the transformed coordinate

$$B_1' := \sqrt{2}\left(B_1 - \frac{1}{2}B_2\right), \tag{17.300}$$

which encompasses all the dependence on B_1, and enters in the exponent to give a normally distributed random variable with zero mean and variance $1/N$. We now continue to complete squares and factor out the dependence on B_2, B_3, and so on. At the nth step, we have

$$\begin{aligned}\sum_{j=1}^{N}(B_j - B_{j-1})^2 &= B_1'^2 + \cdots + B_{n-1}'^2 + c_nB_n^2 - 2B_nB_{n+1} + B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \cdots + B_{n-1}'^2 \\ &\qquad + c_n\left(B_n - \frac{1}{c_n}B_{n+1}\right)^2 + \left(2 - \frac{1}{c_n}\right)B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2 \\ &= B_1'^2 + \cdots + B_n'^2 + c_{n+1}B_{n+1}^2 + \cdots + B_{N-2}^2 - 2B_{N-2}B_{N-1} + 2B_{N-1}^2\end{aligned} \tag{17.301}$$

where in the basis step above we began with $c_1 = 2$, and we have defined

$$c_{n+1} := \left(2 - \frac{1}{c_n}\right) \tag{17.302}$$

[15] Gies *et al.*, *op. cit.*

and

$$B'_{n+1} := \sqrt{c_n}\left(B_n - \frac{1}{c_n}B_{n+1}\right). \tag{17.303}$$

This has the same form as the $(n+1)$th step, which inductively completes all the squares. With these variables, the probability distribution is

$$P(B'_1, \ldots, B'_{N-1}) \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N-1} B'^2_j\right)\right], \tag{17.304}$$

so that again the B'_j are independent, Gaussian random numbers of zero mean and variance $1/N$. These can be chosen independently, and Eq. (17.303) can be solved to give B_n in terms of B'_n and B_{n+1}:

$$B_n = \frac{B'_n}{\sqrt{c_n}} + \frac{B_{n+1}}{c_n}. \tag{17.305}$$

Thus, the bridge coordinates should be generated in a backwards recurrence, given the forward recurrence for the coefficients c_n. This is even more conveniently given in terms of standard-normal deviates z_j, which can replace the B'_j:

$$B_n = \frac{z_n}{\sqrt{Nc_n}} + \frac{B_{n+1}}{c_n}. \tag{17.306}$$

Note also that the recurrence (17.302) has the solution

$$c_n := \frac{n+1}{n}. \tag{17.307}$$

These relations give the bridge directly in terms of easily generated deviates z_n.

Note that in principle we must also involve the Jacobian determinant of the coordinate transformation from B_j to B'_j. Effectively, we have taken the exponent (17.298), which we can write as a quadratic form as

$$\sum_{j=1}^{N}(B_j - B_{j-1})^2 = B_a A_{ab} B_b, \tag{17.308}$$

where (A_{ab}) is a square, tridiagonal matrix of dimension $N-1$ with a twos along the diagonal and ones for every other nonzero element. Since the matrix is symmetric, it is diagonalized by an orthogonal transformation (P_{ab}), so that $B_a A_{ab} B_b = B'_a D_{ab} B'_b$, where (D_{ab}) is diagonal (and in fact has only one nonzero eigenvalue, whose value is $N-1$), and $B'_a := P_{ab}B_b$. We have effectively performed this diagonalization in constructing the above recurrence relations. Notice then that the Jacobian determinant is just the determinant of (P_{ab}), which is just a constant factor that only affects the normalization of the probability density. Thus, we have justified our coordinate transformation to a new Gaussian distribution.

To summarize, the algorithm to generate a Brownian bridge of N steps (i.e., $N+1$ points $B_0, \ldots B_N$, where $B_0 = B_N = 0$) is:

1. Generate standard normal random numbers (zero mean, unit variance) z_n from $n = 1, \ldots, N-1$.

2. Generate the positions B_n for $n = N, \ldots, 1$, according to the *backwards* recurrence

$$\begin{aligned} &B_N = 0 \\ &B_n = z_n\sqrt{\frac{n}{N(n+1)}} + \left(\frac{n}{n+1}\right)B_{n+1}, \qquad n = N-1, \ldots, 1 \\ &B_0 = 0. \end{aligned} \tag{17.309}$$

We can similarly write a *forward* recurrence

$$\begin{aligned} B_0 &= 0 \\ B_n &= z_n\sqrt{\frac{c_n}{N}} + c_n B_{n-1}, \qquad n = 1, \ldots, N-1 \\ B_N &= 0, \end{aligned} \tag{17.310}$$

where to simplify the notation, we have defined the recurrence coefficient

$$c_n := \frac{N-n}{N-n+1}. \tag{17.311}$$

This forward scheme gives finitely sampled Brownian bridges with the same statistics.

This algorithm gives a simulated (finite) realization of a closed, stochastic Wiener path in terms of easily generated standard-normal random deviates. Note that this algorithm is easily generalized to generate samples of a Brownian bridge $B_{a\to b}(t)$ running from a to b in unit time: the B_n in Eqs. (17.310) should be thought of as $B_n - b$ (i.e., measured in terms of their distance to the "target" b), so that

$$\begin{aligned} B_0 &= a \\ B_n &= z_n\sqrt{\frac{c_n}{N}} + c_n(B_{n-1} - b) + b, \qquad n = 1, \ldots, N-1 \\ B_N &= b, \end{aligned} \tag{17.312}$$

will act as a recurrence for this "open" bridge, with the c_n as before. In the case of a more general Brownian bridge $B_{t(a\to b)}(t')$ running from a to b in time t, this recurrence is further generalized by changing the first term on the right-hand side of the second equation of Eqs. (17.312) from $z_n\sqrt{c_n/N}$ to $z_n\sqrt{c_n \Delta t}$, where $\Delta t = t/N$ (the c_n coefficients again stay the same).

17.7.2 Finite Bridge Generation: Inhomogeneous Case

A slightly more complicated variation on the above recurrences arises when we allow for time-dependent drift and diffusion rates, according to

$$dy(t) = \alpha(t)\,dt + \sigma(t)\,dW(t), \tag{17.313}$$

and again impose the boundary condition $y(1) = y(0)$. We will treat this SDE as an Itō SDE for the purposes of finite differences, but strictly speaking this shouldn't matter since the noise is still additive. In finite form, we have the multivariate probability density

$$P(y_1, \ldots, y_{N-1}) \propto \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N} \frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2}\right)\right], \tag{17.314}$$

where again by construction $y_0 = 0$ and $y_N = 0$ are not dependent variables. We now thus have the exponent sum

$$\begin{aligned} \sum_{j=1}^{N} \frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} &= \frac{(y_1 - \alpha_0/N)^2}{\sigma_0^2} + \frac{(y_2 - y_1 - \alpha_1/N)^2}{\sigma_1^2} + \sum_{j=3}^{N} \frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2} \\ &= \frac{\bar{y}_1^2}{\sigma_0^2} + \frac{(y_2 - \bar{y}_1 - \alpha_1/N - \alpha_0/N)^2}{\sigma_1^2} + \sum_{j=3}^{N} \frac{(y_j - y_{j-1} - \alpha_{j-1}/N)^2}{\sigma_{j-1}^2}, \end{aligned} \tag{17.315}$$

where we have defined

$$\bar{y}_1 := y_1 - \frac{\alpha_0}{N} \tag{17.316}$$

in order to begin eliminating the mean drifts. Continuing in this process, we have

$$\sum_{j=1}^{N}\frac{(y_j-y_{j-1}-\alpha_{j-1}/N)^2}{\sigma_{j-1}^2}=\frac{\bar{y}_1}{\sigma_0^2}+\frac{(\bar{y}_2-\bar{y}_1)^2}{\sigma_1^2}+\cdots+\frac{(\bar{y}_N-\bar{y}_{N-1})^2}{\sigma_{N-1}^2}, \tag{17.317}$$

where

$$\bar{y}_n:=y_n-\frac{1}{N}\sum_{j=0}^{n-1}\alpha_j,\qquad n\in 1,\ldots,N, \tag{17.318}$$

again remembering $y_N=0$. Completing the first square as before,

$$\begin{aligned}\sum_{j=1}^{N}\frac{(y_j-y_{j-1}-\alpha_{j-1}/N)^2}{\sigma_{j-1}^2}&=\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma_1^2}\right)\bar{y}_1^2-\frac{2\bar{y}_1\bar{y}_2}{\sigma_1^2}+\frac{\bar{y}_2^2}{\sigma_1^2}+\frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2}+\cdots+\frac{(\bar{y}_N-\bar{y}_{N-1})^2}{\sigma_{N-1}^2}\\&=\left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_0^2\sigma_1^2}\right)\left[\bar{y}_1-\left(\frac{\sigma_0^2}{\sigma_0^2+\sigma_1^2}\right)\bar{y}_2\right]^2+\left[\frac{1}{\sigma_1^2}-\frac{\sigma_0^2}{\sigma_1^2(\sigma_0^2+\sigma_1^2)}\right]\bar{y}_2^2\\&\qquad+\frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2}+\cdots+\frac{(\bar{y}_N-\bar{y}_{N-1})^2}{\sigma_{N-1}^2}\\&=y_1'^2+\left[\frac{1}{\sigma_1^2}-\frac{\sigma_0^2}{\sigma_1^2(\sigma_0^2+\sigma_1^2)}\right]\bar{y}_2^2+\frac{(\bar{y}_3-\bar{y}_2)^2}{\sigma_2^2}+\cdots+\frac{(\bar{y}_N-\bar{y}_{N-1})^2}{\sigma_{N-1}^2},\end{aligned} \tag{17.319}$$

where we have defined

$$y_1':=\sqrt{\left(\frac{\sigma_0^2+\sigma_1^2}{\sigma_0^2\sigma_1^2}\right)}\left[\bar{y}_1-\left(\frac{\sigma_0^2}{\sigma_0^2+\sigma_1^2}\right)\bar{y}_2\right]. \tag{17.320}$$

At the nth stage of completing the square, we must handle terms of the form

$$\begin{aligned}c_n\bar{y}_n^2-\frac{2\bar{y}_n\bar{y}_{n+1}}{\sigma_n^2}+\left(\frac{1}{\sigma_n^2}+\frac{1}{\sigma_{n+1}^2}\right)\bar{y}_{n+1}^2&=c_n\left(\bar{y}_n-\frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}\right)^2+\left(\frac{1}{\sigma_n^2}+\frac{1}{\sigma_{n+1}^2}-\frac{1}{c_n\sigma_n^4}\right)\bar{y}_{n+1}^2\\&=y_n'^2+c_{n+1}\bar{y}_{n+1}^2,\end{aligned} \tag{17.321}$$

where we have defined the decoupled square

$$y_n':=\sqrt{c_n}\left(\bar{y}_n-\frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}\right) \tag{17.322}$$

and the recursion

$$c_{n+1}=\frac{1}{\sigma_n^2}+\frac{1}{\sigma_{n+1}^2}-\frac{1}{c_n\sigma_n^4}, \tag{17.323}$$

thus inductively completing all the squares. Again, the y_n' are Gaussian numbers, such that we may solve to find the shifted positions

$$\bar{y}_n=\frac{y_n'}{\sqrt{c_n}}+\frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}, \tag{17.324}$$

or in terms of standard-normal deviates,

$$\bar{y}_n=\frac{z_n}{\sqrt{Nc_n}}+\frac{1}{c_n\sigma_n^2}\bar{y}_{n+1}. \tag{17.325}$$

Then solving Eq. (17.318),

$$y_n=\bar{y}_n+\frac{1}{N}\sum_{j=0}^{n-1}\alpha_j \tag{17.326}$$

we find the actual bridge positions.

To summarize, the algorithm is, to generate a stochastic, constrained path of N points $y_1,\ldots y_N$, where $y_0=y_N=0$:

1. Begin with means α_n and standard-deviations σ_n for $n \in 0, \ldots, N = 1$.

2. Generate the coefficients c_n for $n = 1, \ldots, N-1$, according to the recurrence

$$c_1 = \frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}, \qquad c_{n+1} = \frac{1}{\sigma_n^2} + \frac{1}{\sigma_{n+1}^2} - \frac{1}{c_n \sigma_n^4}. \tag{17.327}$$

   If many paths are to be generated, these coefficients only need to be generated once.

3. Generate standard normal random numbers (zero mean, unit variance) z_n from $n = 1, \ldots, N-1$.

4. Generate the shifted positions $\bar{y}_n$ for $n = N, \ldots, 1$, according to the *backwards* recurrence

$$\begin{aligned} \bar{y}_N &= -\sum_{j=0}^{N-1} \alpha_j \\ \bar{y}_n &= \frac{z_n}{\sqrt{N c_n}} + \frac{\bar{y}_{n+1}}{c_n \sigma_n^2}. \end{aligned} \tag{17.328}$$

5. Generate the path positions y_n for $n = 1, \ldots, N$, using

$$y_n = \bar{y}_n + \frac{1}{N} \sum_{j=0}^{n-1} \alpha_j. \tag{17.329}$$

This gives a simulated realization of a closed, stochastic path with nonuniform drift and diffusion. Note that a path with *constant* $\alpha(t) \neq 0$ and $\sigma(t) = 1$ is equivalent to the homogeneous path generated by Eqs. (17.310) in the previous section, owing to the conditioning.

17.7.3 SDE and Integral Representations of the Brownian Bridge

The definition (17.287) of the Brownian bridge involves the pro-rated subtraction of the global drift of a Wiener path. However, given the completed-square construction of Eq. (17.7.1), we can also derive a *local* representation of the Brownian bridge as the solution of an SDE. Recall the *backward* recurrence, Eq. (17.309), which we may solve for B_{n+1}:

$$B_{n+1} = \left(\frac{n+1}{n}\right) B_n - z_n \sqrt{\frac{n+1}{nN}}. \tag{17.330}$$

We can further rewrite this as

$$B_{n+1} - B_n = \left(\frac{1}{n}\right) B_n - \Delta W_n \sqrt{\frac{n+1}{n}}, \tag{17.331}$$

where $\Delta W_n = z_n/\sqrt{N}$, and then

$$B_{n+1} - B_n = \left(\frac{N}{n}\right) B_n \Delta t - \Delta W_n \sqrt{\frac{n+1}{n}}, \tag{17.332}$$

where $\Delta t = 1/N$. Passing over to the continuous-time limit, we let $\Delta t \longrightarrow dt$, $n/N \longrightarrow t$, $\Delta W_n \longrightarrow dW(t)$, and we note that $(n+1)/n$ is only different from unity in a vanishingly small interval of small t, a correction we will ignore in the continuum limit:

$$dB = \frac{B}{t}\, dt - dW. \tag{17.333}$$

Note that due to the backward recurrence, this SDE has the "initial" condition $B(1) = 0$, and should be integrated *backwards* in time. We can fix this by letting $t \longrightarrow 1 - t$, so that $dt \longrightarrow -dt$ and $dW \longrightarrow -dW$, and thus

$$dB = -\left(\frac{B}{1-t}\right) dt + dW. \tag{17.334}$$

(SDE form of Brownian bridge)

Thus, we have a representation of a Brownian bridge as an SDE solution. Here, the SDE is similar to an Ornstein–Uhlenbeck process, where the damping rate increases with time, diverging at $t = 1$ to ensure the return of the bridge.

The solution to Eq. (17.334) is (see Problem 17.7)

$$B(t) = (1-t)\int_0^t \frac{dW(t')}{1-t'}, \tag{17.335}$$

(integral form of Brownian bridge)

which can be verified by differentiating this expression. According to this definition, the Brownian bridge is clearly a Gaussian process, and the correlation function computed from this definition

$$\langle\langle B(t)\,B(t')\rangle\rangle = \min(t,t') - tt'$$

(correlation function of Brownian bridge) (17.336)

matches the same correlation function as computed from the first definition (17.287) (see Problem 17.8). This is sufficient to characterize the Gaussian process, and thus the two definitions are equivalent (at least in the sense of ensemble averages, not in the sense of individual paths).

17.7.4 State-Dependent Diffusion in Brownian Bridges

Suppose we consider an inhomogeneous bridge problem that is slightly different from the inhomogeneous equation (17.313):

$$dy(t) = \alpha(y,t)\,dt + \sigma(y,t)\,dW(t).$$

(state-dependent boundary-value SDE) (17.337)

This is more complicated than our previous problem, since the drift and diffusion coefficients depend on the state of the system. While this equation is straightforward to integrate numerically, it is not at all straightforward to do this integration subject to a bridge condition. This problem could be handled iteratively. For example, start by generating a solution with zero drift and constant diffusion, use this fiducial trajectory to generate the drift and diffusion coefficients, which then generates a new path; continue this process until convergence is reached. That is, *if* convergence is reached. In principle, this is a many-dimensional root-finding procedure, but a more stable method such as Newton iteration can be numerically cumbersome.

We will, however, treat in more detail the somewhat simpler problem of state-dependent Stratonovich diffusion,

$$dy(t) = \sigma(y)\circ dW(t), \tag{17.338}$$

(bridge with state-dependent diffusion)

subject to the closure condition $y(T) = y(0) = 0$. We will show that the solution to this SDE is equivalent to the initial-value SDE

$$dy(t) = \sigma(y)\circ dB(t), \tag{17.339}$$

(solution for state-dependent diffusion)

where $B(t)$ is a standard Brownian bridge, *without* imposing the boundary condition at $t = T$. Since $dB(t)$ has the same *local* statistics (i.e., statistics of increments) as the Wiener process $dW(t)$, we need only verify the closure of the path.

Now we turn to the closure of the solution to the SDE (17.339). First, we rewrite this SDE as

$$\frac{1}{\sigma(y)}\circ dy(t) = dB(t), \tag{17.340}$$

still emphasizing the Stratonovich nature of this SDE. Now for some function $S(y)$, where $y(t)$ is defined by the SDE (17.339), the Stratonovich chain rule (17.200) gives

$$dS(y) = S'(y)\circ dy(t), \tag{17.341}$$

where we treat $dB(t)$ equivalently to $dW(t)$ as far as the chain rule is concerned [$dB(t)$ being a particular realization of $dW(t)$]. Suppose we take

$$S'(y) = \frac{1}{\sigma(y)}, \tag{17.342}$$

or that is, $S(y)$ is the antiderivative of $1/\sigma(y)$. Then Eq. (17.340) becomes

$$S'(y) \circ dy(t) = dB(t). \tag{17.343}$$

Integrating both sides from $t = 0$ to 1 gives

$$S[y(1)] - S[y(0)] = B(1) - B(0) = 0, \tag{17.344}$$

where we have again used the Stratonovich chain rule, so that the left-hand side was just the integral of $dS[y(t)]$. Thus,

$$S[y(1)] = S[y(0)], \tag{17.345}$$

which implies

$$y(1) = y(0) = 0 \tag{17.346}$$

as desired, so long as $S(y)$ is an invertible function. Since it is the antiderivative of $1/\sigma(y)$, this is guaranteed if $\sigma(y)$ is everywhere positive and finite, which is a reasonable restriction on the form of the SDE (17.338). However, the Stratonovich nature of this SDE is crucial: an Itō equation naïvely of the same form has a drift term when converted to a Stratonovich equation, and our solution here does not cover that case.

In fact, the integration procedure applied to arbitrary time t instead of $t = 1$ gives

$$S[y(t)] = B(t), \tag{17.347}$$

where we have chosen $S[y(0)] = 0$. Explicitly, this means that we have chosen

$$S(y) = \int_{y(0)}^{y} \frac{dy'}{\sigma(y')}. \tag{17.348}$$

Then the explicit solution in terms of a Brownian bridge is

$$y(t) = S^{-1}[B(t)],$$

(explicit solution for state-dependent diffusion bridge) (17.349)

where as we have already stated, under reasonable assumptions S^{-1} exists. (For an example solution where $\sigma(y)$ is a step function, see Problem 17.12.)Note that the solutions here can be generalized to the case of a return at time $t = T$ instead of $t = 1$ by the replacements

$$B(t) \longrightarrow \sqrt{T}\, B(t/T), \qquad dB(t) \longrightarrow \sqrt{T}\, dB(t/T) \tag{17.350}$$

in the above solutions.

17.7.4.1 Drift

Returning to the more general SDE (17.337), we see that the presence of a drift term $\alpha(y, t)\, dt$ is more difficult because even when integrating with respect to a bridge, closure of the solution is no longer guaranteed: the drift will in general make the solution ''miss'' the initial point, because in the formal bridge solution (interpreting the SDE as a Stratonovich equation),

$$y(t) = \int_0^t \alpha(y, t')\, dt' + \int_0^t \sigma(y, t') \circ dB(t'), \tag{17.351}$$

the first term is not guaranteed to vanish, and in fact the *second* term is *also* no longer guaranteed to vanish due to the influence of the first term on $y(t)$. Additionally, any explicit time dependence in $\sigma(y, t)$ in general causes the solution (17.339) in the drift-free case to fail. As we already mentioned, an iterative procedure to generate a closed solution may work for this, but in practice iteration tends not to converge well if the SDE coefficients change rapidly with y.

An alternate approach is analogous to ''shooting'' methods for ODEs with boundary conditions. The idea here is to note that while the solution (17.351) may not close when integrated with respect to a a bridge

$B(t)$, it will close for *some* bridge $B_{0\to c}(t)$ connecting 0 to c, as defined by Eq. (17.292). The extra drift compensates for any drift introduced by the SDE coefficients, and the existence of a value c that closes the path is guaranteed provided that the SDE is "reasonable enough" to guarantee continuous solutions (and solutions that vary continuously as the input noise varies). In general, this closing value c must be found numerically, via a root-finding algorithm.

A concern with this method is the *measure* with which we generate the solutions. When we perform the analogous procedure with *additive* noise by closing a Wiener path to form a bridge as in Eq. (17.287), there is no problem: each Wiener path is associated with a unique bridge (with many paths associated with a single bridge), and choosing Wiener paths with the usual measure results in bridges being chosen with the correct measure (all Brownian bridges being equally likely). However, the nonlinear transformation in solving Eq. (17.337), as well as the *unique* association of a bridge $B_{0\to c}(t)$ with each solution $y(t)$, where c is different for each solution, makes things more complicated. In particular, when generating Wiener paths, the relative probability of generating a bridge from 0 to c is the usual probability density for a Wiener path to end up at c after time $t = 1$:

$$P[B_{0\to c}(t)] = \frac{1}{\sqrt{2\pi}}\, e^{-c^2/2}. \tag{17.352}$$

Therefore, if we require a bridge $B_{0\to c_\alpha}(t)$ to generate a particular closed solution $y_\alpha(t)$, the relative probability for this trajectory to occur is given by

$$w_\alpha = \frac{1}{\sqrt{2\pi}}\, e^{-c_\alpha^2/2}. \tag{17.353}$$

Thus, to generate an ensemble of solutions of Eq. (17.337), each generated solution should only be kept with a relative probability w_α (e.g., by the rejection method). Alternately, when computing an ensemble average with respect to the solutions $y_\alpha(t)$, the average should be computed as a *weighted* average, where the weight of each member of the ensemble is w_α. This procedure is valid for Itō or Stratonovich equations, provided that appropriate integration methods are used in each case. Finally, note that this measure business is not a concern for the solutions (17.339), since $c = 0$ for every solution, and thus they all occur with the same probability.

One final possibility for an algorithm in the case where the SDE coefficients are time-independent [i.e., $\alpha(y,t) = \alpha(y)$ and $\sigma(y,t) = \sigma(y)$], and a solution bridging $y(0) = a$ to $y(T) = b$ is as follows.[16] Generate numerical solutions $y_a(t)$ and $y_b(t)$, such that $y_a(0) = a$ and $y_b(0) = b$. If there is a crossing of the two paths at an appropriate time, i.e., $y_a(\tau) = y_b(T-\tau)$ for some t, to within the resolution of the time discretization, then the two paths can be spliced together to realize a bridge (i.e., the solution is $y_a(t)$ for $t < \tau$, and $y_b(t)$ thereafter). If there is no such crossing, the paths are rejected and the process is repeated until successful.

17.7.4.2 Lamperti Transform

The above considerations of rescaling stochastic processes can be elegantly viewed in the framework of the *Lamperti transform*,[17] which is a transformation that equalizes the variances within a stochastic process. This idea applies to SDEs as follows. Consider the Itō SDE

$$dy = \alpha(y)\, dt + \beta(y)\, dW, \tag{17.354}$$

and the transformation

$$z = S(y). \tag{17.355}$$

Then the Itō chain rule (17.193) gives

$$dz = \left[S'(y)\,\alpha + \frac{1}{2}S''(y)\,\beta^2\right] dt + S'(y)\,\beta\, dW, \tag{17.356}$$

[16] Stefano M. Iacus, *Simulation and Inference for Stochastic Differential Equations* (Springer, 2008), Section 2.13.

[17] after J. W. Lamperti, "Semi-stable stochastic processes," *Transactions of the American Mathematical Society* **104**, 62 (1962) (doi: 10.1090/S0002-9947-1962-0138128-7). See Stefano M. Iacus, *op. cit.*, Section 1.11.4.

and the Lamperti transformation obtains by choosing

$$S'(y) = \frac{1}{\beta(y)}, \tag{17.357}$$

or

$$z = S(y) = \int_{y_0}^{y} \frac{dy'}{\beta(y')}. \tag{17.358}$$
(Lamperti transform)

Also using $S''(y) = -\beta'/\beta^2$, Eq. (17.356) then becomes

$$dz = \left(\frac{\alpha}{\beta} - \frac{\beta'}{2}\right) dt + dW, \tag{17.359}$$
(Lamperti-transformed process)

which now is driven by additive noise. The complexity of the multiplicative noise in the dy SDE is thus moved from the stochastic to the deterministic term, and reconstructing the dynamics in z requires inverting $z = S(y)$ in Eq. (17.358), assuming that it is indeed invertible (which happens, for example, if β has the same sign over the domain of y).

Recall from the Itō–Stratonovich conversion (17.184) that a Stratonovich diffusion equation

$$dy = \beta(y) \circ dW \tag{17.360}$$

is equivalent to the Itō equation (17.354) provided $\alpha = \beta'\beta/2$. In Eq. (17.359) this leads to a vanishing drift term, and $dW = dz$, or $z(t) = W(t)$, so that this SDE can be solved in terms of the inversion of the function S—something not in general possible if the drift term remains in Eq. (17.354).

17.7.4.3 Temporal Rescaling

A similar transformation, closer to the original transformation introduced by Lamperti to study self-similar processes[18] is a trajectory-dependent temporal rescaling, with a transformation from time t to t', such that

$$dt = dt'\, \beta^2[y(t)]. \tag{17.361}$$
(temporal rescaling)

In the Itō SDE (17.354), this transformation leads to the equivalent SDE

$$dy(t') = \frac{\alpha(y)}{\beta^2(y)}\, dt' + dW(t'). \tag{17.362}$$
(rescaled Itō SDE)

Note that, as in Eq. (17.359), the variable diffusion rate disappears, and we are left with additive noise. However, the drift term is different, and in particular there is not a second-order Itō correction term.

On the other hand, the Stratonovich SDE (17.360) transforms simply to

$$dy(t') = dW(t'), \tag{17.363}$$
(rescaled Stratonovich SDE)

with no drift, just as in the case of the Lamperti transform. As in our discussion in Section 17.7.4, the interpretation here is simple if we consider Brownian bridges. In the Itō case, if the noise $dW(t)$ is a standard Brownian bridge, then the closure of the bridge is preserved under the temporal rescaling if and only if $\alpha = 0$. On the other hand, the Stratonovich SDE *always* preserves the closure of the path for any temporal rescaling. Note, however, that in the Itō case, even an Itō equation that is equivalent to the Stratonovich equation ($\alpha = \beta'\beta/2$) does not in general preserve path closure under temporal rescaling.

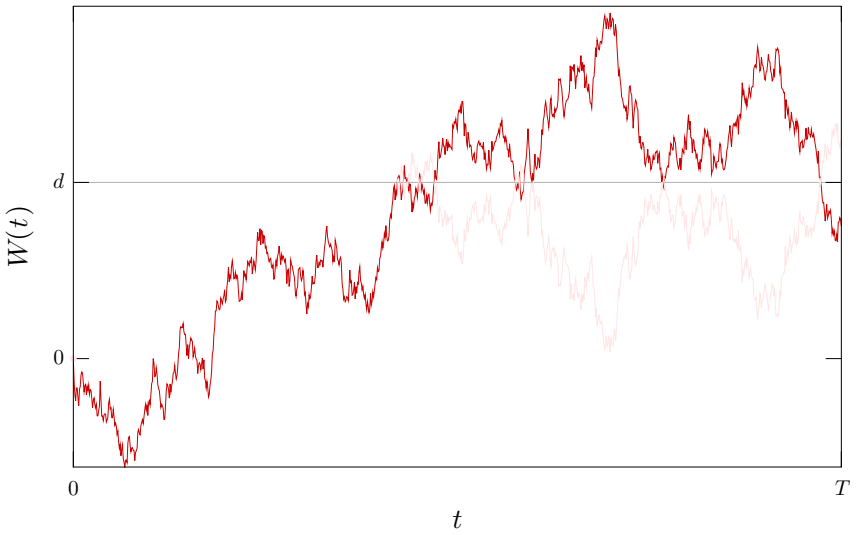
[18] J. W. Lamperti, *op. cit.*; see also Krzysztof Burnecki, Makoto Maejima, and Aleksander Weron "The Lamperti transformation for self-similar processes," *Yokohama Mathematical Journal* **44**, 25 (1997).

17.8 Boundary Crossings

Now we treat the **boundary-crossing problem**: given a continuous random walk $W(t)$ with $W(0) = 0$, what is the probability that $W(t)$ will cross a "boundary" at d after some time t? This is a useful analysis, for example, in financial mathematics, and a wide range of other areas.[19]

17.8.1 Wiener Process

Consider a Wiener path $W(t)$, where as usual $W(0) = 0$. What is the probability that $W(t)$ will cross a threshold $d > 0$ over this time interval from 0 to t? One approach to this problem is via the **Reflection Principle**.[20] The basic idea is illustrated in the plot below of a Wiener path. Once the path hits the threshold at d, it continues to diffuse on. However, note that we can construct an *equally likely* path by mirroring the path after the crossing about d, as shown in the light, mirrored trace. This is true when the probability distribution for the increments is symmetric, which is of course the case for standard Brownian motion. Note that we are also implicitly assuming that there is a well-defined crossing, that is, given that $W(t) > d$, there is some time $t' < t$ such that $W(t') = d$, which is guaranteed by the continuity of the Wiener process.



Now let's apply this to the crossing probability, which we will write

$$P_{\text{cross}}(d,t) = P(\tau_d \le t), \tag{17.364}$$

where τ_d is the time of the first crossing of $W(t)$ through d:

$$\tau_d := \inf\{t : W(t) \ge d\}. \tag{17.365}$$

Then we can partition the probability according to whether $W(t)$ ends up below or above the boundary:

$$P_{\text{cross}}(d,t) = P[\tau_d \le t \wedge W(t) < d] + P[\tau_d \le t \wedge W(t) \ge d]. \tag{17.366}$$

We can write out the first term via conditional probabilities according to $P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$, and in the second term, $\tau_d \le t$ is automatically satisfied if $W(t) \ge d$ by path continuity:

$$P_{\text{cross}}(d,t) = P[W(t) < d \,|\, \tau_d \le t]\, P[\tau_d \le t] + P[W(t) \ge d]. \tag{17.367}$$

[19] David Siegmund, "Boundary crossing probabilities and statistical applications," *The Annals of Statistics* **14**, 361 (1986) (doi: 10.1214/aos/1176349928).

[20] A particularly readable reference for the Reflection Principle appears in Joseph T. Chang, *Stochastic Processes*, available at `http://www.stat.yale.edu/~jtc5/251/stochastic-processes.pdf`. We follow his arguments here for single-boundary crossing probabilities for Wiener paths and bridges. For the Reflection Principle for Wiener paths, see also Kurt Jacobs, *Stochastic Processes for Physicists: Understanding Noisy Systems* (Cambridge, 2010).

According to the Reflection Principle,

$$P[W(t) < d \,|\, \tau_d \leq t] = \frac{1}{2}. \tag{17.368}$$

By construction in the figure, for each path that has crossed through t and ends up with $W(t) < d$, there is an equally likely (mirrored) path with $W(t) > d$. So the probability of ending up above or below d is the same, given that the boundary is crossed. Then we have

$$P_{\text{cross}}(d,t) = \frac{1}{2}P_{\text{cross}}(d,t) + P[W(t) \geq d], \tag{17.369}$$

and solving for $P_{\text{cross}}(d,t)$,

$$P_{\text{cross}}(d,t) = 2P[W(t) \geq d]. \tag{17.370}$$

Now $W(t)$ is normally distributed with variance t, so we have

$$P_{\text{cross}}(d,t) = 2\int_d^\infty dW\, \frac{1}{\sqrt{2\pi t}} e^{-W^2/2t}, \tag{17.371}$$

or finally[21]

$$P_{\text{cross}}(d,t) = \text{erfc}\left(\frac{d}{\sqrt{2t}}\right)$$

(crossing probability of $W(t)$ through d in time t) (17.372)

for the crossing probability past d in time t. Note that this probability converges to unity as $t \longrightarrow \infty$; even though sample paths can start off in the negative direction, given enough time, they will tend to return to the origin and go far enough into the positive direction to cross the boundary anyway. For small t, this probability reduces to

$$P_{\text{cross}}(d,t) \approx e^{-d^2/2t}\sqrt{\frac{2t}{\pi d^2}}, \tag{17.373}$$

so that the crossing probability is exponentially suppressed as $t \longrightarrow 0$.

The result (17.372) can also be interpreted as a cumulative probability distribution for the crossing to occur before time t. Then the probability density for the **first-passage time** τ_d is given by[22]

$$f_{\tau_d}(x) = \partial_t \text{erfc}\left(\frac{d}{\sqrt{2t}}\right)\bigg|_{t=x} = \frac{d}{\sqrt{2\pi x^3}}\, e^{-d^2/2x},$$

(probability density for first-passage time) (17.374)

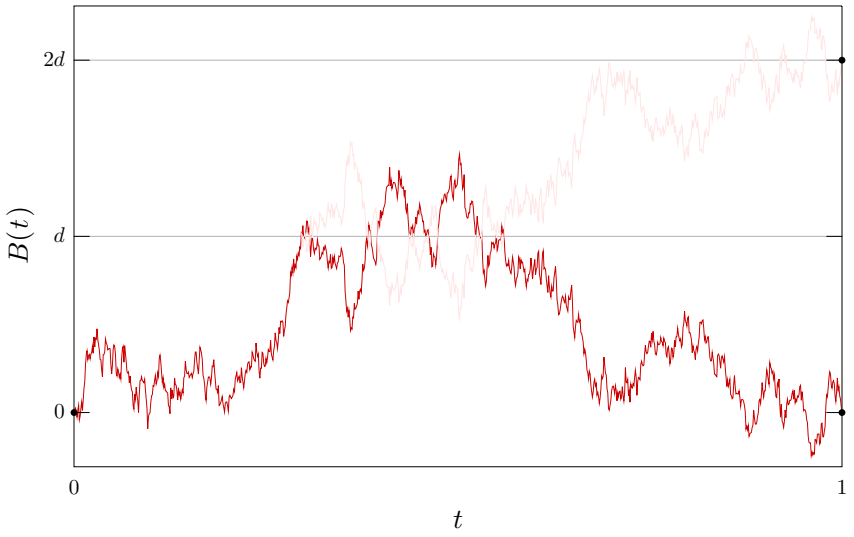
where we have simply differentiated the probability (17.372). Note that the probability density is suppressed exponentially at short times, as $e^{-d^2/2\tau_\text{d}}$, but decays only as $\tau_\text{d}^{-3/2}$ at large times. Thus, this distribution has no mean or variance, meaning that may well take a long time to cross a boundary, although it will almost surely happen. However, the *most likely* value of τ_d (i.e., that maximizes the probability density) is $d^2/3$.

17.8.2 Standard Brownian Bridge

A similar argument works to compute boundary-crossing probabilities for the standard Brownian bridge $B(t)$. Here, however, the Reflection Principle is slightly different. The path crossing the boundary at d must return to zero. The equivalent path that is mirrored after the crossing then must return to $2d$ instead of 0, as shown below.

[21] cf. Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhäuser, 2002), p. 153, formula 1.1.4.

[22] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 198, formula 2.0.2.



Here we will represent the bridge $B(t)$ as an ordinary Wiener path $W(t)$, but subject to the condition $W(1) = 0$. Thus, our crossing probability is the conditional probability

$$P_{\text{cross}}(d) = P[\tau_d \leq 1 \,|\, W(1) = 0]. \tag{17.375}$$

Using again the conditional probability relation in the form to $P(A \wedge B)/P(B) = P(A|B)$, we have

$$P_{\text{cross}}(d) = \frac{P[\tau_d \leq 1 \wedge W(1) = 0]}{P[W(1) = 0]}. \tag{17.376}$$

To compute the numerator,

$$P[\tau_d \leq 1 \wedge W(1) = 0] = P[W(1) = 0 \,|\, \tau_d \leq 1]\, P[\tau_d \leq 1], \tag{17.377}$$

and using the Reflection Principle,

$$\begin{aligned} P[\tau_d \leq t \wedge W(1) = 0] &= P[W(1) = 2d \,|\, \tau_d \leq 1]\, P[\tau_d \leq 1] \\ &= P[W(1) = 2d \wedge \tau_d \leq 1] \\ &= P[W(1) = 2d]. \end{aligned} \tag{17.378}$$

Thus, the crossing probability becomes

$$P_{\text{cross}}(d) = \frac{P[W(1) = 2d]}{P[W(1) = 0]}. \tag{17.379}$$

To interpret these probabilities carefully, the probability $P[W(1) = x]$ refers to the probability of $W(1)$ being between x and $x + dx$; that is, we are referring to probability *densities* here, so that we are taking the ratios of two zero probabilities. Since $P[W(1) = x] = \exp(-x^2/2)\, dx/\sqrt{2\pi}$, then we have[23]

$$P_{\text{cross}}(d) = e^{-2d^2} \tag{17.380}$$

(crossing probability of $B(t)$ through d)

for the crossing probability of the bridge.

Note that the crossing probability is the same as the probability for the peak of the bridge to cross d:

$$P_{\text{cross}}(d) = P\{\sup[B(t)] \geq d\} = 1 - P\{\sup[B(t)] < d\}. \tag{17.381}$$

[23] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 154, formula 1.2.8.

Thus, the probability density function for the maximum of $B(t)$ is

$$f_{\sup\{B(t)\}}(x) = \partial_x P\{\sup[B(t)] < x\} = -\partial_x P_{\text{cross}}(x), \tag{17.382}$$

or[24]

$$f_{\sup\{B(t)\}}(x) = 4xe^{-2x^2} \qquad (x \geq 0)$$

(probability density for Brownian-bridge maximum) (17.383)

We can also compute the moments of this distribution, via

$$\left\langle\!\left\langle [\sup\{B(t)\}]^n \right\rangle\!\right\rangle = \int_0^\infty x^n f_{\sup\{B(t)\}}(x) = \int_0^\infty 4x^{n+1} e^{-2x^2}, \tag{17.384}$$

with the result

$$\left\langle\!\left\langle [\sup\{B(t)\}]^n \right\rangle\!\right\rangle = 2^{-n/2}\Gamma\left(1 + \frac{n}{2}\right),$$

(moments for Brownian-bridge maximum) (17.385)

which gives $(1/4)\sqrt{\pi/2}$, 1/2, $(3/8)\sqrt{\pi/2}$, and 1/2, for $n = 1$, 2, 3, and 4, respectively. Note that the results here are easy to generalize for a Brownian bridge that is pinned such that $B(T) = 0$ instead of $B(1) = 0$, since the pinning essentially rescales the size of the bridge by $\sqrt{T}$, which is equivalent to scaling any distances by $1/\sqrt{T}$. Thus, for example, the crossing probability (17.380) is given by letting $d^2 \longrightarrow d^2/T$, and the peak density (17.383) is given by letting $x^2 \longrightarrow x^2/T$ and $x\,dx \longrightarrow x\,dx/T$ in $f_{\sup\{B(t)\}}(x)\,dx$.

17.8.3 Brownian Bridge

The treatment above for the standard Brownian bridge is easy to generalize to a bridge $B_{0\to c}(t)$ that connects $B_{0\to c}(0) = 0$ to final point $B_{0\to c}(1) = c$, where we obtain

$$P_{\text{cross}}(d) = \frac{P[W(1) = 2d - c]}{P[W(1) = c]} = e^{-(2d-c)^2/2 + c^2/2} \tag{17.386}$$

such that Eq. (17.380) generalizes to

$$P_{\text{cross}}(d, c) = e^{-2d(d-c)} \qquad (d \geq 0, c).$$

(crossing probability of $B_{0\to c}(t)$ through d) (17.387)

Then Eq. (17.383) similarly becomes

$$f_{\sup\{B_{0\to c}(t)\}}(x) = 2(2x - c)e^{-2x(x-c)} \qquad (x \geq 0, c).$$

(probability density for Brownian-bridge maximum) (17.388)

These results can be generalized to a bridge pinned to c at time t by scaling $d \longrightarrow d/\sqrt{t}$ and $c \longrightarrow c/\sqrt{t}$.

17.8.4 First-Passage Time of the Brownian Bridge

In analyzing the crossing probability for the Wiener path, we were able to simply derive the result (17.374) for the density of the first-passage time τ_d. It is only slightly more complicated to do this for a Brownian bridge, so we will carry out the derivation here. For the Brownian bridge, we can define the first-passage time as

$$\tau_d := \inf\{\tau \leq t : B_{0\to c}(\tau) \geq d\}, \tag{17.389}$$

where $d \geq 0$ and the running time of the bridge is t, such that $B_{0\to c}(t) = c$. Then the crossing probability up to time τ (*before* the final time t) is

$$P_{\text{cross}}(d, \tau \leq t) = P[\tau_d \leq \tau \leq t | W(t) = c], \tag{17.390}$$

[24] cf. Eq. (2) in Jim Pitman and Marc Yor, "On the distribution of ranked heights of excursions of a Brownian bridge," *Annals of Probability* **29**, 361 (2001) (doi: 10.1214/aop/1008956334).

when written in terms of a Wiener path. We don't know this crossing probability, but from Eq. (17.387) we *do* know the crossing time of the bridge up to the final running time t. To rewrite Eq. (17.390) in terms of this crossing probability, we can proceed as follows:

$$\begin{aligned} P_{\text{cross}}(d, \tau \le t) &= \sqrt{2\pi t}\, e^{c^2/2t}\, P[\tau_d \le \tau \le t \wedge W(t) = c] \\ &= \sqrt{2\pi t}\, e^{c^2/2t} \int_{-\infty}^{\infty} dz\, P[\tau_d \le \tau \le t \wedge W(\tau) = z]\, P[W(t) = c | W(\tau) = z]. \end{aligned} \tag{17.391}$$

The second factor in the integrand is just a Gaussian probability density,

$$P[W(t) = c | W(\tau) = z] = \frac{1}{\sqrt{2\pi(t-\tau)}}\, e^{-(c-z)^2/2(t-\tau)}, \tag{17.392}$$

and the first factor is given by Eq. (17.387) as

$$\begin{aligned} P[\tau_d \le \tau \le t \wedge W(\tau) = z] &= \frac{1}{\sqrt{2\pi\tau}}\, e^{-z^2/2\tau} P[\tau_d \le \tau \le t | W(\tau) = z] \\ &= \frac{1}{\sqrt{2\pi\tau}}\, e^{-z^2/2\tau} e^{-2d(d-z)/\tau} \\ &= \frac{1}{\sqrt{2\pi\tau}}\, e^{-(2d-z)^2/2\tau} \qquad (z \le d) \end{aligned} \tag{17.393}$$

for the case where the intermediate point z is *not* past the boundary d, and

$$\begin{aligned} P[\tau_d \le \tau \le t \wedge W(\tau) = z] &= \frac{1}{\sqrt{2\pi\tau}}\, e^{-z^2/2\tau} P[\tau_d \le \tau \le t | W(\tau) = z] \\ &= \frac{1}{\sqrt{2\pi\tau}}\, e^{-z^2/2\tau} \qquad (z \ge d) \end{aligned} \tag{17.394}$$

for the case where the path has already crossed the boundary by time τ. Putting these expressions into Eq. (17.391) and carrying out the resulting two integrals for the cases $z \le d$ and $z \ge d$, we obtain the crossing probability

$$P_{\text{cross}}(d, \tau \le t) = \frac{1}{2} \operatorname{erfc}\left(\frac{\tau(d-c) + d(t-\tau)}{\sqrt{2t\tau(t-\tau)}} \right) + \frac{1}{2}\, e^{-2d(d-c)/t} \operatorname{erfc}\left(\frac{\tau(c-d) + d(t-\tau)}{\sqrt{2t\tau(t-\tau)}} \right).$$

(first-passage cumulative probability, Brownian bridge) (17.395)

To check the normalization of this probability note that $\operatorname{erfc}(x) \longrightarrow 0$ as $x \longrightarrow \infty$ and $\operatorname{erfc}(x) \longrightarrow 2$ as $x \longrightarrow -\infty$. Then as $\tau \longrightarrow t$, Eq. (17.395) becomes

$$P_{\text{cross}}(d, t) = \Theta(c-d) + e^{-2d(d-c)/t}\, \Theta(d-c). \tag{17.396}$$

which gives the correct crossing probability from Eq. (17.387). The probability density for the first-passage time is then simply given by differentiating with respect to τ:[25]

$$f_{\tau_d}(x) = \frac{\sqrt{t}}{\sqrt{2\pi x^3 (t-x)}}\, d\, e^{-(d\,t - cx)^2/2tx(t-x)}.$$

(probability density for Brownian-bridge first-passage time) (17.397)

Note that in the limit of large t, this reduces to the Wiener-process result (17.374), as it should. This expression is also unnormalized if $c < d$: in this case the normalization is given by Eq. (17.396), since the first-passage time is undefined in the event that a path does not cross the boundary.

[25] L. Beghin and E. Orsingher, "On the maximum of the generalized Brownian bridge," *Lithuanian Mathematical Journal* **39**, 157 (1999) (doi: 10.1007/BF02469280), Eq. (2.15).

The moments of the first-passage time can be written in terms of the density (17.397) as

$$\left\langle\!\left\langle \tau_d^{\,n} \right\rangle\!\right\rangle = \sqrt{\frac{t}{2\pi}}\, d \int_0^t dx\, \frac{x^{n-3/2}}{\sqrt{t-x}}\, e^{-(d\,t-cx)^2/2tx(t-x)}. \tag{17.398}$$

Changing variables to $y = t/x - 1$ gives the alternate expression

$$\left\langle\!\left\langle \tau_d^{\,n} \right\rangle\!\right\rangle = \frac{t^{n-1/2} d}{\sqrt{2\pi}} \int_0^\infty dy\, \frac{(1+y)^{-n}}{\sqrt{y}}\, e^{-[(1+y)d-c]^2/2ty}. \tag{17.399}$$

For $n = 0$, this expression gives the correct normalization

$$\left\langle\!\left\langle \tau_d^{\,0} \right\rangle\!\right\rangle = e^{-d(d-c+|d-c|)/t}. \tag{17.400}$$

The integral is easier to evaluate for $n < 0$ than $n > 0$. For example

$$\left\langle\!\left\langle \tau_d^{\,-1} \right\rangle\!\right\rangle = \frac{d|d-c| + t + d^2}{td^2}\, e^{-d(d-c+|d-c|)/t} \tag{17.401}$$

gives the first inverse moment—the well-defined value here and of the other inverse moments is an indication of how heavily the value $\tau_d = 0$ is suppressed.

The quantity $(t - \tau_d)$ is the time *after* the first passage time. In the case $c = 0$, for a closed bridge, this is also the statistic for the *last* passage time, because of the time-reversal symmetry of the bridge. Moments of the last passage time can be calculated in the same way as the regular moments. Adapting the same variable change leading to Eq. (17.399), the post-first-passage-time moments are

$$\left\langle\!\left\langle (t-\tau_d)^n \right\rangle\!\right\rangle = \frac{t^{n-1/2} d}{\sqrt{2\pi}} \int_0^\infty dy\, y^{n-1/2} (1+y)^{-n}\, e^{-[(1+y)d-c]^2/2ty}. \tag{17.402}$$

Then, for example, the first inverse moment is given by

$$\left\langle\!\left\langle (t-\tau_d)^{-1} \right\rangle\!\right\rangle = \frac{d + |d-c|}{|d-c|t}\, e^{-d(d-c+|d-c|)/t}. \tag{17.403}$$

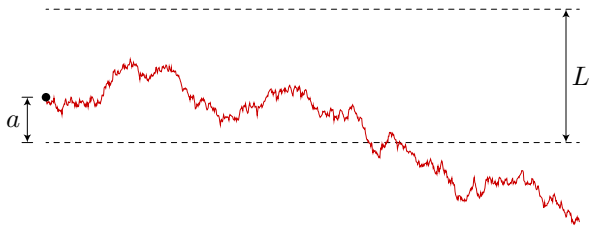
Note that, in the case of $c = 0$, that the mean of the inverse *fractional* last-passage time $t/(t - \tau_d)$ happens to have the same form as the boundary-crossing probability $e^{-2d^2/t}$ from Eq. (17.380), but with an extra factor of 2 (recall that this applies to the case where paths that do *not* cross the boundary count as zero in this ensemble average).

17.9 Escape Probability

Similar to the boundary-crossing problem is the *escape* problem, which is concerned whether a stochastic process leaves an *interval*. There are implicitly two boundaries involved, and the idea is to see whether the process touches *either* boundary. [This is closely related to whether the process touches *both* boundaries, and we can also calculate this via $P(A \wedge B) = P(A) + P(B) - P(A \vee B)$, where the individual probabilities $P(A)$ and $P(B)$ are given by the appropriate single-boundary-crossing probabilities.]

17.9.1 Wiener Process

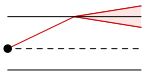
We will set up the problem as follows: a Wiener process $W(t)$ begins between two barriers separated by distance L, and the starting point is a distance a from one of the barriers. For concreteness, we take the Wiener process' starting point to be distance a from the *lower* barrier, and with $W(0) = 0$, the barriers define the interval $[-a, L - a]$.



Of course, we would obtain the same answers by instead using the interval $[a - L, a]$. The question is, what is the probability to touch either boundary, and thus to escape the interval, in time t?

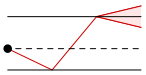
We will approach this problem in the same way as the boundary-crossing, making use of the boundary-crossing probability (17.372) in the process.[26] Actually we will first consider a slightly different problem, which is, what is the probability that $W(t)$ touches the *upper* boundary *first*? That is, we only count the escapes where the *first* escape is through the upper boundary. To compute this, we will define some events (sets of outcomes). First, we define event U_1 to be set of all outcomes where the process touches the upper boundary:

U_1 :



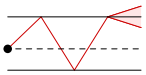
We are illustrating the trajectory schematically here; the trajectory may be much more complicated and touch either boundary many more times than we have indicated. The conical "spray" of trajectories to the right indicates that we don't particularly care what happens to the trajectory afterwards. Now we are only interested in the cases where the process touches the upper boundary first, but we have included cases where the process touches the lower boundary and *then* the upper boundary, since U_1 includes *any* outcome that touches the upper boundary. We will denote this set L_1:

L_1 :



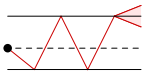
To properly count the events we want, we should delete events in L_1. But not *all* of them! In L_1 we included paths that touch the upper boundary before touching the lower boundary, and we want to count these. We will denote this set U_2:

U_2 :



But again, in this set, we are counting paths that touch the lower boundary before the indicated touchings, and we don't want to count these. We will denote this set L_2.

L_2 :



Continuing in this way, we should define the set U_j to be the set of all paths that touch the upper boundary j times, with $j - 1$ touchings of the lower boundary "interleaved," and L_j to be the set of all paths that touch the lower boundary and then alternating between the upper boundary and lower boundary, with j touchings of each boundary. (Once a boundary touches a boundary, it is okay for it to touch it again before touching the other boundary in these definitions.)

Thus, we have argued that the set of all paths that touch the upper boundary *first* is

$$A_{\text{upper first}} = U_1 - L_1 + U_2 - L_2 + \ldots \tag{17.404}$$

The probabilities of the events on the right-hand side are easy to compute, using the Reflection Principle. (Indeed, notice how the "reflections" pop up here, in a way analogous to the infinity of reflections in a Fabry–Perot cavity.) The idea as before is to "unwrap" trajectories via reflections, and the resulting probability

[26] The method here was applied to the Wiener process in J. L. Doob, "Heuristic Approach to the Kolmogorov–Smirnov Theorems," *The Annals of Mathematical Statistics* **20**, 393 (1949) (doi: 10.1214/aoms/1177729991); and T. W. Anderson, "A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size," *The Annals of Mathematical Statistics* **31**, 165 (1960) (doi: 10.1214/aoms/1177705996).

is just the single-boundary crossing probability (17.372), where the distance d is the total vertical distance traversed in each diagram (noting that the distances to the boundaries are $L-a$ and a above and below the dashed line, respectively. Thus, for example,

$$\begin{aligned} P(U_1) &= \operatorname{erfc}\left(\frac{L-a}{\sqrt{2t}}\right) \\ P(L_1) &= \operatorname{erfc}\left(\frac{L+a}{\sqrt{2t}}\right) \\ P(U_2) &= \operatorname{erfc}\left(\frac{3L-a}{\sqrt{2t}}\right) \\ P(L_2) &= \operatorname{erfc}\left(\frac{3L+a}{\sqrt{2t}}\right), \end{aligned} \tag{17.405}$$

and so on. The probability to touch the upper boundary before the lower boundary is then

$$\begin{aligned} P_{\text{upper first}}(t) &= \operatorname{erfc}\left(\frac{L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{L+a}{\sqrt{2t}}\right) + \operatorname{erfc}\left(\frac{3L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{3L+a}{\sqrt{2t}}\right) + \dots \\ &= \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{(2j-1)L-a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{(2j-1)L+a}{\sqrt{2t}}\right)\right]. \end{aligned} \tag{17.406}$$

The probability to touch the *lower* boundary *before* the upper boundary is simply given by the replacement $a \longrightarrow L-a$ in the above expression:

$$\begin{aligned} P_{\text{lower first}}(t) &= \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{(2j-2)L+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{(2j)L-a}{\sqrt{2t}}\right)\right] \\ &= \operatorname{erfc}\left(\frac{a}{\sqrt{2t}}\right) + \sum_{j=1}^{\infty}\left[\operatorname{erfc}\left(\frac{2jL+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{2jL-a}{\sqrt{2t}}\right)\right]. \end{aligned} \tag{17.407}$$

The escape probability is the sum of the above two probabilities,

$$P_{\text{escape}}(t) = P_{\text{upper first}}(t) + P_{\text{lower first}}(t), \tag{17.408}$$

since they represent two disjoint sets of outcomes:[27]

$$\begin{aligned} P_{\text{escape}}(t) &= \operatorname{erfc}\left(\frac{a}{\sqrt{2t}}\right) + \sum_{j=1}^{\infty}(-1)^j\left[\operatorname{erfc}\left(\frac{jL+a}{\sqrt{2t}}\right) - \operatorname{erfc}\left(\frac{jL-a}{\sqrt{2t}}\right)\right] \\ &= \sum_{j=-\infty}^{\infty}(-1)^j \operatorname{sgn}(j+0^+)\operatorname{erfc}\left(\frac{|a+jL|}{\sqrt{2t}}\right). \end{aligned}$$

(escape probability for Wiener process) (17.409)

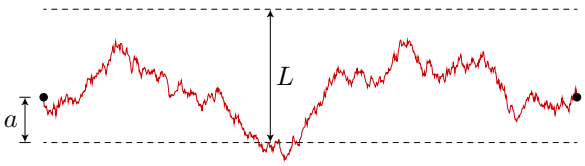
Note that the 0^+ is included in the *sgn* function so that the $j=0$ term is positive.

17.9.2 Standard Brownian Bridge

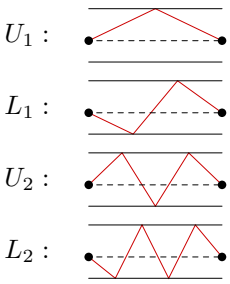
The calculation for the Brownian bridge goes in essentially the same was as for the Wiener path.[28] We set up the problem with the same dimensions as before, the only difference being the constraint $B(1)=0$ on the Wiener path $B(t)$.

[27] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 167, formula 1.7.4(2), and see p. 641 for the function definition.

[28] Bruno Casella and Gareth O. Roberts, "Exact Monte Carlo Simulation of Killed Diffusions," *Advances in Applied Probability* **40**, 273 (2008) (doi: 10.1239/aap/1208358896). (Note the typo in the expression for q_j.) Also Alexandros Beskos, Stefano Peluchetti, and Gareth Roberts, "ϵ-Strong Simulation of the Brownian Path," arXiv.org preprint (arXiv: 1110.0110v1).



Again, we consider the altered problem of: what is the probability that $W(t)$ touches the upper boundary first? Again, we define events U_j and L_j as in the Wiener-path case, except that now the final points of the paths are pinned down:



and so on. Continuing in this way, we again define the set U_j to be the set of all paths that touch the upper boundary j times, with $j-1$ touchings of the lower boundary "interleaved," and L_j to be the set of all paths that touch the lower boundary and then alternating between the upper boundary and lower boundary, with j touchings of each boundary.

Thus, the set of all paths that touch the upper boundary first is again

$$A_{\text{upper first}} = U_1 - L_1 + U_2 - L_2 + \dots \tag{17.410}$$

The probabilities of the events on the right-hand side are easy to compute, using the Reflection Principle. The probability in each diagram is just the single-boundary (bridge) crossing probability (17.380), where the distance d is *half* the total vertical distance traversed in each diagram. Thus, for example,

$$\begin{aligned} P(U_1) &= e^{-2(L-a)^2} \\ P(L_1) &= e^{-2L^2} \\ P(U_2) &= e^{-2(2L-a)^2} \\ P(L_2) &= e^{-2(2L)^2}, \end{aligned} \tag{17.411}$$

and so on. The probability to touch the upper boundary before the lower boundary is then

$$\begin{aligned} P_{\text{upper first}} &= e^{-2(L-a)^2} - e^{-2L^2} + e^{-2(2L-a)^2} - e^{-2L^2} + \dots \\ &= \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2} - e^{-2(jL)^2} \right]. \end{aligned} \tag{17.412}$$

The probability to touch the lower boundary before the upper boundary is again simply given by the replacement $a \longrightarrow L - a$ in the above expression:

$$\begin{aligned} P_{\text{lower first}} &= \sum_{j=1}^{\infty} \left[e^{-2[(j-1)L+a)^2} - e^{-2(jL)^2} \right] \\ &= e^{-2a^2} + \sum_{j=1}^{\infty} \left[e^{-2(jL+a)^2} - e^{-2(jL)^2} \right] \end{aligned} \tag{17.413}$$

The escape probability is once again the sum of the above two probabilities,

$$P_{\text{escape}} = P_{\text{upper first}} + P_{\text{lower first}}. \tag{17.414}$$

The resulting expression is

$$\begin{aligned} P_{\text{escape}} &= e^{-2a^2} + \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2} + e^{-2(jL+a)^2} - 2e^{-2(jL)^2} \right] \qquad (0 < a < L) \\ &= 1 + \sum_{j=-\infty}^{\infty} \left[e^{-2(a+jL)^2} - e^{-2(jL)^2} \right]. \end{aligned}$$

(escape probability, standard Brownian bridge) (17.415)

Notice that besides the obvious difference in the functions appearing here compared to the Wiener case, the *structure* is different: an a-independent term (the last term here) does not appear in the Wiener case.

17.9.3 Brownian Bridge

It is not difficult to generalize the above escape probability to a more general Brownian bridge $B_{0\to c}(t)$, pinned to c at $t = 1$. In the derivation, we summed over terms of the form $\exp(-2d^2)$ from Eq. (17.380) for various distances d. According to (17.387) we just need to change these to terms of the form $\exp[-2d(d-c)]$, with

$$P_{\text{escape}} = e^{-2a(a-c)} + \sum_{j=1}^{\infty} \left[e^{-2(jL-a)(jL-a-c)} + e^{-2(jL+a)(jL+a-c)} - 2e^{-2(jL)(jL-c)} \right] \qquad (0 < a < L;\ -a < c < L-a)$$

(escape probability, Brownian bridge) (17.416)

as the result.

17.10 Dirichlet Problem and Connection to Electrostatics

17.10.1 Laplace Equation

Another interesting context in which boundary crossing and escape arises is in the Laplace equation, subject to Dirichlet boundary conditions. That is, suppose we want the solution to the Laplace equation

$$\nabla^2 \phi(\mathbf{r}) = 0, \qquad \phi(\mathbf{r}) = \phi_\partial(\mathbf{r}) \quad \forall_{\mathbf{r}\in\partial D}. \tag{17.417}$$

(Dirichlet problem)

on some bounded domain D, where $\phi_\partial(\mathbf{r})$ fixes the solution on the domain boundary ∂D. (That is, find the electrostatic potential in a charge-free region, given the potential/voltage on a bounding surface.)

We will now show that the solution can be written as the path average

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_\partial)] \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_\partial)}, \tag{17.418}$$

(stochastic solution)

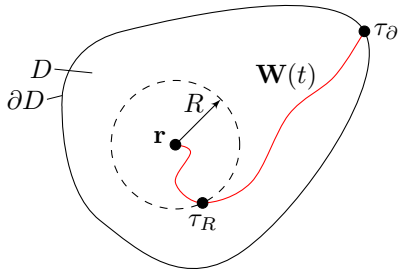
where $\mathbf{W}(t)$ is a vector Wiener process, and τ_∂ is the first passage time of $\mathbf{W}(t)$ through ∂D. That is, we start a bunch of Wiener paths from $\mathbf{r}$, let them go until they hit the boundary, and then take the average of the boundary values where the paths hit. Actually, this solution requires that the boundary be sufficiently nice, which is certainly true for physical boundaries in electrostatic problems. (More technically, the surface should satisfy the "Poincaré cone condition," which basically says that at each point on the surface, you can attach a code of finite angle and length that doesn't intersect the boundary except at the attachment point. This rules out, for example, interior boundaries of arbitrarily small area, or a sufficiently severe "kink" in the surface.[29])

That the stochastic solution (17.418) has the correct boundary values is reasonably obvious, because as the point $\mathbf{r}$ moves close to a point on the boundary, the paths starting from that point will hit the

[29] Kiyosi Itô and Henry P. McKean, Jr., *Diffusion Processes and their Sample Paths* (Springer, 1974), pp. 257, 261-4 (doi: 10.1007/978-3-642-62025-6).

nearest surface with probability approaching unity. This can be seen, for example, from the first-passage-time density for the Wiener path in Eq. (17.374), where the peak of the density shifts to zero as $d \longrightarrow 0$, or that the crossing probability (17.372) in any finite time converges to unity as $d \longrightarrow 0$. However, the boundary function ϕ_∂ should be continuous, as the paths from $\mathbf{r}$ will always average over a small region of the boundary, even as the source point approaches the boundary.

Now to show that the expression (17.418) satisfies the Laplace equation. Suppose we draw a sphere of radius R, centered at $\mathbf{r}$, such that the sphere lies entirely within D, as shown below.



The idea is that the Wiener path must cross the sphere before it crosses the boundary. Then suppose we rewrite Eq. (17.418) as

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \left\langle\!\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_R) + \Delta\mathbf{W}(\tau_\partial)] \right\rangle\!\!\right\rangle_{\Delta\mathbf{W}(\tau_\partial)} \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}, \tag{17.419}$$

where τ_R is the first crossing time of the sphere, and

$$\Delta\mathbf{W}(\tau_\partial) := \mathbf{W}(\tau_\partial) - \mathbf{W}(\tau_R). \tag{17.420}$$

We haven't really done much here, except to split (pathwise) the time interval into pre- and post-τ_R, and we are explicitly taking the ensemble average after τ_R separately from the ensemble average over all possible first crossings of the sphere represented by τ_R. However, since the path $\Delta\mathbf{W}(\tau_\partial)$ after touching the sphere acts itself like a Wiener path, we can use the solution (17.418) to replace the inner ensemble average:

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_R)] \right\rangle\!\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{17.421}$$

Now since the Wiener path is equally likely to have its first touching point $\mathbf{W}(\tau_R)$ at any point on the sphere, we can simplify the notation a bit to write

$$\phi(\mathbf{r}) = \left\langle\!\!\left\langle \phi(\mathbf{r} + \mathbf{R}) \right\rangle\!\!\right\rangle_{|\mathbf{R}|=R}, \tag{17.422}$$

where $\mathbf{R}$ determines some point on the sphere, and the ensemble average is simply a uniform average over the surface of the sphere. That $\phi(\mathbf{r})$ is the average value of ϕ on a sphere centered on $\mathbf{r}$ is a necessary and sufficient condition for ϕ to be a **harmonic function** (i.e., a solution to the Laplace equation).[30]

17.10.2 Laplacian as Spherical Average

We can make the above argument about the averaging property of harmonic functions more precise by working it out somewhat more explicitly, and in the process arrive at a useful representation of the Laplacian. First, consider the Taylor expansion

$$\phi(\mathbf{r} + \mathbf{R}) = \phi(\mathbf{r}) + \big[\partial_\alpha \phi(\mathbf{r})\big] R_\alpha + \frac{1}{2}\big[\partial_\alpha \partial_\beta \phi(\mathbf{r})\big] R_\alpha R_\beta + O(R^3), \tag{17.423}$$

[30] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice Hall, 2013), Section 3.1.4, p. 117. The idea is to consider a point charge outside the sphere, and show that the averaging statement is true for this case. Then the same statement must be true for any collection of charges, and conversely for any solution of the Laplace equation, since any physical solution may be regarded as being produced by some charge distribution.

where repeated indices are summed. Averaging over all orientations of $\mathbf{R}$ at fixed distance R gives

$$\left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} = \phi(\mathbf{r}) + \frac{1}{2}\left\langle\!\left\langle \left[\partial_\alpha^2 \phi(\mathbf{r})\right] R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} + O(R^3), \tag{17.424}$$

where all first derivatives have vanished under the symmetric average. Now we can rewrite the last term using

$$\begin{aligned} \left\langle\!\left\langle \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} &= \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] \left\langle\!\left\langle R_\alpha^2 \right\rangle\!\right\rangle_{|\mathbf{R}|=R} \\ &= \sum_{\alpha=1}^{d} \left[\partial_\alpha^2 \phi(\mathbf{r})\right] \frac{R^2}{d} \\ &= \frac{R^2}{d} \nabla^2 \phi(\mathbf{r}), \end{aligned} \tag{17.425}$$

where we are writing out the sum explicitly now over d dimensions, and we used the fact that R_α^2 is independent of α once averaged over all orientations. Then putting this result into Eq. (17.424) and taking the limit $R \longrightarrow 0$

$$\nabla^2 \phi(\mathbf{r}) = \lim_{R \to 0^+} \frac{2d}{R^2} \left[\left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} - \phi(\mathbf{r}) \right],$$

(Laplacian as spherical average) (17.426)

which is a representation of the Laplacian operator in terms of an average over a small sphere around $\mathbf{r}$. This equation immediately implies that Eq. (17.422) is equivalent to the Laplace equation (17.417), completing our proof of the Wiener-path solution.

17.10.3 Poisson Equation

The same basic approach[31] works for the Poisson equation

$$\nabla^2 \phi(\mathbf{r}) = \rho(\mathbf{r}), \qquad \phi(\mathbf{r}) = \phi_\partial(\mathbf{r}) \quad \forall_{\mathbf{r} \in \partial D}, \tag{17.427}$$

(Dirichlet Poisson problem)

which is the same as the original problem (17.417), but with the addition of a source $\rho(\mathbf{r})$ (a factor of $1/\epsilon_0$ is absorbed into the source for simplicity). The solution is the same as before, but with the addition of a source-averaging term:

$$\phi(\mathbf{r}) = \left\langle\!\left\langle \phi[\mathbf{r} + \mathbf{W}(\tau_\partial)] - \frac{1}{2} \int_0^{\tau_\partial} dt\, \rho[\mathbf{r} + \mathbf{W}(t)] \right\rangle\!\right\rangle_{\mathbf{W}(\tau_\partial)}.$$

(stochastic Poisson solution) (17.428)

The same argument leading to Eq. (17.419) applies, but the second term splits into two parts, before and after τ_R. The result is then

$$\phi(\mathbf{r}) = \left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} - \left\langle\!\left\langle \frac{1}{2} \int_0^{\tau_R} dt\, \rho[\mathbf{r} + \mathbf{W}(t)] \right\rangle\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{17.429}$$

The portion of the integral from τ_R to τ_∂ was already absorbed into the solution in the first term here. Rearranging and multiplying by $2d/R^2$,

$$\frac{2d}{R^2} \left[\left\langle\!\left\langle \phi(\mathbf{r}+\mathbf{R}) \right\rangle\!\right\rangle_{|\mathbf{R}|=R} - \phi(\mathbf{r}) \right] = \frac{d}{R^2} \left\langle\!\left\langle \int_0^{\tau_R} dt\, \rho[\mathbf{r} + \mathbf{W}(t)] \right\rangle\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{17.430}$$

[31] For a rigorous version of the argument here, see Sidney C. Port and Charles J. Stone, *Brownian Motion and Classical Potential Theory* (Academic Press, 1978) (ISBN: 0124335942), Proposition 5.2 on p. 14 and Section 4.6 on p. 114.

Taking the limit $R \longrightarrow 0$, we can use Eq. (17.426) on the left-hand side, and on the right-hand side, we can treat ρ as being a constant with respect to the integral. Thus

$$\nabla^2 \phi(\mathbf{r}) = \rho(\mathbf{r}) \frac{d}{R^2} \left\langle\!\left\langle \tau_R \right\rangle\!\right\rangle_{\mathbf{W}(\tau_R)}. \tag{17.431}$$

At this point, we have dropped the limit $R \longrightarrow 0$ on the right-hand side, as it is not necessary. The statistic involved here is the mean of the first-passage time through a sphere of radius R in d dimensions.

Now to calculate the remaining expectation value. First, recall that for a vector Wiener process in d dimensions, the Euclidean norm is given on average by

$$\left\langle\!\left\langle \|\mathbf{W}(t)\|^2 \right\rangle\!\right\rangle = td, \tag{17.432}$$

since there are d independent directions of displacement, each contributing t to the variance. Now what we would like to show is that this is still true if t is replaced by the stopping time τ_R, in the sense that

$$\left\langle\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle = \left\langle\!\left\langle \tau_R \right\rangle\!\right\rangle d, \tag{17.433}$$

or indeed any other stopping time. First, let's take t to be some very large time, such that almost certainly $\tau_R < t$ (i.e., $0 < \tau_R < t$). For some paths this will not be true, such that the argument below will miss them; in these cases we can take τ_R to be equal to t, which will produce an error that vanishes in the limit $t \longrightarrow \infty$. Then starting with Eq. (17.432),

$$\begin{aligned} td &= \left\langle\!\left\langle \|\mathbf{W}(t)\|^2 \right\rangle\!\right\rangle \\ &= \left\langle\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R) + \mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle \\ &= \left\langle\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle + \left\langle\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle - \left\langle\!\left\langle [\mathbf{W}(t) - \mathbf{W}(\tau_R)] \cdot \mathbf{W}(\tau_R) \right\rangle\!\right\rangle, \end{aligned} \tag{17.434}$$

where the last term is zero, because the parts of the path $\mathbf{W}(t)$ before and after τ_R are independent, and at least the first factor has zero mean. Then taking on the first term on the right-hand side, we can think of all paths that start at a particular stopping point $\mathbf{W}(\tau_R)$, such that the ensemble average of $\|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|$ is just $(t - \tau_R)d$. Then continue the ensemble average over all stopping times τ_R, so that

$$\left\langle\!\left\langle \|\mathbf{W}(t) - \mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle = \left\langle\!\left\langle (t - \tau_R) d \right\rangle\!\right\rangle = td - \left\langle\!\left\langle \tau_R \right\rangle\!\right\rangle d. \tag{17.435}$$

Putting this together with Eq. (17.434), we find the desired result

$$\left\langle\!\left\langle \|\mathbf{W}(\tau_R)\|^2 \right\rangle\!\right\rangle = \left\langle\!\left\langle \tau_R \right\rangle\!\right\rangle d, \tag{17.436}$$

of course taking the limit $t \longrightarrow \infty$ (which only appears in the sense that we forced $\tau_R \leq t$) so that the ensemble averages involving τ_R are correct. By definition of τ_R, the left-hand side is R^2, so

$$\left\langle\!\left\langle \tau_R \right\rangle\!\right\rangle = \frac{R^2}{d}. \tag{17.437}$$

Putting this into Eq. (17.430), we see that it reduces to the Poisson equation (17.427) as desired.

17.11 Feynman–Kac Formula

To introduce a very powerful method of computing expectation values for SDEs, we will introduce the **Feynman–Kac formula**,[32] which solves diffusion problems in terms of integrals over solutions to SDEs.

[32] R. P. Feynman, "Space-Time Approach to Non-Relativistic Quantum Mechanics," *Reviews of Modern Physics* **20**, 367 (1948) (doi: 10.1103/RevModPhys.20.367); M. Kac, "On Distributions of Certain Wiener Functionals," *Transactions of the American Mathematical Society* **65**, 1 (1949) (doi: 10.1090/S0002-9947-1949-0027960-X); M. Kac, "On Some Connections between Probability Theory and Differential and Integral Equations," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, 1951), p. 189 (`http://projecteuclid.org/euclid.bsmsp/1200500229`).

There are numerous forms of this formula, and we will develop a relatively simple form[33] that considers a forced diffusion equation for the distribution $f(x,t)$,

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x,t)f + g(x,t), \tag{17.438}$$
(PDE for Feynman–Kac formula)

subject to the initial condition

$$f_0(x) = f(x,0). \tag{17.439}$$
(initial condition for Feynman–Kac formula)

The Feynman–Kac formula gives the solution of (17.438) for $t > 0$ as

$$f(x,t) = \left\langle\!\!\left\langle f_0[x+W(t)]\,\exp\!\left(-\int_0^t dt'\,V[x+W(t'),t-t']\right)\right.\right.$$
$$\left.\left. + \int_0^t dt'\,g[x+W(t'),t-t']\,\exp\!\left(-\int_0^{t'} dt''\,V[x+W(t''),t-t'']\right)\right\rangle\!\!\right\rangle, \tag{17.440}$$
(Feynman–Kac formula)

where the ensemble average is over all realizations of $W(t)$. Before continuing to prove this formula, notice that it reduces correctly to $f_0(x)$ as $t \longrightarrow 0$. Also, for a simple, undriven diffusion,

$$\partial_t f = \frac{1}{2}\partial_x^2 f, \tag{17.441}$$

the formula reduces to

$$\begin{aligned} f(x,t) &= \left\langle\!\!\left\langle f_0[x+W(t)]\right\rangle\!\!\right\rangle \\ &= \int_{-\infty}^{\infty} dW\,f_0(W)\,\frac{1}{\sqrt{2\pi t}}e^{-(W-x)^2/2t}, \end{aligned} \tag{17.442}$$

which is the convolution of $f_0(x)$ with a Gaussian of variance t, as we expect. The extra terms V and g introduce damping (or "killing" of diffusing particles) and particle sources, respectively, that clearly make the solution more complicated.

17.11.1 Proof: Simple Diffusion

But now to prove the Feynman–Kac formula (17.440), which we will do in increasingly complex stages. First, take again the simple case where $V = g = 0$, so that we have the simple diffusion equation,

$$\partial_t f = \frac{1}{2}\partial_x^2 f, \tag{17.443}$$

with solution

$$f(x,t) = \left\langle\!\!\left\langle f_0[x+W(t)]\right\rangle\!\!\right\rangle. \tag{17.444}$$

We will show explicitly that this is the solution by simply differentiating it with respect to time. Technically, we will just compute the differential $df(t)$, regarding x as a fixed parameter,

$$\partial_t f\,dt = \left\langle\!\!\left\langle \partial_x f_0[x+W(t)]\,dW\right\rangle\!\!\right\rangle + \frac{1}{2}\left\langle\!\!\left\langle \partial_x^2 f_0[x+W(t)]\,dt\right\rangle\!\!\right\rangle, \tag{17.445}$$

where we have differentiated according to the Itō rule. The dW term vanishes in the ensemble average,

$$\partial_t f\,dt = \frac{1}{2}\left\langle\!\!\left\langle f_0''[x+W(t)]\right\rangle\!\!\right\rangle dt, \tag{17.446}$$

and differentiating (17.444) with respect to x to evaluate the averages on the right-hand side, we see that the first derivatives cancel, and the second-derivative term gives the remaining term we need in Eq. (17.443), after canceling factors of dt.

[33] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 103.

17.11.2 Proof: Diffusion with Damping

Now consider the case where we have the damping term $V(x,t)$, but $g = 0$. Then the PDE we wish to generate is

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x,t) f, \tag{17.447}$$

and we want to show that this PDE is solved by

$$f(x,t) = \left\langle\!\!\left\langle f_0[x + W(t)] \, \exp\!\left(-\int_0^t dt' \, V[x + W(t'), t - t'] \right) \right\rangle\!\!\right\rangle. \tag{17.448}$$

The procedure here is somewhat more complicated than for the simple diffusion in Eqs. (17.443) and (17.444). To keep the calculation organized, consider the quantity[34]

$$M(t) = f[x + W(t), t'' - t] \, \exp\left(-\int_0^t dt' \, V[x + W(t'), t'' - t'] \right), \tag{17.449}$$

where we assume $f(x,t)$ to satisfy the PDE (17.447), with x and t'' effectively fixed parameters. This is something like the solution (17.448) with the initial condition replaced by the time-dependent solution. We will now compute the differential of M:

$$\begin{aligned} dM(t) = \Bigg(& - \partial_t f[x + W(t), t'' - t] \, dt + \partial_x f[x + W(t), t'' - t] \, dW \\ & + \frac{1}{2}\partial_x^2 f[x + W(t), t'' - t] \, dt \Bigg) e^{-\int_0^t dt' \, V[x + W(t'), t'' - t]} \\ & + f[x + W(t), t'' - t] \, V[x + W(t), t'' - t] \, e^{-\int_0^t dt' \, V[x + W(t'), t'' - t]} \, dt. \end{aligned} \tag{17.450}$$

The first, third, and fourth terms here vanish together since we assumed $f(x,t)$ to satisfy (17.447). Thus,

$$dM(t) = \partial_x f[x + W(t), t'' - t] \, \exp\!\left(-\int_0^t dt' \, V[x + W(t'), t'' - t'] \right) dW. \tag{17.451}$$

In particular, $\langle\!\langle dM(t) \rangle\!\rangle = 0$, so $M(t)$ is a martingale, which says that $M(t)$ tracks (at least locally) the average behavior of our solution. We will make use of this as follows. Note that evaluating $\langle\!\langle M(t) \rangle\!\rangle$ at $t = t''$, we have

$$\left\langle\!\!\left\langle M(t'') \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle f[x + W(t''), 0] \, \exp\!\left(-\int_0^{t''} dt' \, V[x + W(t'), t'' - t'] \right) \right\rangle\!\!\right\rangle, \tag{17.452}$$

which is just the desired solution (17.448) with the replacement $t \longrightarrow t''$, while evaluating $\langle\!\langle M(t) \rangle\!\rangle$ at $t = 0$ gives the general solution

$$\left\langle\!\!\left\langle M(0) \right\rangle\!\!\right\rangle = f(x, t''). \tag{17.453}$$

It follows from integrating $\langle\!\langle dM(t) \rangle\!\rangle = 0$ from $t = 0$ to t'' that we have a constant of the motion,

$$\left\langle\!\!\left\langle M(0) \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle M(t'') \right\rangle\!\!\right\rangle, \tag{17.454}$$

which directly leads to Eq. (17.448).

[34] The proof here follows the basic idea of Richard Durrett, *Stochastic Calculus: A Practical Introduction* (CRC Press, 1996), pp. 137-41.

17.11.3 Proof: Diffusion with Source

Next, we proceed to the case where we have the source term $g(x,t)$, but $V = 0$. Then the PDE we wish to generate is

$$\partial_t f = \frac{1}{2}\partial_x^2 f + g(x,t), \tag{17.455}$$

and we want to show that this PDE is solved by

$$f(x,t) = \left\langle\!\!\left\langle f_0[x + W(t)] + \int_0^t dt'\, g[x + W(t'), t - t'] \right\rangle\!\!\right\rangle. \tag{17.456}$$

Again, consider the quantity[35]

$$M(t) = f[x + W(t), t'' - t] + \int_0^t dt'\, g[x + W(t'), t'' - t'], \tag{17.457}$$

with $f(x,t)$ satisfying Eq. (17.455). The differential is

$$\begin{aligned} dM(t) &= -\partial_t f[x + W(t), t'' - t]\, dt + \partial_x f[x + W(t), t'' - t]\, dW + \frac{1}{2}\partial_x^2 f[x + W(t), t'' - t]\, dt \\ &\qquad + g[x + W(t), t'' - t] \\ &= \partial_x f[x + W(t), t'' - t]\, dW. \end{aligned} \tag{17.458}$$

Again, we have that $M(t)$ is a martingale,

$$d\Big\langle\!\!\Big\langle M(t) \Big\rangle\!\!\Big\rangle = 0, \tag{17.459}$$

which upon integration gives

$$\Big\langle\!\!\Big\langle M(0) \Big\rangle\!\!\Big\rangle = \Big\langle\!\!\Big\langle M(t'') \Big\rangle\!\!\Big\rangle, \tag{17.460}$$

where

$$\Big\langle\!\!\Big\langle M(0) \Big\rangle\!\!\Big\rangle = f(x,t''), \tag{17.461}$$

and

$$\Big\langle\!\!\Big\langle M(t'') \Big\rangle\!\!\Big\rangle = \left\langle\!\!\left\langle f[x + W(t), 0] + \int_0^{t''} dt'\, g[x + W(t'), t'' - t'] \right\rangle\!\!\right\rangle, \tag{17.462}$$

thus establishing the desired solution (17.456).

17.11.4 Proof: Diffusion with Damping and Source

Now we return to the general case, Eqs. (17.438) and (17.440). In analogy with the simpler cases, we consider the candidate quantity

$$\begin{aligned} M(t) = f[x + W(t), t'' - t] \exp\left(- \int_0^t dt'\, V[x + W(t'), t'' - t'] \right) \\ + \int_0^t dt'\, g[x + W(t'), t'' - t'] \exp\left(- \int_0^{t'} dt'''\, V[x + W(t'''), t'' - t'''] \right). \end{aligned} \tag{17.463}$$

[35] Richard Durrett, *op. cit.*, pp. 130-6.

The differential is

$$\begin{aligned} dM(t) = \Bigg(& -\partial_t f[x+W(t),t''-t]\,dt + \partial_x f[x+W(t),t''-t]\,dW \\ & + \frac{1}{2}\partial_x^2 f[x+W(t),t''-t]\,dt\Bigg) \exp\left(-\int_0^t dt'\,V[x+W(t'),t''-t']\right) \\ & - f[x+W(t),t''-t]\,V[x+W(t),t''-t]\,\exp\left(-\int_0^t dt'\,V[x+W(t'),t''-t']\right) dt \\ & + g[x+W(t),t''-t]\,\exp\left(-\int_0^t dt'''\,V[x+W(t'''),t''-t''']\right) dt \\ = {} & \partial_x f[x+W(t),t''-t]\,dW, \end{aligned} \tag{17.464}$$

after using the PDE (17.438) as usual to eliminate terms. Once again, $M(t)$ is a martingale,

$$d\Big\langle\!\Big\langle M(t)\Big\rangle\!\Big\rangle = 0, \tag{17.465}$$

so that

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t'')\Big\rangle\!\Big\rangle, \tag{17.466}$$

where

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = f(x,t''), \tag{17.467}$$

and

$$\begin{aligned} \Big\langle\!\Big\langle M(t'')\Big\rangle\!\Big\rangle = \Bigg\langle\!\Bigg\langle & f[x+W(t),0]\,\exp\left(-\int_0^{t''} dt'\,V[x+W(t'),t''-t']\right) \\ & + \int_0^{t''} dt'\,g[x+W(t'),t''-t']\,\exp\left(-\int_0^{t'} dt'''\,V[x+W(t'''),t''-t''']\right)\Bigg\rangle\!\Bigg\rangle, \end{aligned} \tag{17.468}$$

which establishes the solution (17.440).

17.11.5 Other Forms

Other forms of the Feynman–Kac theorem are common, for example, that specify *final* conditions for the diffusion equation, or that employ more complicated diffusions. As a simple example of the latter, consider

$$\partial_t f = \alpha(x)\,\partial_x f + \frac{1}{2}\beta^2(x)\,\partial_x^2 f, \tag{17.469}$$

(generalized diffusion equation)

which is solved by

$$f(x,t) = \Big\langle\!\Big\langle f_0[y(t)]\Big\rangle\!\Big\rangle_{y(0)=x}, \tag{17.470}$$

(Feynman–Kac path average)

where again $f_0(x) = f(x,0)$, and we assume state-dependent drift and diffusion,

$$dy = \alpha(y)\,dt + \beta(y)\,dW, \qquad y(0) = x, \tag{17.471}$$

(trajectories for Feynman–Kac equation)

with no explicit time dependence in the SDE. Defining

$$M(t) = f[y(t), t'-t], \tag{17.472}$$

the differential is

$$\begin{aligned} dM(t) &= -\partial_t f[y(t), t'-t]\,dt + \partial_x f[y(t), t'-t]\,dy + \frac{1}{2}\partial_x^2 f[y(t), t'-t]\,(dy)^2 \\ &= -\partial_t f[y(t), t'-t]\,dt + \alpha(y, t'-t)\,\partial_x f[y(t), t'-t]\,dt + \beta(y, t'-t)\,\partial_x f[y(t), t'-t]\,dW \\ &\qquad + \frac{1}{2}\beta^2(y, t'-t)\,\partial_x^2 f[y(t), t'-t]\,dt \\ &= \beta(y, t'-t)\,\partial_x f[y(t), t'-t]\,dW. \end{aligned} \tag{17.473}$$

As usual, $M(t)$ is a martingale, meaning $d\langle\langle M(t)\rangle\rangle = 0$, so that $\langle\langle M(0)\rangle\rangle = \langle\langle M(t')\rangle\rangle$, where

$$\left\langle\left\langle M(0)\right\rangle\right\rangle = f[y(0), t'] = f(x, t'), \tag{17.474}$$

and

$$\left\langle\left\langle M(t')\right\rangle\right\rangle = \left\langle\left\langle f[y(t'), 0]\right\rangle\right\rangle, \tag{17.475}$$

establishing (17.470).

17.11.5.1 General form for State-Dependent Diffusion

For the same state-dependent drift and diffusion represented by the SDE (17.470), but adding decay and source terms to the SDE as in

$$\partial_t f = \alpha(x)\,\partial_x f + \frac{1}{2}\beta^2(x)\,\partial_x^2 f - V(x,t)f + g(x,t),$$

(PDE for generalized Feynman–Kac formula) (17.476)

the corresponding Feynman–Kac formula becomes

$$\begin{aligned} f(x,t) = \left\langle\left\langle f_0[y(t)]\,\exp\left(-\int_0^t dt'\,V[y(t'), t-t']\right)\right.\right. \\ \left.\left. + \int_0^t dt'\,g[y(t'), t-t']\,\exp\left(-\int_0^{t'} dt''\,V[y(t''), t-t'']\right)\right\rangle\right\rangle_{y(0)=x}. \end{aligned}$$

(generalized Feynman–Kac formula) (17.477)

The proof for this goes in the same way by considering the martingale function

$$\begin{aligned} M(t) = f[y(t), t''-t]\,\exp\left(-\int_0^t dt'\,V[y(t'), t''-t']\right) \\ + \int_0^t dt'\,g[y(t'), t''-t']\,\exp\left(-\int_0^{t'} dt'''\,V[y(t'''), t''-t''']\right), \end{aligned} \tag{17.478}$$

where the differential is

$$\begin{aligned}
dM(t) &= \left(-\partial_t f[y(t),t''-t]\,dt + \partial_x f[y(t),t''-t]\,dy + \frac{1}{2}\partial_x^2 f[y(t),t''-t]\,dy^2\right) \\
&\qquad \times \exp\left(-\int_0^t dt'\,V[y(t'),t''-t']\right) \\
&\qquad - f[y(t),t''-t]\,V[y(t),t''-t]\,\exp\left(-\int_0^t dt'\,V[y(t'),t''-t']\right)dt \\
&\qquad + g[y(t),t''-t]\,\exp\left(-\int_0^t dt'''\,V[y(t'''),t''-t''']\right)dt \\
&= \left(-\partial_t f[y(t),t''-t]\,dt + \Big[\alpha[y(t)]\,dt + \beta[y(t)]\,dW\Big]\partial_x f[y(t),t''-t]\right. \\
&\qquad \left. + \frac{1}{2}\beta^2(y,t)\,\partial_x^2 f[y(t),t''-t]\,dt\right)\exp\left(-\int_0^t dt'\,V[y(t'),t''-t']\right) \\
&\qquad - f[y(t),t''-t]\,V[y(t),t''-t]\,\exp\left(-\int_0^t dt'\,V[y(t'),t''-t']\right)dt \\
&\qquad + g[y(t),t''-t]\,\exp\left(-\int_0^t dt'''\,V[y(t'''),t''-t''']\right)dt \\
&= \beta[y(t)]\,\partial_x f[y(t),t''-t]\,dW.
\end{aligned} \tag{17.479}$$

We have already implemented the diffusion equation (17.476) as usual to show that this is a martingale, $\langle\langle dM(t)\rangle\rangle = 0$. It is important to notice here that this argument does not carry through if α and β depend explicitly on time, because then the time dependences do not match in the correct way to permit the diffusion equation to cancel all the deterministic terms of $dM(t)$.

Then setting

$$\Big\langle\Big\langle M(0)\Big\rangle\Big\rangle = \Big\langle\Big\langle M(t'')\Big\rangle\Big\rangle, \tag{17.480}$$

where

$$\Big\langle\Big\langle M(0)\Big\rangle\Big\rangle = f[y(0),t''], \tag{17.481}$$

and

$$\begin{aligned}
\Big\langle\Big\langle M(t'')\Big\rangle\Big\rangle = \Big\langle\Big\langle & f[y(t''),0]\,\exp\left(-\int_0^{t''} dt'\,V[y(t'),t''-t']\right) \\
& + \int_0^{t''} dt'\,g[y(t'),t''-t']\,\exp\left(-\int_0^{t'} dt'''\,V[y(t'''),t''-t''']\right)\Big\rangle\Big\rangle,
\end{aligned} \tag{17.482}$$

and dropping the primes on t'', we arrive at the result (17.478).

17.11.5.2 Time-Dependent Drift and Diffusion: First Form

Again, for

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW,$$

(trajectories for Feynman–Kac equation) (17.483)

the approach from the previous section does not carry through because the explicit time dependence of α and β runs forward in time, but the time dependence of f runs backwards in time. However, both should be the same, as they appear in the PDE. Thus consider the alternate martingale function

$$\begin{aligned}
M(t) = f[y(t''-t),t''-t]\,&\exp\left(-\int_0^t dt'\,V[y(t''-t'),t''-t']\right) \\
&+ \int_0^t dt'\,g[y(t''-t'),t''-t']\,\exp\left(-\int_0^{t'} dt'''\,V[y(t''-t'''),t''-t''']\right),
\end{aligned} \tag{17.484}$$

with differential

$$\begin{aligned}
dM(t) &= \Bigg(-\partial_t f[y(t''-t),t''-t]\,dt + \partial_x f[y(t''-t),t''-t]\,dy + \frac{1}{2}\partial_x^2 f[y(t''-t),t''-t]\,dy^2\Bigg) \\
&\qquad \times \exp\Bigg(-\int_0^t dt'\,V[y(t''-t'),t''-t']\Bigg) \\
&\qquad - f[y(t''-t),t''-t]\,V[y(t''-t),t''-t]\,\exp\Bigg(-\int_0^t dt'\,V[y(t''-t'),t''-t']\Bigg)\,dt \\
&\qquad + g[y(t''-t),t''-t]\,\exp\Bigg(-\int_0^t dt'''\,V[y(t''-t'''),t''-t''']\Bigg)\,dt \\
&= \Bigg(-\partial_t f[y(t''-t),t''-t]\,dt - \Big[\alpha(y,t''-t)\,dt + \beta[y(t),t''-t]\,dW(t''-t)\Big]\partial_x f[y(t),t''-t] \\
&\qquad + \frac{1}{2}\beta^2(y,t''-t)\,\partial_x^2 f[y(t),t''-t]\,dt\Bigg)\exp\Bigg(-\int_0^t dt'\,V[y(t''-t'),t''-t']\Bigg) \\
&\qquad - f[y(t''-t),t''-t]\,V[y(t''-t),t''-t]\,\exp\Bigg(-\int_0^t dt'\,V[y(t''-t'),t''-t']\Bigg)\,dt \\
&\qquad + g[y(t''-t),t''-t]\,\exp\Bigg(-\int_0^t dt'''\,V[y(t''-t'''),t''-t''']\Bigg)\,dt \\
&= -\beta(y,t''-t)\,\partial_x f[y(t''-t),t''-t]\,dW(t''-t).
\end{aligned} \tag{17.485}$$

Note that all the evolution now is explicitly backwards, including the time dependence of the trajectories $y(t''-t)$. In the last step, we used the PDE

$$\partial_t f = -\alpha(x,t)\,\partial_x f + \frac{1}{2}\beta^2(x,t)\,\partial_x^2 f - V(x,t)f + g(x,t),$$

(PDE for generalized Feynman–Kac formula) (17.486)

which gives the relevant diffusion equation. As usual, setting

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle M(t'')\Big\rangle\!\Big\rangle, \tag{17.487}$$

where

$$\Big\langle\!\Big\langle M(0)\Big\rangle\!\Big\rangle = f[y(t''),t''], \tag{17.488}$$

and

$$\begin{aligned}
\Big\langle\!\Big\langle M(t'')\Big\rangle\!\Big\rangle = \Bigg\langle\!\Bigg\langle & f[y(0),0]\,\exp\Bigg(-\int_0^{t''} dt'\,V[y(t''-t'),t''-t']\Bigg) \\
& + \int_0^{t''} dt'\,g[y(t''-t'),t''-t']\,\exp\Bigg(-\int_0^{t'} dt'''\,V[y(t''-t'''),t''-t''']\Bigg)\Bigg\rangle\!\Bigg\rangle,
\end{aligned} \tag{17.489}$$

gives the generalized Feynman–Kac formula

$$\begin{aligned}
f(x,t) = \Bigg\langle\!\Bigg\langle & f_0[y(0)]\,\exp\Bigg(-\int_0^t dt'\,V[y(t-t'),t-t']\Bigg) \\
& + \int_0^t dt'\,g[y(t-t'),t-t']\,\exp\Bigg(-\int_0^{t'} dt''\,V[y(t-t''),t-t'']\Bigg)\Bigg\rangle\!\Bigg\rangle_{y(t)=x}.
\end{aligned}$$

(generalized Feynman–Kac formula) (17.490)

The time dependence can be simplified somewhat by writing

$$f(x,t) = \left\langle\!\!\left\langle f_0[y(0)] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) + \int_0^t dt'\, g[y(t'),t'] \exp\!\left(-\int_{t-t'}^t dt''\, V[y(t''),t'']\right) \right\rangle\!\!\right\rangle_{y(t)=x}.$$

(generalized Feynman–Kac formula) (17.491)

However, the form (17.490) is somewhat better for interpreting the formula, because it explicitly traces each trajectory $y(t)$, ''beginning'' at $y(t) = x$, *backwards* in time to the initial condition $y(0)$. The formula then weights the average in the first term by the initial distribution f_0 at the corresponding state $y(0)$. The second term corresponds to the creating of a trajectory by the source $g(x,t)$ at time t' (summed over all possible t'), propagating backwards to $t = 0$. In both terms, the trajectory amplitudes were damped by $V(x,t)$, integrated over the temporal extent of each path. The backwards-propagating nature of the solutions is reflected in the above PDE, where the drift coefficient has a minus sign compared to the earlier, ''forward-propagating'' PDE (17.476). (The sign change does not apply to the diffusion term, since the backward propagation is conditioned on a *final* condition, not an *initial* condition.) Note also that while more general than Eq. (17.477) in accounting for explicit time dependence in the SDE coefficients, Eq. (17.477) is much more convenient, from the point of view of simulation—the trajectories here have to be propagated backwards from a particular final point (in which case the SDE acts as an anticipating SDE), and weighted according to the corresponding zero-time point.

17.11.5.3 Time-Dependent Drift and Diffusion: Second Form

An alternative, ''forward-time'' version of the generalized Feynman–Kac formula arises by considering the forward-propagating martingale function

$$M(t) = f[y(t),t] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) + \int_0^t dt'\, g[y(t'),t'] \exp\!\left(-\int_0^{t'} dt''\, V[y(t''),t'']\right). \tag{17.492}$$

The differential is

$$\begin{aligned} dM(t) &= \left(\partial_t f[y(t),t]\, dt + \partial_x f[y(t),t]\, dy + \frac{1}{2}\partial_x^2 f[y(t),t]\, dy^2\right) \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) \\ &\quad - f[y(t),t]\, V[y(t),t] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) dt \\ &\quad + g[y(t),t] \exp\left(-\int_0^t dt''\, V[y(t''),t'']\right) dt \\ &= \left(\partial_t f[y(t),t]\, dt + \Big[\alpha(y,t)\, dt + \beta(y,t)\, dW\Big] \partial_x f[y(t),t] \right. \\ &\quad \left. + \frac{1}{2}\beta^2(y,t)\, \partial_x^2 f[y(t),t]\, dt\right) \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) \\ &\quad - f[y(t),t]\, V[y(t),t] \exp\!\left(-\int_0^t dt'\, V[y(t'),t']\right) dt \\ &\quad + g[y(t),t] \exp\!\left(-\int_0^t dt''\, V[y(t''),t'']\right) dt \\ &= \beta(y,t)\, \partial_x f[y(t),t]\, dW(t), \end{aligned} \tag{17.493}$$

where in the last step we used the PDE

$$-\partial_t f = \alpha(x,t)\, \partial_x f + \frac{1}{2}\beta^2(x,t)\, \partial_x^2 f - V(x,t) f + g(x,t).$$

(PDE for generalized Feynman–Kac formula) (17.494)

Then as we have done so many times, we set

$$\left\langle\!\left\langle M(0) \right\rangle\!\right\rangle = \left\langle\!\left\langle M(t) \right\rangle\!\right\rangle, \tag{17.495}$$

where

$$\left\langle\!\left\langle M(0) \right\rangle\!\right\rangle = f[y(0), 0], \tag{17.496}$$

and

$$\begin{aligned} \left\langle\!\left\langle M(t) \right\rangle\!\right\rangle = \Bigg\langle\!\!\Bigg\langle & f[y(t), t] \exp\left(-\int_0^t dt' \, V[y(t'), t'] \right) \\ & + \int_0^t dt' \, g[y(t'), t'] \exp\left(-\int_0^{t'} dt'' \, V[y(t''), t''] \right) \Bigg\rangle\!\!\Bigg\rangle, \end{aligned} \tag{17.497}$$

leading to the alternate generalized formula

$$\begin{aligned} f_0(x) = \Bigg\langle\!\!\Bigg\langle & f[y(t), t] \exp\left(-\int_0^t dt' \, V[y(t'), t'] \right) \\ & + \int_0^t dt' \, g[y(t'), t'] \exp\left(-\int_0^{t'} dt'' \, V[y(t''), t''] \right) \Bigg\rangle\!\!\Bigg\rangle_{y(0)=x}. \end{aligned}$$

(generalized Feynman–Kac formula) (17.498)

Like Eq. (17.491), this form is in a sense less convenient for simulation than Eq. (17.477), because it obtains the initial distribution from the final distribution. The reverse-time nature of this formula is reflected in the negative time derivative in the PDE (17.494).

Note also that the diffusion equation (17.494) has the form of the Kolmogorov backward equation (17.144). Thus, the derivation of the above is effectively an alternate derivation of the Kolmogorov backward equation, for the evolution of $P(x, t|x_0, t_0)$ with t_0, given a final distribution at t. This is implicit in the form of the solution (17.498), which gives the initial distribution at t_0, given the final distribution at t, such that the time derivative in the PDE (17.494) can be regarded here as a derivative with respect to t_0, and the distribution $f(x, t)$ should be regarded as equivalent to $P(x, t|x_0, t_0)$.

17.12 Sojourn Times

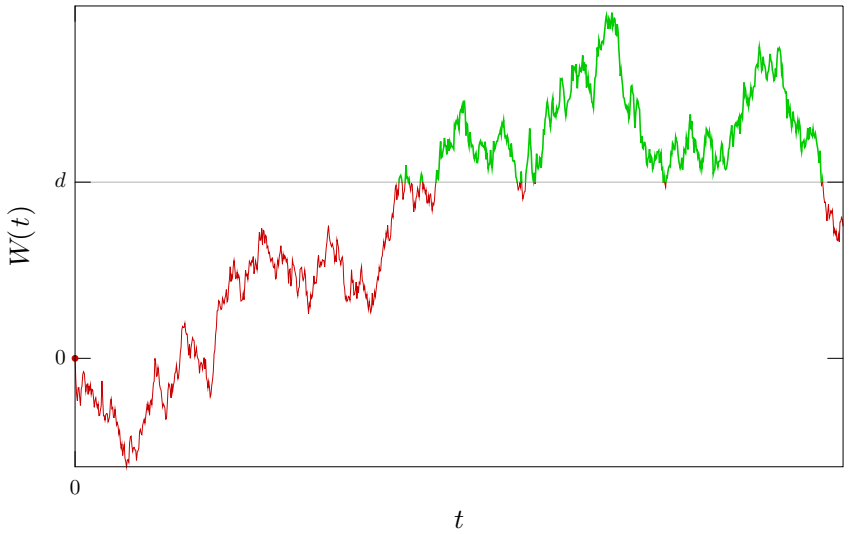
17.12.1 Wiener Process

As an example application of the Feynman–Kac formula, we wish to consider the **sojourn time** of a Wiener path $W(t)$ past a boundary at d, which is defined as the functional

$$T_\mathrm{s}[W(t); d] := \int_0^t dt' \, \Theta[W(t') - d] \qquad (0 \le T_\mathrm{s}[W(t); d] \le t), \tag{17.499}$$

(sojourn time)

or the total time that $W(t)$ spends across the boundary. In the plot below, this counts the portion of time that the path is highlighted in green.



We will only consider the case $d \geq 0$, since we can get the $d < 0$ probabilities as complements to the probabilities we calculate.

Let $f_{T_s}(x)$ for $0 \leq x \leq t$ denote the probability density for the sojourn time $T_s[W(t); d]$ of $W(t)$ above d, with cumulative probability $P(T_s \leq x)$, satisfying

$$f_{T_s}(x) = \partial_x P(T_s \leq x). \tag{17.500}$$

Then the Laplace transform of $f_{T_s}(x)$ is

$$\int_0^t dx\, e^{-sx}\, f_{T_s}(x) = \left\langle\!\!\left\langle \exp\{-sT_s[W(t); d]\} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \exp\left[-s\int_0^t dt'\, \Theta[W(t') - d]\right] \right\rangle\!\!\right\rangle. \tag{17.501}$$

Our goal will be to compute the expectation value on the right via the Feynman–Kac formula, and then to obtain $f_{T_s}(x)$ by inverting the Laplace transform.[36]

Consider now the driven diffusion equation

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x)f - \lambda f + g(x), \tag{17.502}$$

where $V(x)$ is the occupation function we seek here (i.e., the step function)—we will take it to be the constant s past the barrier, and 0 before it:

$$V(x) = s\Theta(x - d). \tag{17.503}$$

The Feynman–Kac formula (17.440) gives the solution of this equation as

$$\begin{aligned} f(x,t) = \Bigg\langle\!\!\Bigg\langle & f_0[x + W(t)] \exp\left(-\lambda t - \int_0^t dt'\, V[x + W(t')]\right) \\ & + \int_0^t dt'\, g[x + W(t')] \exp\left(-\lambda t' - \int_0^{t'} dt''\, V[x + W(t'')]\right) \Bigg\rangle\!\!\Bigg\rangle. \end{aligned} \tag{17.504}$$

We will seek the steady-state solution $f(x) := f(x, t \longrightarrow \infty)$. In particular, $f(x)$ must be independent of $f_0(x)$, so we will take this initial condition to be zero. Thus, we have the steady state

$$\begin{aligned} f(x) &= \left\langle\!\!\left\langle \int_0^\infty dt\, \exp\left(-\lambda t - \int_0^t dt'\, V[x + W(t')]\right) \right\rangle\!\!\right\rangle \\ &= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\, \Theta[x + W(t') - d]\right) \right\rangle\!\!\right\rangle, \end{aligned} \tag{17.505}$$

[36] here we are adapting the method of Gerard Hooghiemstra, "On explicit occupation time distributions for Brownian processes," *Statistics & Probability Letters* **56**, 405 (2002) (doi: 10.1016/S0167-7152(02)00037-8).

after dropping some primes and setting $g(x) = 1$. This contains our desired expectation value from Eq. (17.501) upon setting $x = 0$, except that we still here have *another* Laplace transform of the desired expectation value. This is then the solution of the steady-state version of Eq. (17.502):

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\Theta(x-d)f(x) + 1. \tag{17.506}$$

For $x > d$, the ODE is

$$f'' = 2(\lambda + s)f - 2, \tag{17.507}$$

with the case of $x < d$ given by setting $s = 0$. Then setting $h = f - 1/(\lambda + s)$, we have $h'' = 2(\lambda + s)h$, so that for $x > d$,

$$h(x) = Ae^{-\sqrt{2(\lambda+s)}\,x}, \tag{17.508}$$

or

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{1}{\lambda+s} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x < d) \end{cases} \tag{17.509}$$

for some undetermined constants A and B, where we have chosen the bounded solution on either side of d. Demanding continuity of $f(x)$ at $x = d$ gives

$$Ae^{-\sqrt{2(\lambda+s)}\,d} + \frac{1}{\lambda+s} = Be^{\sqrt{2\lambda}\,d} + \frac{1}{\lambda}, \tag{17.510}$$

and continuity of $f'(x)$ at $x = d$ gives

$$-\sqrt{2(\lambda+s)}Ae^{-\sqrt{2(\lambda+s)}\,d} = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d}. \tag{17.511}$$

The solution of these two equations fixes the coefficients as

$$A = \frac{se^{\sqrt{2(\lambda+s)}\,d}}{(\lambda+s)[\lambda + \sqrt{\lambda(\lambda+s)}]}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]}. \tag{17.512}$$

Now, we can equate Eqs. (17.505) and (17.509) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\Theta[W(t') - d]\right)\right\rangle\!\!\right\rangle = B + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]}. \tag{17.513}$$

Then using Eq. (17.501) on the left-hand side,

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\{-sT_{\rm s}[W(t)]\}\right\rangle\!\!\right\rangle = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]}. \tag{17.514}$$

Now we can use the formula for the inverse Laplace transform [Problem 15.4]

$$y(t) = \frac{1}{2\pi i}\int_{0^+-i\infty}^{0^++i\infty} ds\, e^{s\tau}\,\mathscr{L}[y](s), \tag{17.515}$$

where $\mathscr{L}[y](s)$ is the Laplace transform of $y(t)$:

$$\mathscr{L}[y](s) := \int_0^\infty dt\, e^{-st}\, y(t). \tag{17.516}$$

Then we have

$$\left\langle\!\!\left\langle \exp\{-sT_{\rm s}[W(t)]\}\right\rangle\!\!\right\rangle = \frac{1}{2\pi i}\int_{0^+-i\infty}^{0^++i\infty} d\lambda\, e^{\lambda t}\left[\frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda[\lambda + s + \sqrt{\lambda(\lambda+s)}]}\right]. \tag{17.517}$$

Evaluating the first term is simple by rotating the integration direction and then completing the contour integral around the great half-plane:

$$\frac{1}{2\pi i}\int_{0^+-i\infty}^{0^++i\infty} d\lambda\,\frac{e^{\lambda t}}{\lambda} = \frac{1}{2\pi i}\int_{-\infty-i0^+}^{\infty-i0^+} d\lambda\,\frac{e^{i\lambda t}}{\lambda} = 1, \qquad (t>0). \tag{17.518}$$

The second term is more involved, and to evaluate it we will stick to ''recipes'' for known Laplace transforms. We will need[37]

$$\mathscr{L}\left[\frac{e^{-st/2}}{\sqrt{s}} I_{1/2}(st/2)\right](\lambda) = \frac{1}{\sqrt{\lambda}\left(\lambda + s + \sqrt{\lambda(\lambda+s)}\right)}, \tag{17.519}$$

where $I_\nu(x)$ is a modified Bessel function, and[38]

$$\mathscr{L}\left[\frac{1}{\sqrt{\pi t}} e^{-k^2/4t}\right](\lambda) = \frac{1}{\sqrt{\lambda}} e^{-k\sqrt{\lambda}}, \tag{17.520}$$

along with the convolution theorem for Laplace transforms,[39]

$$\mathscr{L}\left[\int_0^t d\tau\, f(t-\tau) g(\tau)\right] = \mathscr{L}[f]\,\mathscr{L}[g], \tag{17.521}$$

which combine to give the Laplace transform appropriate to the second term in Eq. (17.514),

$$\mathscr{L}\left[\int_0^t d\tau\,\frac{e^{-k^2/4(t-\tau)}}{\sqrt{\pi(t-\tau)}}\frac{e^{-s\tau/2}}{\sqrt{s}} I_{1/2}(s\tau/2)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\lambda\left(\lambda + s + \sqrt{\lambda(\lambda+s)}\right)}, \tag{17.522}$$

provided we take $k = -\sqrt{2}\,d$ and insert an overall factor of $-s$. Thus, Eq. (17.514) becomes

$$\int_0^t dx\, e^{-sx}\, f_{T_s}(x) = 1 - \sqrt{\frac{s}{\pi}}\int_0^t d\tau\,\frac{e^{-d^2/2(t-\tau)-s\tau/2}}{\sqrt{t-\tau}} I_{1/2}(s\tau/2) \tag{17.523}$$

after using Eq. (17.501) to replace the ensemble average on the left-hand side. Now we will use an integral representation of the Bessel function,[40]

$$I_{1/2}(x) = \sqrt{\frac{x}{2\pi}}\int_{-1}^1 du\, e^{-ux} = \sqrt{\frac{2x}{\pi}} e^x \int_0^1 dv\, e^{-2vx} = \sqrt{\frac{2x}{\pi}}\frac{e^x}{t}\int_0^t dv\, e^{-2vx/t}, \tag{17.524}$$

where we have changed variables by setting $1+u = 2v$ and then letting $v \longrightarrow v/t$. Then Eq. (17.523) becomes

$$\begin{aligned}
\int_0^t dx\, e^{-sx}\, f_{T_s}(x) &= 1 - \frac{s}{\pi t}\int_0^t d\tau \int_0^t dv\, e^{-vs\tau/t} e^{-d^2/2(t-\tau)}\sqrt{\frac{\tau}{t-\tau}} \\
&= 1 + \frac{1}{\pi}\int_0^t d\tau\left(e^{-s\tau}-1\right)\frac{e^{-d^2/2(t-\tau)}}{\sqrt{\tau(t-\tau)}} \\
&= 1 - I(d,t) + \frac{1}{\pi}\int_0^t d\tau\, e^{-s\tau}\frac{e^{-d^2/2(t-\tau)}}{\sqrt{\tau(t-\tau)}} \\
&= \int_0^t dx\, e^{-sx}\,\delta(x-0^+)\,[1-I(d,t)] + \frac{1}{\pi}\int_0^t dx\, e^{-sx}\frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}},
\end{aligned} \tag{17.525}$$

[37] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 1024, Eq. (29.3.54).
[38] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.84).
[39] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1020, Eq. (29.2.8).
[40] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 376, Eq. (9.6.18).

where in the last step we set $\tau = x$, and we have also named the integral

$$I(d,t) := \frac{1}{\pi}\int_0^t dx\,\frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}}, \tag{17.526}$$

which we will evaluate shortly. Eq. (17.525) now equates (finite-time) Laplace transforms. Differentiating with respect to time undoes them, which then yields

$$f_{T_\mathrm{s}}(x) = [1 - I(d,t)]\,\delta(x - 0^+) + \frac{1}{\pi}\frac{e^{-d^2/2(t-x)}}{\sqrt{x(t-x)}}. \tag{17.527}$$

Clearly, this result is normalized, since integration over x produces canceling terms of $I(d,\tau)$ from the δ function and the integral. The δ function indicates that there is a (possibly) finite probability for having zero sojourn time, whereas positive sojourn times have infinitesimal probability, as we expect for a distribution. In particular, the probability for having zero sojourn time is just

$$\begin{aligned} P(T_\mathrm{s} = 0) &= \lim_{\epsilon\to 0}\int_0^\epsilon dx\, f_{T_\mathrm{s}}(x) \\ &= 1 - I(d,t). \end{aligned} \tag{17.528}$$

However, we have already calculated this: the probability to *not* sojourn across the boundary at d is equivalent to the non-crossing probability for a boundary at d. This is the complement of the crossing probability from Eq. (17.372), and thus

$$P(T_\mathrm{s} = 0) = 1 - P_\mathrm{cross}(d,t) = 1 - \mathrm{erfc}\left(\frac{d}{\sqrt{2t}}\right) = \mathrm{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{17.529}$$

Thus, in computing the boundary-crossing probability before, we have essentially used a path-integral method to evaluate the integral (17.526), with the result

$$I(d,t) = \mathrm{erfc}\left(\frac{d}{\sqrt{2t}}\right). \tag{17.530}$$

Putting this result into Eq. (17.525), we finally have the probability density for the sojourn time of $W(t)$ beyond d:[41]

$$f_{T_\mathrm{s}}(x) = \mathrm{erf}\left(\frac{d}{\sqrt{2t}}\right)\delta(x - 0^+) + \frac{e^{-d^2/2(t-x)}}{\pi\sqrt{x(t-x)}} \qquad (0 \le x \le t, d \ge 0). \tag{17.531}$$

(probability density for sojourn time of $W(t)$ past d)

The probability density for the time the particle *stays under* the barrier at d is given by the replacement $x \longrightarrow t - x$. The cumulative probability is given by simply integrating this from 0 to x:

$$P(T_\mathrm{s} \le x) = \mathrm{erf}\left(\frac{d}{\sqrt{2t}}\right) + \frac{1}{\pi}\int_0^x dx'\,\frac{e^{-d^2/2(t-x')}}{\sqrt{x'(t-x')}} \qquad (0 \le x \le t, d \ge 0). \tag{17.532}$$

(cumulative probability for sojourn time of $W(t)$ past d)

For the case $d \le 0$, we simply have $P(T_\mathrm{s} \le x, -|d|) = 1 - P(T_\mathrm{s} \le x, |d|)$, since there are only two, nonintersecting regions in which to sojourn.

[41] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 156, formula 1.4.4, though note the absence of the first (δ-function) term; Paul Lévy, "Sur certains processus stochastiques homogènes," *Compositio Mathematica* **7**, 283 (1940), p. 327, Eq. (58) (`http://www.numdam.org/item?id=CM_1940__7__283_0`); Marc Yor, "The Distribution of Brownian Quantiles," *Journal of Applied Probability* **32**, 405 (1995), Eq. (3.a) (doi: 10.2307/3215296); Lajos Takács, "Sojourn Times," *Journal of Applied Mathematics and Stochastic Analysis* **9**, 415 (1996), Eq. (3) (doi: 10.1155/S1048953396000366).

The moments are given simply by integrating the density (17.531) as

$$\begin{aligned}\left\langle\!\left\langle {T_{\rm s}}^n \right\rangle\!\right\rangle &= \int_0^t dx\, x^n f_{T_{\rm s}}(x) \\ &= \int_0^t dx\, x^n \frac{e^{-d^2/2(t-x)}}{\pi\sqrt{x(t-x)}} \qquad (d \geq 0).\end{aligned} \tag{17.533}$$

A somewhat nicer integral expression comes from changing variables via $x = t(1-\sigma^2)$, such that the moments becomes

$$\left\langle\!\left\langle {T_{\rm s}}^n \right\rangle\!\right\rangle = \frac{2t^n}{\pi} \int_0^1 d\sigma \left(1-\sigma^2\right)^{n-1/2} e^{-d^2/2t\sigma^2} \qquad (d \geq 0). \tag{17.534}$$

Thus, for example, the mean sojourn time is

$$\left\langle\!\left\langle T_{\rm s} \right\rangle\!\right\rangle = \frac{1}{2}(d^2+t)\,\mathrm{erfc}\left[\frac{d}{\sqrt{2t}}\right] - \sqrt{\frac{td^2}{2\pi}}\, e^{-d^2/2t} \qquad (d \geq 0). \tag{17.535}$$

Again, for $d < 0$, this expression for $d \geq 0$ can be adapted by subtracting it from t.

17.12.1.1 Arcsine Laws

One well-known special case of the probability distribution (17.532) is the case of $d = 0$

$$P(T_{\rm s} \leq x) = \frac{1}{\pi} \int_0^x dx' \, \frac{1}{\sqrt{x'(t-x')}}. \tag{17.536}$$

Evaluating this integral gives

$$P(T_{\rm s} \leq x) = \frac{2}{\pi} \sin^{-1}\sqrt{\frac{x}{t}} \qquad (0 \leq x \leq t, d = 0), \tag{17.537}$$

(Levy's arcsine law)

which is known as **Lévy's arcsine law**.[42] This gives the probability distribution of the time a Wiener path spends on a particular side of its starting point. The density following from Eq. (17.536) is

$$f(x) = \frac{1}{\pi\sqrt{x(t-x)}}. \tag{17.538}$$

(density for Levy arcsine law)

Since this density diverges at $x = 0$ and $x = t$, it says that the path is most likely to spend essentially all its time on one side or the other of the origin.

17.12.1.2 Alternate Arcsine Law: Last Crossing time

The same arcsine law also applies to other statistics related to Wiener processes. To explore the first statistic, suppose we fix a time t, and ask what was the most recent time before t when $W(t)$ crossed the origin?[43] That is, consider the statistic

$$\tau_{\rm L0}(t) := \sup\{t' : t' \in [0,t] \wedge W(t') = 0\}. \tag{17.539}$$

Then let's compute the cumulative probability $P(\tau_{\rm L0}(t) \leq \tau)$. For $\tau_{\rm L0}(t) \leq \tau$ to occur, the Wiener path must not touch $W = 0$ for all $t' > \tau$ (up to t). We could have $W(\tau) = 0$ or not. Then let's integrate over all possible values x of $W(\tau)$, using the Gaussian probability density for $W(\tau)$:

$$P\big[\tau_{\rm L0}(t) \leq \tau\big] = \int_{-\infty}^{\infty} dx\, f_{W(\tau)}(x)\, P[W(\tau < t' \leq t) \neq 0 | W(\tau) = x], \tag{17.540}$$

[42] Paul Lévy, "Sur un problème de M. Marcinkiewicz," *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **208**, 318 (1939) (`http://gallica.bnf.fr/ark:/12148/bpt6k3160g/f300.image`); Paul Lévy, "Sur certains processus stochastiques homogènes," *op. cit.*

[43] The proof here is from Jim Pitman's course notes, `http://www.stat.berkeley.edu/~pitman/s205s03/lecture18.pdf`, Theorem 18.4.

where of course the condition "$W(\tau) = x$" means $W(\tau)$ is between x and $x + dx$. Since the Gaussian probability density is even, we can change to evaluating only half of the integral $x \geq 0$:

$$P\big[\tau_{\text{L0}}(t) \leq \tau\big] = 2\int_0^\infty dx\, f_{W(\tau)}(x)\, P\big[W(t') < x\ \forall_{t' \in [0,t-\tau)}\big]. \tag{17.541}$$

Note we have also changed the second probability statement: it was that a Wiener process, starting at x, should not touch the origin for a time interval of length $t - \tau$. This is the same as the probability that a Wiener process, starting at 0, should not touch x over an interval of the same duration. The latter probability is just the complement of a crossing probability, which we can obtain from the complement of Eq. (17.372). Putting in this probability as an error function, and also putting in the Gaussian density, we have

$$P\big[\tau_{\text{L0}}(t) \leq \tau\big] = 2\int_0^\infty dx\, \frac{1}{\sqrt{2\pi\tau}}\, e^{-x^2/2\tau}\, \text{erf}\left(\frac{x}{\sqrt{2(t-\tau)}}\right). \tag{17.542}$$

Evaluating this integral, we have

$$P\big[\tau_{\text{L0}}(t) \leq \tau\big] = \frac{2}{\pi}\sin^{-1}\sqrt{\frac{\tau}{t}}.$$

(second arcsine law for last crossing time) (17.543)

Note that this implies the density (17.538) for $\tau_{\text{L0}}(t)$, which is someone odd in that it is a symmetric function over $[0, t]$. This means that the most recent crossing most likely occurred close to $t' = t$ or close to $t' = 0$, and the symmetry is peculiar, given that $W(t')$ itself does not share the symmetry. Note that while the last crossing is well-defined, asking something like when was the *next-to-last* crossing occur is much more complicated: the crossings of $W(t')$ form a set of measure zero, but are nevertheless uncountable, much like the Cantor set. Thus, the last crossing is not isolated, but rather has other, arbitrarily close crossings.

17.12.1.3 Alternate Arcsine Law: Maximum-Value Time

The remaining statistic governed by the arcsine law is the time $\tau_{\max}$ at which a Wiener path attains its maximum value over the time interval $[0, t]$. To see this, it is first necessary to make some comments about the maxima of a Wiener process.[44] First, note that over any time interval $[a, b]$, the maximum of $W(t)$ over the interval does not occur at either endpoint (with unit probability). This follows at the lower endpoint a from noting the equivalence (via some shifting) of $W(t)$ over this interval to a Wiener path starting at $t = 0$, and the crossing probability (17.372): over an interval of any finite duration, the probability of crossing a barrier just above the starting point is arbitrarily close to 1. The maximum does not occur at the upper endpoint b for the same reason (i.e., the same argument under a time-reversal). By subdividing the $[a, b]$ into smaller and smaller intervals, each of which has its own local maximum, we can see that the local maxima are dense (but countable). Also by subdividing $[a, b]$ into two intervals (which do not overlap except at the boundary point), each subinterval will have a local maximum. But because the Wiener process progresses independently in each interval, the local maximum can take on any value, and in particular the probability for attaining the *same* local maximum in each interval is zero. The point is that the maximum over $[a, b]$ is unique, so that $\tau_{\max}$ is well-defined.

Now consider a Wiener process $W(t)$, with the **maximum-value process**

$$M(t) := \max_{0 \leq t' \leq t} W(t'), \tag{17.544}$$

representing the running maximum value attained by $W(t)$ up to time t. Defining the difference process

$$R(t) := M(t) - W(t), \tag{17.545}$$

we will aim to show that $R(t)$ is statistically equivalent[45] to the **absolute-value process** or **reflection Brownian motion** $|W(t)|$ (about which we will have more to explore in Section 17.13.4.1). To do this,

[44] Peter Mörters and Yuval Peres, *Brownian Motion* (Cambridge, 2010) (ISBN: 0521760186).

[45] This argument follows that of Peter Mörters and Yuval Peres, *op. cit.*, at least until near the end.

consider a fixed time $t_0 > 0$, and an evolution over a time interval Δt. The Wiener process changes in this interval by

$$\Delta W_0(\Delta t) := W(t_0 + \Delta t) - W(t_0), \tag{17.546}$$

and the maximum changes by

$$\Delta M_0(\Delta t) := \max_{0 \le \delta t \le \Delta t} \Delta W_0(\delta t). \tag{17.547}$$

The idea will be to show that $R(t_0 + \Delta t)$ is statistically equivalent to $|R(t_0) + \Delta W_0(\Delta t)|$. First, note that the maximum at $t_0 + \Delta t$ is

$$M(t_0 + \Delta t) = \max\Big\{M(t_0), W(t_0) + \Delta M_0(\Delta t)\Big\}. \tag{17.548}$$

That is, the final maximum is either the initial maximum, or the initial value of $W(t)$ plus the maximum added over the interval Δt, whichever is larger. Now writing

$$\begin{aligned} R(t_0 + \Delta t) &= M(t_0 + \Delta t) - W(t_0 + \Delta t) \\ &= \max\Big\{M(t_0), W(t_0) + \Delta M_0(\Delta t)\Big\} - \Big[W(t_0) + \Delta W_0(\Delta t)\Big] \\ &= \max\Big\{R(t_0), \Delta M_0(\Delta t)\Big\} - \Delta W_0(\Delta t). \end{aligned} \tag{17.549}$$

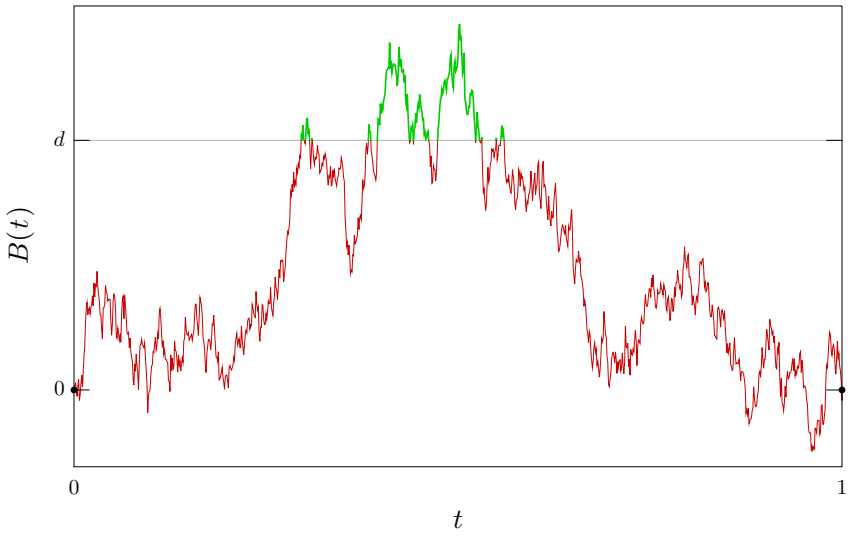
Now to analyze the two cases here. If $R(t_0)$ is the larger value in the maximum, then this reduces to $R(t_0 + \Delta t) = R(t_0) - \Delta W_0(\Delta t)$. Since $\Delta W_0(\Delta t)$ and $-\Delta W_0(\Delta t)$ are statistically equivalent, $R(t_0 + \Delta t)$ is equivalent to $R(t_0) + \Delta W_0(\Delta t)$ in this case, and furthermore the resulting $R(t_0 + \Delta t) \ge 0$ because the maximum value that we are subtracting from $R(t_0)$ is less than $R(t_0)$, by the definition of the maximum and $\Delta M_0(\Delta t)$. In the other case, $\Delta M_0(\Delta t) > R(t_0)$, which means that $R(t_0) - \Delta W_0(\Delta t)$ crosses through zero, and so this case requires more careful treatment. In fact it is useful to identify the time $\delta t_{\max}$ at which the maximum value of $\Delta W_0(\Delta t)$ occurs, so that $\Delta W_0(\delta t_{\max}) = \Delta M_0(\Delta t)$. This is the time when R reaches zero, $R(t_0 + \delta t_{\max}) = 0$, according to Eqs. (17.549). This is intuitive, as $R(t)$ should hit zero anytime $W(t)$ attains a local maximum, according to its definition. Then continuing the evolution, $R(t_0 + \Delta t)$ is the magnitude of the difference between $W(t_0 + \delta t_{\max})$ and $W(t_0 + \Delta t)$. So $R(t_0 + \Delta t)$ is still positive as required, and if we apply the reflection principle to $\Delta W_0(\Delta t)$ at this time $t_0 + \delta t_{\max}$, the reflected version of $R(t_0 + \Delta t)$ is exactly equivalent to a Wiener process that happened to cross through zero at time $t_0 + \delta t_{\max}$.

Returning to the main point, since $M(t) - W(t)$ behaves like $|W(t)|$, the *last* zero of which correspond to the time $\tau_{\max}$ when $W(t)$ achieves its maximum value. Since $|W(t)|$ has the same set of zeros as $W(t)$, and the last zero has the arcsine distribution, the time $\tau_{\max}$ of the maximum likewise has the arcsine distribution (17.543). Somewhat counterintuitively, then, the Wiener process tends to achieve its maximum value near the beginning or near the end of the time interval in question.

17.12.2 Standard Brownian Bridge

Now we will repeat the above analysis to compute the probability distributions for the sojourn time of a Brownian bridge above a boundary at d.[46] As before, in the plot here, this counts the portion of time that the path is highlighted in green; the only difference is the pinning of the path at $t = 0$ to $B(t) = 0$.

[46] this is the first calculation in Gerard Hooghiemstra, *op. cit.*



We will again take $d \geq 0$ in the analysis here. As before, we define the sojourn time for a bridge $B(t)$ as the functional

$$T_s[B(t);d] := \int_0^1 dt\, \Theta[B(t)-d],$$

(sojourn time for standard Brownian bridge) (17.550)

or the total time that $B(t)$ spends across the boundary. The calculation here is similar to the calculation for the Wiener path, and we will refer to that calculation frequently to save effort. The main difference is that we will take $g(x) = e^{ikx}$ in the diffusion equation (17.502) instead of $g(x) = 1$, and then we will integrate with respect to k to introduce a δ function that will pin the endpoint of the Wiener path to make a bridge. This will lead to somewhat more complicated algebra, but nothing conceptually different.

The procedure is the same as for the Wiener path up to Eq. (17.505), which we will leave as

$$f(x) = \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle g[x+W(t)] \exp\left(-s\int_0^t dt'\, \Theta[x+W(t')-d]\right)\right\rangle\!\!\right\rangle. \tag{17.551}$$

Now for the steady-state PDE (17.506); this time, taking $g(x) = e^{ikx}$ leads to

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\Theta(x-d)f(x) + e^{ikx}, \tag{17.552}$$

which for $x > d$ is

$$f'' = 2(\lambda+s)f - 2e^{ikx}, \tag{17.553}$$

again with the case of $x < d$ given by setting $s = 0$. Then setting $h = f - 2e^{ikx}/[2(\lambda+s)+k^2]$, we have

$$\begin{aligned} h'' &= f'' + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)f - 2e^{ikx} + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)h + \frac{4(\lambda+s)e^{ikx}}{2(\lambda+s)+k^2} - 2e^{ikx} + \frac{2k^2 e^{ikx}}{2(\lambda+s)+k^2} \\ &= 2(\lambda+s)h. \end{aligned} \tag{17.554}$$

so that for $x > d$,

$$h(x) = Ae^{-\sqrt{2(\lambda+s)}\,x}, \tag{17.555}$$

or

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{2e^{ikx}}{2(\lambda+s)+k^2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{2e^{ikx}}{2\lambda+k^2} & (x < d), \end{cases} \tag{17.556}$$

where A and B must be determined by requiring continuity of f and f' at $x = d$. Solving the two resulting equations gives

$$A = \frac{2\sqrt{2}\,s\left(\sqrt{2\lambda} - ik\right)e^{\sqrt{2(\lambda+s)}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}, \qquad B = -\frac{2\sqrt{2}\,s\left(\sqrt{2(\lambda+s)} + ik\right)e^{-\sqrt{2\lambda}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}. \tag{17.557}$$

Now, we can equate Eqs. (17.551) and (17.556) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle e^{ikW(t)}\,\exp\left(-s\int_0^t dt'\,\Theta[W(t')-d]\right)\right\rangle\!\!\right\rangle = B + \frac{2}{2\lambda+k^2}. \tag{17.558}$$

Integrating with respect to k and then dividing through by 2π,

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \delta[W(t)]\,\exp\left(-s\int_0^t dt'\,\Theta[W(t')-d]\right)\right\rangle\!\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk\left(B + \frac{2}{2\lambda+k^2}\right). \tag{17.559}$$

The δ function here picks out only the Wiener paths that return to $W(t) = 0$ (i.e., standard Brownian bridges).

We will treat this carefully to obtain the correct normalization for the ensemble average. In the following, we will use t' as a dummy variable and t as the temporal endpoint, so that $0 \le t' \le t$. We have an ensemble average of a functional $F[W(t')]$ (note that $\delta[W(t)]$ only operates on the final value $W(t)$, and so is not itself a functional),

$$\begin{aligned} \left\langle\!\!\left\langle \delta[W(t)]\,F[W(t')]\right\rangle\!\!\right\rangle &= \sum_{W(t')} P[W(t')]\,\delta[W(t)]\,F[W(t')] \\ &= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')\wedge W(t)=0]}{dW}\,F[W(t')], \end{aligned} \tag{17.560}$$

where we have expanded into a sum over all possible paths, weighted by the probability $P[W(t')]$ of $W(t')$ to occur, and then implemented the δ function in selecting the probabilities for Wiener paths satisfying $0 \le W(t) < dW$, which we write as $W(t) = 0$ for short. The factor of $1/dW$ comes from thinking of $\delta(W = 0) = 1/dW$, with $\delta(W \ne 0) = 0$ for proper normalization of the δ function. Then we change to a conditional probability via $P(A \wedge B) = P(A|B)P(B)$:

$$\begin{aligned} \left\langle\!\!\left\langle \delta[W(t)]\,F[W(t')]\right\rangle\!\!\right\rangle &= \sum_{\{W(t')\}} \frac{P[W(t')|W(t)=0]\,P[W(t)=0]}{dW}\,F[W(t')] \\ &= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')|W(t)=0]\,dW/\sqrt{2\pi t}}{dW}\,F[W(t')] \\ &= \sum_{\{W(t')|W(t)=0\}} \frac{P[W(t')|W(t)=0]}{\sqrt{2\pi t}}\,F[W(t')] \\ &= \sum_{\{B_t(t')\}} \frac{P[B_t(t')]}{\sqrt{2\pi t}}\,F[B_t(t')] \\ &= \frac{1}{\sqrt{2\pi t}}\left\langle\!\!\left\langle F[B_t(t')]\right\rangle\!\!\right\rangle, \end{aligned} \tag{17.561}$$

where we have switched to Brownian bridges, which are equivalent to pinned Wiener paths, and we are using the notation $B_T(t) = \sqrt{T}\,B(t/T)$ for a bridge that is pinned at time $t = T$.

Changing the ensemble average in Eq. (17.559) to Brownian bridges, we have

$$\int_0^\infty \frac{dt}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\Theta[B_t(t')-d]\right)\right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left(B+\frac{2}{2\lambda+k^2}\right), \tag{17.562}$$

where we are using the notation $B_T(t) = \sqrt{T}\,B(t/T)$ for a bridge that closes at time T. Then using Eq. (17.501) on the left-hand side, adapted for the bridge, and (17.557) on the right-hand side,

$$\begin{aligned}\int_0^\infty \frac{dt}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_t]\right\}\right\rangle\!\!\right\rangle &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left[\frac{2}{2\lambda+k^2}-\frac{2\sqrt{2}\,s\left(\sqrt{2(\lambda+s)}+ik\right)e^{-\sqrt{2\lambda}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}\right]\\ &= \sqrt{\frac{\pi}{\lambda}}-\sqrt{\pi}\,\frac{e^{-2\sqrt{2\lambda}\,d}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right).\end{aligned} \tag{17.563}$$

To invert the Laplace transform here, we will use the formula[47]

$$\mathscr{L}\left[\frac{1}{t}e^{-st/2}I_1(st/2)\right](\lambda) = \frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}} \tag{17.564}$$

along with formula (17.520) (with $k = 2\sqrt{2}\,d$) and the Laplace convolution rule (17.521) to write

$$\mathscr{L}\left[\int_0^t d\tau\,\frac{1}{\sqrt{\pi\tau(t-\tau)}}e^{-2d^2/(t-\tau)}e^{-s\tau/2}I_1(s\tau/2)\right] = \frac{e^{-2\sqrt{2\lambda}\,d}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right), \tag{17.565}$$

which takes care of the second term. The first term on the right-hand side of (17.563) is also covered by the formula (17.520) with $k = 0$. Thus, undoing the Laplace transform in Eq. (17.563) gives

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_t]\right\}\right\rangle\!\!\right\rangle = 1-\int_0^t d\tau\,\frac{1}{\tau}\sqrt{\frac{t}{t-\tau}}e^{-2d^2/(t-\tau)}e^{-s\tau/2}I_1(s\tau/2). \tag{17.566}$$

We can simplify things by considering only a standard Brownian bridge $B(t)$ by setting $t = 1$. Doing this and using Eq. (17.501) on the left-hand side,

$$\int_0^1 dx\,e^{-sx}\,f_{T_{\rm s}}(x) = 1-\int_0^1 d\tau\,\frac{1}{\tau\sqrt{1-\tau}}\,e^{-2d^2/(1-\tau)}e^{-s\tau/2}I_1(s\tau/2). \tag{17.567}$$

We will proceed as before by using the integral representation of the Bessel function,[48]

$$I_1(x) = \frac{x}{\pi}\int_{-1}^1 du\,\sqrt{1-u^2}\,e^{-ux} = \frac{4x}{\pi}e^x\int_0^1 dv\,\sqrt{v(1-v)}\,e^{-2vx}, \tag{17.568}$$

where we have changed variables by setting $1 + u = 2v$. Using this in Eq. (17.567), we find

$$\int_0^1 dx\,e^{-sx}\,f_{T_{\rm s}}(x) = 1-\frac{2s}{\pi}\int_0^1 d\tau\int_0^1 dv\,\sqrt{\frac{v(1-v)}{1-\tau}}e^{-2d^2/(1-\tau)}\,e^{-s\tau v}, \tag{17.569}$$

and letting $v = x/\tau$,

$$\begin{aligned}\int_0^1 dx\,e^{-sx}\,f_{T_{\rm s}}(x) &= 1-\frac{2s}{\pi}\int_0^1 d\tau\int_0^\tau dx\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}\,e^{-sx}\\ &= \int_0^1 dx\,e^{-sx}\,\delta(x-0^+)-\frac{2s}{\pi}\int_0^1 dx\,e^{-sx}\int_x^1 d\tau\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}.\end{aligned} \tag{17.570}$$

[47] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1024, Eq. (29.3.52).
[48] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 376, Eq. (9.6.18).

Note how we changed the integration limits when interchanging the order of integration in the last step, in order to integrate over the same triangular area in the (x,τ)-plane. We can regard $f_{T_s}(x)$ here as vanishing for $x > 1$, since it is nonsensical to consider the standard bridge past $t = 1$. If we regard the last term in the same way, we can extend the upper integration limits to ∞ and write out the Laplace transforms as

$$\mathscr{L}\left[f_{T_s}(x)\right] = \mathscr{L}\left[\delta(x-0^+)\right] - s\mathscr{L}\left[\frac{2}{\pi}\int_x^1 d\tau\,\frac{1}{\tau^2}\sqrt{\frac{x(\tau-x)}{1-\tau}}\,e^{-2d^2/(1-\tau)}\right]. \tag{17.571}$$

Recall [Eq. (5.144)] that the Laplace transform of a derivative satisfies

$$s\mathscr{L}[f(x)](s) = \mathscr{L}[f'(x)](s) + f(0), \tag{17.572}$$

so that

$$\begin{aligned}\mathscr{L}\left[f_{T_s}(x)\right] &= \mathscr{L}\left[\delta(x-0^+)\right] - \mathscr{L}\left[\partial_x\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right] - I(d)\\ &= \mathscr{L}\left\{\left[1-I(d)\right]\delta(x-0^+)\right\} - \mathscr{L}\left[\partial_x\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right],\end{aligned} \tag{17.573}$$

where

$$I(d) := \lim_{x\to 0}\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}. \tag{17.574}$$

Thus, we can write the sojourn-time density as

$$f_{T_s}(x) = \left[1-I(d)\right]\delta(x-0^+) - \partial_x\left[\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right]. \tag{17.575}$$

Again, the $[1-I(d)]$ coefficient of the δ function is associated with the probability of having a zero sojourn time. This is the probability of a bridge to *not* touch the boundary. We have already calculated the touching probability in Eq. (17.380), so

$$I(d) = e^{-2d^2}, \tag{17.576}$$

Thus, our first complete expression for the sojourn-time density is

$$f_{T_s}(x) = \left[1-e^{-2d^2}\right]\delta(x-0^+) - \partial_x\left[\frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2}\right] \qquad (0\le x\le 1;\ d>0),$$

(probability density for bridge sojourn time) (17.577)

with cumulative probability[49]

$$P(T_s \le x) = 1 - \frac{2\sqrt{x}}{\pi}\int_x^1 d\tau\,\sqrt{\frac{\tau-x}{1-\tau}}\,\frac{e^{-2d^2/(1-\tau)}}{\tau^2} \qquad (0\le x\le 1;\ d>0),$$

(cumulative probability distribution for bridge sojourn time) (17.578)

though we will continue a bit more in massaging the expression here into nicer forms.

Next, we change variables via

$$\tau = \frac{x(1+\sigma^2)}{1+x\sigma^2}, \tag{17.579}$$

with the result

$$f_{T_s}(x) = \left[1-e^{-2d^2}\right]\delta(x-0^+) - \partial_x\left[\frac{4(1-x)}{\pi}\,e^{-2d^2/(1-x)}\int_0^\infty d\sigma\,\frac{\sigma^2}{(1+\sigma^2)^2}\,e^{-2\sigma^2 d^2 x/(1-x)}\right].$$

(probability density for bridge sojourn time) (17.580)

[49] Gerard Hooghiemstra, *op. cit.*, Eq. (6).

Note that here we can identify

$$\begin{aligned} I(d) &= -\lim_{x\to 0} \frac{4(1-x)}{\pi}\, e^{-2d^2/(1-x)} \int_0^\infty d\sigma\, \frac{\sigma^2}{(1+\sigma^2)^2}\, e^{-2(1+s^2x)d^2} \\ &= \frac{4}{\pi}\, e^{-2d^2} \int_0^\infty d\sigma\, \frac{\sigma^2}{(1+\sigma^2)^2} \\ &= \frac{4}{\pi}\, e^{-2d^2} \int_0^\infty d\sigma\, \frac{\sigma^2}{(1+\sigma^2)^2} \\ &= e^{-2d^2}, \end{aligned} \tag{17.581}$$

which serves as a separate verification of the boundary-crossing result (17.380). Additionally, we can now integrate Eq. (17.580) from 0 to x to obtain the cumulative distribution for the sojourn time. In doing so, the integral of the last term, evaluated at the lower integration limit, will cancel the e^{-2d^2} from the first term, and the result is[50]

$$P(T_{\rm s} \le x) = 1 - \frac{4(1-x)}{\pi}\, e^{-2d^2/(1-x)} \int_0^\infty d\sigma\, \frac{\sigma^2}{(1+\sigma^2)^2}\, e^{-2\sigma^2 d^2 x/(1-x)} \qquad (0 \le x \le 1;\ d \ge 0).$$

(cumulative probability distribution for bridge sojourn time) (17.582)

Finally, we can evaluate the σ integral, with the result[51]

$$P(T_{\rm s} \le x) = 1 + 4d\sqrt{\frac{x(1-x)}{2\pi}}\, e^{-2d^2/(1-x)} - (1 - x + 4d^2x)\, e^{-2d^2}\, \mathrm{erfc}\left(\sqrt{\frac{2d^2x}{1-x}}\right) \qquad (0 \le x \le 1;\ d \ge 0).$$

(cumulative probability distribution for bridge sojourn time) (17.583)

The same integration applied to (17.580) gives the explicit probability density[52]

$$f_{T_{\rm s}}(x) = \left[1 - e^{-2d^2}\right] \delta(x - 0^+) + \sqrt{\frac{8d^2(1-x)}{\pi x}}\, e^{-2d^2/(1-x)} + (1 - 4d^2)\, e^{-2d^2}\, \mathrm{erfc}\left(\sqrt{\frac{2d^2x}{1-x}}\right).$$

(probability density for bridge sojourn time) (17.584)

As usual, these formulae can be extended to a bridge $B_T(t) = \sqrt{T}B(t/T)$ pinned to zero at $t = T$ by the replacements $d \longrightarrow d/\sqrt{T}$ and $x \longrightarrow x/T$. Also, recall that we explicitly assumed $d \ge 0$; for $d < 0$, this formula for $f_{T_{\rm s}}(x)$ can be regarded as the density for $1 - T_{\rm s}$.

We can then compute the moments for the sojourn time as

$$\begin{aligned} \left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle &= \int_0^1 dx\, x^n f_{T_{\rm s}}(x) \\ &= 1 - n \int_0^1 dx\, x^{n-1} P(T_{\rm s} \le x) \\ &= n \int_0^1 dx\, x^{n-1} [1 - P(T_{\rm s} \le x)] \\ &= \frac{2n}{\pi} \int_0^1 dx\, x^{n-1/2} \int_x^1 d\tau\, \sqrt{\frac{\tau - x}{1-\tau}}\, \frac{e^{-2d^2/(1-\tau)}}{\tau^2} \\ &= \frac{2n}{\pi} \int_0^1 d\tau\, \frac{e^{-2d^2/(1-\tau)}}{\tau^2\sqrt{1-\tau}} \int_0^\tau dx\, x^{n-1/2} \sqrt{\tau - x}, \end{aligned} \tag{17.585}$$

where we integrated by parts to integrate with respect to the cumulative distribution, for which we used the

[50] Gerard Hooghiemstra, *op. cit.*, before Eq. (10).

[51] Lajos Takács, "The Distribution of the Sojourn Time for the Brownian Excursion," *Methodology and Computing in Applied Probability* **1**, 7 (1999), Eq. (6) (doi: 10.1023/A:1010060107265).

[52] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

form (17.578). After performing the x integration, the result is[53]

$$\left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle = \frac{n\Gamma[n+1/2]}{\sqrt{\pi}(n+1)!}\int_0^1 d\tau\,\frac{\tau^{n-1}e^{-2d^2/(1-\tau)}}{\sqrt{1-\tau}}. \tag{17.586}$$

Notice that this moment formula also arises from the moment-generating function in the form (17.566), where the power-series expansion of $e^{-s\tau/2}I_1(s\tau/2)$ in s yields the individual moments here. A somewhat easier form to handle arises by changing variables via $\tau = 1-\sigma^2$, with the result

$$\left\langle\!\left\langle (T_{\rm s})^n \right\rangle\!\right\rangle = \frac{2n\Gamma[n+1/2]}{\sqrt{\pi}(n+1)!}\int_0^1 d\sigma\,(1-\sigma^2)^{n-1}e^{-2d^2/\sigma^2}. \tag{17.587}$$
(sojourn-time moments)

The integral here does not have a simple general form, but it can be readily evaluated for particular n. For example, we have

$$\begin{aligned}
\left\langle\!\left\langle T_{\rm s} \right\rangle\!\right\rangle &= \frac{e^{-2d^2}}{2} - \sqrt{\frac{\pi}{2}}\,d\,\mathrm{erfc}\left[\sqrt{2}\,d\right] \\
\left\langle\!\left\langle (T_{\rm s})^2 \right\rangle\!\right\rangle &= \frac{(1+2d^2)\,e^{-2d^2}}{3} - \sqrt{2\pi}d\left(1+\frac{4d^2}{3}\right)\mathrm{erfc}\left[\sqrt{2}\,d\right] \\
\left\langle\!\left\langle (T_{\rm s})^4 \right\rangle\!\right\rangle &= \frac{(12+87d^2+80d^4+16d^6)\,e^{-2d^2}}{60} - \frac{\sqrt{2\pi}d\left(105+420d^2+336d^4+64d^6\right)}{240}\,\mathrm{erfc}\left[\sqrt{2}\,d\right]
\end{aligned} \tag{17.588}$$

for the first, second, and fourth moments of the sojourn time, remembering that $d \geq 0$ in these expressions.

17.12.3 Brownian Bridge

We can now generalize the above treatment to a more general Brownian bridge $B_{0\to c}(t)$, pinned to c at $t=1$, so that in this section we will consider the sojourn-time functional

$$T_{\rm s}[B_{0\to c}(t);d] := \int_0^1 dt\,\Theta[B_{0\to c}(t)-d].$$
(sojourn time for Brownian bridge) (17.589)

To modify the above derivation, after Eq. (17.551), we will choose $g(x) = e^{ik(x-c)}$. The solution of the PDE is then essentially the same, with the result with $\exp(ikx) \longrightarrow \exp[ik(x-c)]$ in (17.556), and $\exp(ikd) \longrightarrow \exp[ik(d-c)]$ in the expression for B in Eqs. (17.557). Then after setting $x=0$ as before, we have

$$\int_0^\infty dt\,e^{-\lambda t}\left\langle\!\left\langle e^{ik[W(t)-c]}\,\exp\left(-s\int_0^t dt'\,\Theta[W(t')-d]\right)\right\rangle\!\right\rangle = B + \frac{2e^{-ikc}}{2\lambda+k^2}. \tag{17.590}$$

in place of (17.590), where we should keep in mind that the form of B is modified here. Then integrating with respect to k and dividing through by 2π gives

$$\int_0^\infty dt\,e^{-\lambda t}\left\langle\!\left\langle \delta[W(t)-c]\,\exp\left(-s\int_0^t dt'\,\Theta[W(t')-d]\right)\right\rangle\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk\left(B+\frac{2e^{-ikc}}{2\lambda+k^2}\right). \tag{17.591}$$

in place of Eq. (17.559). Thus we have introduced the correct modification to force the δ function to pin the Wiener paths that return to $W(t)=c$.

[53] Gerard Hooghiemstra, *op. cit.*, Eq. (10).

Next, we should generalize the result (17.561) to the case of pinning $W(t)$ to c:

$$\begin{aligned}
\left\langle\!\!\left\langle \delta[W(t)-c]\,F[W(t')] \right\rangle\!\!\right\rangle &= \sum_{W(t')} P[W(t')]\,\delta[W(t)-c]\,F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')\wedge W(t)=c]}{dW}\,F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')|W(t)=c]\,P[W(t)=c]}{dW}\,F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} \frac{P[W(t')|W(t)=c]\,e^{-c^2/2t}\,dW/\sqrt{2\pi t}}{dW}\,F[W(t')] \\
&= \sum_{\{W(t')|W(t)=c\}} P[W(t')|W(t)=c]\,\frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\,F[W(t')] \\
&= \sum_{\{B_t(t')\}} P[B_{t(0\to c)}(t')]\,\frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\,F[B_{t(0\to c)}(t')] \\
&= \frac{e^{-c^2/2t}}{\sqrt{2\pi t}} \left\langle\!\!\left\langle F[B_{t(0\to c)}(t')] \right\rangle\!\!\right\rangle.
\end{aligned} \tag{17.592}$$

The change here is straightforward, and involves evaluating the Gaussian probability density for $W(t)$ at c instead of 0.

Then changing the average in Eq. (17.591) to encompass the appropriate Brownian bridges, we have

$$\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\Theta[B_{t(0\to c)}(t')-d]\right) \right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left(B+\frac{2e^{-ikc}}{2\lambda+k^2}\right), \tag{17.593}$$

in place of Eq. (17.562). Carrying out the following integration gives

$$\begin{aligned}
&\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t} \left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\} \right\rangle\!\!\right\rangle \\
&\qquad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left[\frac{2e^{-ikc}}{2\lambda+k^2} - \frac{2\sqrt{2}\,s\left(\sqrt{2(\lambda+s)}+ik\right)e^{-\sqrt{2\lambda}\,d+ik(d-c)}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}\right] \\
&\qquad = \left[\sqrt{\frac{\pi}{\lambda}}\,e^{-\sqrt{2\lambda}\,|c|} - \sqrt{\pi}\,\frac{e^{-\sqrt{2\lambda}\,(2d-c)}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right)\right] \qquad (c\le d,\ d\ge 0)
\end{aligned} \tag{17.594}$$

in place of Eqs. (17.563) in the case $c \le d$. If $c \ge d$ this is instead

$$\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t} \left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\} \right\rangle\!\!\right\rangle = \sqrt{4\pi}\,\frac{e^{-\sqrt{2\lambda}\,d-\sqrt{2(\lambda+s)}\,(c-d)}}{\sqrt{\lambda+s}+\sqrt{\lambda}} \qquad (c\ge d\ge 0). \tag{17.595}$$

The first expression (17.594) is the same result as in Eq. (17.563), except for a factor $\exp(-c^2/2t)$ on the left-hand side, the factor of $\exp(-\sqrt{2\lambda}\,c)$ in the first term on the right-hand side, and the replacement $2d \longrightarrow 2d-c$ in the second term on the right-hand side. To invert the Laplace transform, the second term on the right-hand side inverts in the same way with the replacement $2d \longrightarrow 2d-c$, while Eq. (17.520) applies to the first term with $k = \sqrt{2}\,|c|$, such that Eq. (17.566) becomes

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\} \right\rangle\!\!\right\rangle = 1 - e^{c^2/2t}\int_0^t d\tau\,\frac{1}{\tau}\sqrt{\frac{t}{t-\tau}}e^{-(2d-c)^2/2(t-\tau)}e^{-s\tau/2}I_1(s\tau/2) \qquad (c\le d, d\ge 0), \tag{17.596}$$

or that is, the last term on the right-hand side is multiplied by $\exp(c^2/2t)$, and subject to the replacement $2d \longrightarrow 2d - c$. The same modifications (with $t = 1$ for the Brownian bridge we want here) carry through the rest of the treatment in the previous section. In particular, Eq. (17.574) becomes

$$\begin{aligned} I(d) &:= \lim_{x \to 0} \frac{2\sqrt{x}}{\pi} e^{c^2/2} \int_x^1 d\tau \sqrt{\frac{\tau - x}{1 - \tau}} \frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2} \\ &= e^{c^2/2} e^{-(2d-c)^2/2} \\ &= e^{-2d(d-c)}, \end{aligned} \tag{17.597}$$

which is the correct boundary-crossing probability (17.387) for the same bridge pinned to c. Then we can adapt the probability-density expressions with these modifications, with the result[54]

$$\begin{aligned} f_{T_s}(x) &= \left[1 - e^{-2d(d-c)}\right] \delta(x - 0^+) - \partial_x \left[\frac{2\sqrt{x}}{\pi} e^{c^2/2} \int_x^1 d\tau \sqrt{\frac{\tau - x}{1 - \tau}} \frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2} \right] \\ &= \left[1 - e^{-2d(d-c)}\right] \delta(x - 0^+) \\ &\qquad - \partial_x \left[\frac{4(1-x)}{\pi} e^{c^2/2 - (2d-c)^2/2(1-x)} \int_0^\infty d\sigma \frac{\sigma^2}{(1+\sigma^2)} e^{-\sigma^2 (2d-c)^2 x/2(1-x)} \right] \\ &= \left[1 - e^{-2d(d-c)}\right] \delta(x - 0^+) + (2d - c) \sqrt{\frac{2(1-x)}{\pi x}} e^{c^2/2 - (2d-c)^2/2(1-x)} \\ &\qquad + \left[1 - (2d-c)^2\right] e^{-2d(d-c)} \operatorname{erfc}\left(\sqrt{\frac{(2d-c)^2 x}{2(1-x)}} \right) \qquad (0 \le x \le 1;\ c \le d;\ d \ge 0). \end{aligned}$$

(probability density for bridge sojourn time) (17.598)

Similarly, the cumulative-probability expressions become

$$\begin{aligned} P(T_s \le x) &= 1 - \frac{2\sqrt{x}}{\pi} e^{c^2/2} \int_x^1 d\tau \sqrt{\frac{\tau - x}{1 - \tau}} \frac{e^{-(2d-c)^2/2(1-\tau)}}{\tau^2} \qquad (0 \le x \le 1;\ c \le d;\ d \ge 0) \\ &= 1 - \frac{4(1-x)}{\pi} e^{c^2/2 - (2d-c)^2/2(1-x)} \int_0^\infty d\sigma \frac{\sigma^2}{(1+\sigma^2)^2} e^{-\sigma^2 (2d-c)^2 x/2(1-x)} \\ &= 1 + (2d - c) \sqrt{\frac{2x(1-x)}{\pi}} e^{c^2/2 - (2d-c)^2/2(1-x)} \\ &\qquad - \left[1 - x + (2d-c)^2 x\right] e^{-2d(d-c)} \operatorname{erfc}\left(\sqrt{\frac{(2d-c)^2 x}{2(1-x)}} \right), \end{aligned}$$

(cumulative probability distribution for bridge sojourn time) (17.599)

and the moment formula (17.587) becomes

$$\left\langle\!\left\langle (T_s)^n \right\rangle\!\right\rangle = \frac{2n\Gamma[n + 1/2]}{\sqrt{\pi}(n+1)!} \int_0^1 d\sigma\, (1 - \sigma^2)^{n-1} e^{c^2/2 - (2d-c)^2/2\sigma^2} \qquad (c \le d;\ d \ge 0)$$

(sojourn-time moments) (17.600)

under the same replacements. For example, the explicit expression for the mean is

$$\left\langle\!\left\langle T_s \right\rangle\!\right\rangle = \frac{e^{-2d(d-c)}}{2} - \sqrt{\frac{\pi}{8}} (2d - c)\, e^{c^2/2} \operatorname{erfc}\left[\frac{2d - c}{\sqrt{2}} \right] \qquad (c \le d;\ d \ge 0), \tag{17.601}$$

which reduces to the standard-bridge mean in Eqs. (17.588).

[54] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

In the case $c \geq d$, the inversion of the Laplace transform (17.595) is somewhat different. First using[55]

$$\mathscr{L}\left[\frac{e^{-st/2}}{2\sqrt{s}\,t} I_{1/2}(st/2)\right](\lambda) = \mathscr{L}\left[\frac{1-e^{-st}}{2s\sqrt{\pi t^3}}\right](\lambda) = \frac{1}{\sqrt{\lambda+s}+\sqrt{\lambda}}, \tag{17.602}$$

where $I_\nu(x)$ is a modified Bessel function, along with[56]

$$\mathscr{L}\left[\frac{k}{2\sqrt{\pi t^3}}\, e^{-k^2/4t}\right](\lambda) = e^{-k\sqrt{\lambda}}, \tag{17.603}$$

these transforms can be combined via the convolution formula (17.521) to give

$$\mathscr{L}\left[\int_0^t d\tau\, \frac{k}{4\pi s\sqrt{\tau^3(t-\tau)^3}}\, e^{-k^2/4(t-\tau)} \left(1-e^{-s\tau}\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{17.604}$$

Then using the shifted-argument version of Eq. (17.603),

$$\mathscr{L}\left[\frac{k'}{2\sqrt{\pi t^3}}\, e^{-st}\, e^{-k'^2/4t}\right](\lambda) = e^{-k'\sqrt{\lambda+s}}, \tag{17.605}$$

we can again employ convolution to combine this with Eq. (17.604) to give

$$\mathscr{L}\left[\int_0^t d\sigma \int_0^{t-\sigma} d\tau\, \frac{kk'}{8\pi^{3/2} s\sqrt{\tau^3(t-\sigma-\tau)^3\sigma^3}}\, e^{-k^2/4(t-\sigma-\tau)}\, e^{-k'^2/4\sigma} \left(1-e^{-s\tau}\right) e^{-s\sigma}\right](\lambda) = \frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{17.606}$$

Thus, setting $k = \sqrt{2}\,d$ and $k' = \sqrt{2}\,(c-d)$, we may invert the Laplace transform in Eq. (17.595) to give

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2t}\, \frac{d(c-d)\sqrt{t}}{2\pi s} \int_0^t d\sigma \int_0^{t-\sigma} d\tau\, \frac{e^{-d^2/2(t-\sigma-\tau)}\, e^{-(c-d)^2/2\sigma}}{\sqrt{\tau^3(t-\sigma-\tau)^3\sigma^3}} \left(1-e^{-s\tau}\right) e^{-s\sigma} \qquad (c \geq d \geq 0). \tag{17.607}$$

We can simplify this to $t = 1$ for a Brownian bridge that runs over a unit time interval:

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2}\, \frac{d(c-d)}{2\pi s} \int_0^1 d\sigma \int_0^{1-\sigma} d\tau\, \frac{e^{-d^2/2(1-\sigma-\tau)}\, e^{-(c-d)^2/2\sigma}}{\sqrt{\tau^3(1-\sigma-\tau)^3\sigma^3}} \left(1-e^{-s\tau}\right) e^{-s\sigma} \qquad (c \geq d \geq 0). \tag{17.608}$$

Now using

$$\frac{e^{-s\sigma}-e^{-s(\sigma+\tau)}}{s} = \int_\sigma^{\sigma+\tau} dx\, e^{-sx} \tag{17.609}$$

to remove the difference in the final factor, while introducing a new integral, we can shift the order of integration according to

$$\int_0^1 d\sigma \int_0^{1-\sigma} d\tau \int_\sigma^{\sigma+\tau} dx = \int_0^1 d\sigma \int_\sigma^1 dx \int_{x-\sigma}^{1-\sigma} d\tau = \int_0^1 dx \int_0^x d\sigma \int_{x-\sigma}^{1-\sigma} d\tau, \tag{17.610}$$

and shifting $\tau \longrightarrow \tau - \sigma$, Eq. (17.608) becomes

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2}\, \frac{d(c-d)}{2\pi} \int_0^1 dx\, e^{-sx} \int_0^x d\sigma \int_x^1 d\tau\, \frac{e^{-d^2/2(1-\tau)-(c-d)^2/2\sigma}}{\sqrt{(\tau-\sigma)^3(1-\tau)^3\sigma^3}} \qquad (c \geq d \geq 0). \tag{17.611}$$

[55] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1024, Eq. (29.3.53).
[56] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.82).

Now the Laplace transform in s may be inverted to give the probability density:

$$f_{T_s}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi}\int_0^x d\sigma \int_x^1 d\tau\,\frac{e^{-d^2/2(1-\tau)-(c-d)^2/2\sigma}}{\sqrt{(\tau-\sigma)^3(1-\tau)^3\sigma^3}} \qquad (0\le x\le 1; c\ge d\ge 0). \tag{17.612}$$

Changing variables via $\tau \longrightarrow 1-\tau$ gives

$$f_{T_s}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi}\int_0^x d\sigma \int_0^{1-x} d\tau\,\frac{e^{-d^2/2\tau-(c-d)^2/2\sigma}}{\sqrt{(1-\sigma-\tau)^3\sigma^3\tau^3}} \qquad (0\le x\le 1; c\ge d\ge 0), \tag{17.613}$$

and then letting $u = x/\sigma$ and $v = (1-x)/\tau$ leads to the alternate integral expression

$$f_{T_s}(x) = e^{c^2/2}\,\frac{d(c-d)}{2\pi\sqrt{x(1-x)}}\int_1^\infty du \int_1^\infty dv\,\frac{uv\,e^{-d^2v/2(1-x)-(c-d)^2u/2x}}{[uv-(1-x)u-xv]^{3/2}} \qquad (0\le x\le 1; c\ge d\ge 0), \tag{17.614}$$

from where it is difficult to proceed with the integration.

There is a second approach to inverting the Laplace transforms here that will lead to expression with only a single integral, in both the density and the Laplace transform of the density.[57] As an alternate form of the right-hand side in Eq. (17.606), we can consider

$$\begin{aligned}\frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}} &= \frac{1}{s}\left(\sqrt{\lambda+s}-\sqrt{\lambda}\right)e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}} \\ &= \frac{1}{s}\,e^{-k\sqrt{\lambda}}\left(\sqrt{\lambda+s}\,e^{-k'\sqrt{\lambda+s}}\right) - \frac{1}{s}\,e^{-k'\sqrt{\lambda+s}}\left(\sqrt{\lambda}\,e^{-k\sqrt{\lambda}}\right).\end{aligned} \tag{17.615}$$

Using again the Laplace transform (17.603) and its shifted version (17.605), along with[58]

$$\mathscr{L}\left[\frac{k^2-2t}{4\sqrt{\pi t^5}}\,e^{-k^2/4t}\right](\lambda) = \sqrt{\lambda}\,e^{-k\sqrt{\lambda}}, \tag{17.616}$$

and the shifted version

$$\mathscr{L}\left[\frac{k'^2-2t}{4\sqrt{\pi t^5}}\,e^{-st}\,e^{-k'^2/4t}\right](\lambda) = \sqrt{\lambda+s}\,e^{-k'\sqrt{\lambda+s}}, \tag{17.617}$$

the convolution theorem allows us to combine these transforms to form the right-hand side of Eq. (17.615), with the result

$$\mathscr{L}\left[\int_0^t d\tau\,\frac{k(t-\tau)(k'^2-2\tau)-k'\tau[k^2-2(t-\tau)]}{8\pi s\sqrt{\tau^5(t-\tau)^5}}\,e^{-k^2/4(t-\tau)}\,e^{-k'^2/4\tau}\,e^{-s\tau}\right](\lambda) = \frac{e^{-k\sqrt{\lambda}-k'\sqrt{\lambda+s}}}{\sqrt{\lambda+s}+\sqrt{\lambda}}. \tag{17.618}$$

Then again setting $k = \sqrt{2}\,d$ and $k' = \sqrt{2}\,(c-d)$, we may invert the Laplace transform in Eq. (17.595) to give

$$\begin{aligned}\left\langle\!\!\left\langle \exp\left\{-sT_s[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = e^{c^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{d(t-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}} \\ \times\; e^{-d^2/2(t-\tau)}\,e^{-(c-d)^2/2\tau}\,e^{-s\tau} \qquad (c\ge d\ge 0),\end{aligned} \tag{17.619}$$

or for $t = 1$,

$$\begin{aligned}\left\langle\!\!\left\langle \exp\left\{-sT_s[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = \frac{e^{c^2/2}}{\sqrt{2\pi}s}\int_0^1 d\tau\,\frac{d(1-\tau)[(c-d)^2-\tau]-(c-d)\tau[d^2-(1-\tau)]}{\sqrt{\tau^5(1-\tau)^5}} \\ \times\; e^{-d^2/2(1-\tau)}\,e^{-(c-d)^2/2\tau}\,e^{-s\tau} \qquad (c\ge d\ge 0).\end{aligned} \tag{17.620}$$

[57] Vadim Linetsky, "Step Options," *Mathematical Finance* **9**, 55 (2001) (doi: 10.1111/1467-9965.00063). See in particular Eq. (C.9) and the discussion just before Eq. (A.9).

[58] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.87) for $n = 2$, which $H_2(x) = 2(2x^2-1)$.

Then writing the s-dependence as

$$\frac{e^{-s\tau}}{s} = \int_\tau^\infty dx\, e^{-sx} = \int_0^\infty dx\, e^{-sx} - \int_0^\tau dx\, e^{-sx}, \tag{17.621}$$

and then massaging the two resulting integrals (with the rest of the integrand suppressed for brevity) via

$$\int_0^1 d\tau \int_0^\infty dx\, e^{-sx} - \int_0^1 d\tau \int_0^\tau dx\, e^{-sx} = \int_0^\infty dx\, e^{-sx} \int_0^1 d\tau - \int_0^1 dx\, e^{-sx} \int_x^1 d\tau = \int_0^1 dx\, e^{-sx} \int_0^x d\tau, \tag{17.622}$$

where we have made use of the fact that the sojourn-time density has its support on $[0, 1]$. Thus we can now invert the Laplace transform (17.620) of the sojourn-time density to obtain

$$f_{T_s}(x) = \frac{e^{c^2/2}}{\sqrt{2\pi}} \int_0^x d\tau\, \frac{d(1-\tau)[(c-d)^2-\tau] - (c-d)\tau[d^2-(1-\tau)]}{\sqrt{\tau^5(1-\tau)^5}}\, e^{-d^2/2(1-\tau)-(c-d)^2/2\tau} \qquad (0 \le x \le 1;\ c \ge d \ge 0). \tag{17.623}$$

This integral can be performed analytically, with result[59]

$$f_{T_s}(x) = \sqrt{\frac{2}{\pi}}\, \frac{(c-d)x + d(1-x)}{\sqrt{x(1-x)}}\, \exp\left(\frac{c^2}{2} - \frac{(c-d)^2}{2x} - \frac{d^2}{2(1-x)}\right) \qquad (0 \le x \le 1;\ c \ge d \ge 0)$$
$$+ \left[1-(2d-c)^2\right] e^{-2d(d-c)}\, \mathrm{erfc}\left(\frac{(c-d)(1-x) + d\,x}{\sqrt{2x(1-x)}}\right), \tag{17.624}$$

(probability density for bridge sojourn time)

as can be verified by differentiating this expression to obtain the integrand of Eq. (17.623). This then gives the density for the boundary-crossing case of the bridge sojourn time.

The cases with $d < 0$ can be generated from the cases with $d \ge 0$ by replacing $x \longrightarrow 1 - x$. And again, all of these expressions can be generalized to a bridge $B_{t(0\to c)}(t)$ that is pinned to c at time T via the replacements $c \longrightarrow c/\sqrt{t}$, $d \longrightarrow d/\sqrt{t}$, and $x \longrightarrow x/t$. Overall factors of t must be restored to make the density come out with "units" of $1/t$ [including, e.g., an overall factor of $\sqrt{t}$ that came from the path-pinning factor $e^{c^2/2}$ in Eq. (17.593)]. Also, a bridge starting at a different location a than 0 can be obtained by shifting c and d by $-a$. Writing these out explicitly from Eqs. (17.598) and (17.624),

$$f_{T_s}(x) = \left[1 - e^{-2(d-a)(d-c)/t}\right] \delta(x - 0^+) + (2d-a-c)\sqrt{\frac{2(t-x)}{\pi t^3 x}}\, e^{(c-a)^2/2t - (2d-a-c)^2/2(t-x)}$$
$$+ \frac{1}{t}\left[1 - \frac{(2d-a-c)^2}{t}\right] e^{-2(d-a)(d-c)/t}\, \mathrm{erfc}\left(\sqrt{\frac{(2d-a-c)^2 x}{2t(t-x)}}\right) \qquad (0 \le x \le t;\ a \le d;\ c \le d)$$
$$f_{T_s}(x) = \sqrt{\frac{2}{\pi}}\, \frac{(c-d)x + (d-a)(t-x)}{\sqrt{t^3 x(t-x)}}\, e^{(c-a)^2/2t - (c-d)^2/2x - (d-a)^2/2(t-x)} \qquad (0 \le x \le t;\ a \le d \le c)$$
$$+ \frac{1}{t}\left[1 - \frac{(2d-a-c)^2}{t}\right] e^{-2(d-a)(d-c)/t}\, \mathrm{erfc}\left(\frac{(c-d)(t-x) + (d-a)\,x}{\sqrt{2tx(t-x)}}\right), \tag{17.625}$$

(probability density for bridge sojourn time)

[59] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.8.

while to obtain the other cases, we can change the signs of a, c, and d, while replacing x with $t-x$,

$$
\begin{aligned}
f_{T_s}(x) &= \left[1-e^{-2(a-d)(c-d)/t}\right]\delta(t-x-0^+) + (a+c-2d)\sqrt{\frac{2x}{\pi t^3(t-x)}}\,e^{(a-c)^2/2t-(a+c-2d)^2/2x} \\
&\quad + \frac{1}{t}\left[1-\frac{(a+c-2d)^2}{t}\right]e^{-2(a-d)(c-d)/t}\,\mathrm{erfc}\left(\sqrt{\frac{(a+c-2d)^2(t-x)}{2tx}}\right) \quad (0\le x\le t;\ d\le c;\ d\le a) \\
f_{T_s}(x) &= \sqrt{\frac{2}{\pi}}\,\frac{(a-d)x+(d-c)(t-x)}{\sqrt{t^3x(t-x)}}\,e^{(a-c)^2/2t-(a-d)^2/2x-(d-c)^2/2(t-x)} \qquad (0\le x\le t;\ c\le d\le a) \\
&\quad + \frac{1}{t}\left[1-\frac{(a+c-2d)^2}{t}\right]e^{-2(a-d)(c-d)/t}\,\mathrm{erfc}\left(\frac{(a-d)(t-x)+(d-c)\,x}{\sqrt{2tx(t-x)}}\right),
\end{aligned}
$$

(probability density for bridge sojourn time, reflected cases) (17.626)

which takes advantage of the fact that the problem is reflection symmetric if the occupation time is taken to be the non-occupation time.

It is also useful to write out the moment-generating functions corresponding to these probability densities. We have the explicit expression for ($c\ge d\ge 0$) in Eq. (17.619), but we bypassed the solution for ($c\le d, d\ge 0$) by generalizing the treatment from the standard Brownian bridge. In fact, it is best to replace that treatment with one that avoids obtaining a Bessel function, as we will do here. First, to set up the Laplace transforms, we will need to arrive at something of the form

$$
\begin{aligned}
\frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right) &= \frac{e^{-k\sqrt{\lambda}}}{s\sqrt{\lambda}}\left(2\lambda+s-2\sqrt{\lambda(\lambda+s)}\right) \\
&= \frac{2}{s}\left(\sqrt{\lambda}-\sqrt{\lambda+s}\right)e^{-k\sqrt{\lambda}} + \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}.
\end{aligned}
\tag{17.627}
$$

Then using the Laplace-transform formula[60]

$$
\mathscr{L}\left[\frac{1}{2\sqrt{\pi t^3}}\left(e^{-st}-1\right)\right](\lambda) = \sqrt{\lambda}-\sqrt{\lambda+s}, \tag{17.628}
$$

we can combine this with the transform formula (17.603) using the convolution theorem to obtain the formula

$$
\mathscr{L}\left[\int_0^t d\tau\,\frac{k}{4\pi\sqrt{\tau^3(t-\tau)^3}}\,e^{-k^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) = \left(\sqrt{\lambda}-\sqrt{\lambda+s}\right)e^{-k\sqrt{\lambda}}. \tag{17.629}
$$

Then adding this to the transform formula (17.520) to hit the last term on the right-hand side of Eq. (17.627),

$$
\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}\,e^{-k^2/4t} + \int_0^t d\tau\,\frac{k}{2\pi s\sqrt{\tau^3(t-\tau)^3}}\,e^{-k^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right). \tag{17.630}
$$

Changing k to k' and combining this again with Eq. (17.520) gives

$$
\begin{aligned}
&\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}\left(e^{-k^2/4t}-e^{-k'^2/4t}\right) - \int_0^t d\tau\,\frac{k'}{2\pi s\sqrt{\tau^3(t-\tau)^3}}\,e^{-k'^2/4\tau}\left(e^{-s(t-\tau)}-1\right)\right](\lambda) \\
&\qquad = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}} - \frac{e^{-k'\sqrt{\lambda}}}{\sqrt{\lambda}}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right).
\end{aligned}
\tag{17.631}
$$

[60] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1023, Eq. (29.3.36) for $n=2$, which $H_2(x) = 2(2x^2-1)$.

Then using this to invert Eq. (17.594) with $k = \sqrt{2}\,|c|$ and $k' = \sqrt{2}(2d - c)$, we find[61]

$$\left\langle\!\!\left\langle \exp\left\{-sT_{\mathrm{s}}[B_{t(0\to c)}]\right\}\right\rangle\!\!\right\rangle = 1 - e^{-2d(d-c)/t} \qquad (c \le d,\ d \ge 0)$$
$$+ e^{c^2/2t}\frac{\sqrt{t}(2d-c)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(2d-c)^2/2\tau}\left(1 - e^{-s(t-\tau)}\right), \tag{17.632}$$

The non-integral part of the expression here is related to the boundary-touching probability of the bridge. Note that the normalization can be verified by taking the limit $s \longrightarrow 0$ and carrying out the resulting integral, recovering the unit-normalization result. Now putting this together with Eq. (17.619) and putting in a starting point of a,

$$\left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}}\right\rangle\!\!\right\rangle = 1 - e^{-2(d-a)(d-c)/t} \qquad (a \le d;\ c \le d)$$
$$+ e^{(c-a)^2/2t}\frac{\sqrt{t}(2d-a-c)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(2d-a-c)^2/2\tau}\left(1 - e^{-s(t-\tau)}\right)$$
$$\left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}}\right\rangle\!\!\right\rangle = e^{(c-a)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(d-a)(t-\tau)[(c-d)^2-\tau] - (c-d)\tau[(d-a)^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}$$
$$\times\, e^{-(d-a)^2/2(t-\tau) - (c-d)^2/2\tau - s\tau} \qquad (a \le d \le c)$$
$$\left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}}\right\rangle\!\!\right\rangle = \left[1 - e^{-2(a-d)(c-d)/t}\right]e^{-st} \qquad (d \le c;\ d \le a)$$
$$+ e^{(a-c)^2/2t}\frac{\sqrt{t}(a+c-2d)}{\sqrt{2\pi}\,s}\int_0^t d\tau\,\frac{1}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(a+c-2d)^2/2\tau}\left(e^{-s\tau} - e^{-st}\right)$$
$$\left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}}\right\rangle\!\!\right\rangle = e^{(a-c)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(d-c)(t-\tau)[(a-d)^2-\tau] - (a-d)\tau[(d-c)^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}$$
$$\times\, e^{-(d-c)^2/2(t-\tau) - (a-d)^2/2\tau - s\tau} \qquad (c \le d \le a).$$

(sojourn-time moment generating function) (17.633)

The last two cases here follow from the first two via the symmetry of the problem. Specifically, ''flipping'' the geometry by reversing the signs of a, c, and d allows us to calculate the ''complimentary'' generating function $\langle\!\langle e^{-s(t-T_{\mathrm{s}})}\rangle\!\rangle$. Thus in addition to reversing these signs, we also need to reverse the sign of s in the first two expressions and then multiply through by e^{-st} to obtain the desired generating function. In the fourth expression, we also changed integration variables, letting $\tau \longrightarrow t - \tau$, to bring it into a form more similar to the second expression. In this case it turns out to have the same form, but with a and c interchanged.

The singularity of the integrand in Eqs. (17.633) can be problematic, and merits some further discussion. For example, in the case of the second expression with $d = a$, the integral naïvely becomes

$$\left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}}\right\rangle\!\!\right\rangle = e^{(c-a)^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}s}\int_0^t d\tau\,\frac{(c-a)}{\sqrt{\tau^3(t-\tau)^3}}\,e^{-(c-a)^2/2\tau - s\tau} \qquad (a \le c;\ d = a). \tag{17.634}$$

However, this integral is ill-defined because the integrand diverges at $t = \tau$ (note that it is okay at $\tau = 0$ provided $c \neq a$). Similarly, the expression is problematic if instead $d = c$, because the integrand diverges at $\tau = 0$. Thus, some cases above may require some caution in their evaluation. One solution to regularize this example[62] is to multiply through by s in Eq. (17.619), and then take the limit $s \longrightarrow 0$ to obtain

$$e^{c^2/2t}\frac{\sqrt{t}}{\sqrt{2\pi}}\int_0^t d\tau\,\frac{d(t-\tau)[(c-d)^2-\tau] - (c-d)\tau[d^2-(t-\tau)]}{\sqrt{\tau^5(t-\tau)^5}}\,e^{-d^2/2(t-\tau)}\,e^{-(c-d)^2/2\tau} = 0 \qquad (c \ge d \ge 0). \tag{17.635}$$

[61] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 158, formula 1.4.7.

[62] Vadim Linetsky, *op. cit.*, in the discussion after Eq. (C.9).

[That the result vanishes may be somewhat more obvious from performing the same maneuver in Eq. (17.618).] After restoring the initial point a, this is a useful counterpart to the second and fourth expressions in Eqs. (17.633). For example, multiplying Eq. (17.635) by e^{-st}/s and subtracting from the second equation in Eqs. (17.633) leads to a similar expression, but with the replacement $e^{-s\tau} \longrightarrow e^{-s\tau} - e^{-st}$. This cures the problematic case (17.634), which becomes

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{\sqrt{t}}{\sqrt{2\pi}s} \int_0^t d\tau \, \frac{(c-a)}{\sqrt{\tau^3(t-\tau)^3}} \, e^{-(c-a)^2(\tau^{-1}-t^{-1})/2} \left[e^{-s\tau} - e^{-st} \right] \qquad (a \le c;\ d = a) \tag{17.636}$$

after this regularization; this cures the divergence at $t = \tau$, which is now cut off by the difference in exponentials. Note that this is the same expression that follows from the third expression in Eqs. (17.633). To cure the divergence in the case $d = c$, it is better to multiply Eq. (17.635) by $1/s$ and subtract from the second equation in Eqs. (17.633); this again leads to a similar expression, but with the replacement $e^{-s\tau} \longrightarrow e^{-s\tau} - 1$, cutting off the divergence at $\tau = 0$. However, the integrals are still pathological when $a = d = c$, in which case it is easiest to setting $d = 0$ in Eq. (17.566), in which case the result is

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{1 - e^{-st}}{st} \qquad (a = c = d). \tag{17.637}$$

To see how this limit follows directly from Eq. (17.636), note that as $c \longrightarrow a$, the factor of $(c-a)$ makes the integrand small everywhere except for the range of small τ—the divergent factor $\tau^{-3/2}$ is important here, and although it is cut off by the Gaussian factor $e^{-(c-a)^2/2\tau}$, it is still important for $\tau \sim \sqrt{c-a}$. Thus, since the significant part of the integrand moves towards vanishingly small τ, we can expand the integrand in τ to obtain

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle = \frac{1}{\sqrt{2\pi}st} \int_0^t d\tau \left[\frac{(c-a)}{\tau^{3/2}} \, e^{-(c-a)^2(\tau^{-1}-t^{-1})/2} \big(1 - e^{-st}\big) + O\big(\tau^{-1/2}\big) \right] \qquad (a \le c;\ d = a). \tag{17.638}$$

This integral may then be carried out, with result

$$\left\langle\!\left\langle e^{-sT_s} \right\rangle\!\right\rangle \approx e^{(c-a)^2/2t} \left(\frac{1 - e^{-st}}{st} \right) \operatorname{erfc}\left[\frac{c-a}{\sqrt{2t}} \right] \qquad (a \le c;\ d = a). \tag{17.639}$$

At this point, the limit $c \longrightarrow a$ yields Eq. (17.637).

In general, the moments must then be computed by integrating the probability density or differentiating the moment-generating functions listed above. Although we wrote out an integral for $c \le d$ and $d \ge 0$ in Eq. (17.600), the case for $0 \le d \le c$ is not as simple. For the mean sojourn time, however, it is possible to work out a relatively nice expression by integrating the local time. The result is given in Eq. (17.724) as

$$\begin{aligned} \left\langle\!\left\langle T_s[B_{a\to c}(t); d] \right\rangle\!\right\rangle = \frac{t}{2} + \operatorname{sgn}(2d - a - c) \, \frac{t}{2} \left[e^{-2[(d-a)(d-c)\,\Theta(d-a)\,\Theta(d-c) + (a-d)(c-d)\,\Theta(a-d)\,\Theta(c-d)]/t} - 1 \right] \\ - \sqrt{\frac{\pi t}{8}} \, (2d - a - c) \, e^{(c-a)^2/2t} \operatorname{erfc}\left(\frac{|d-a| + |d-c|}{\sqrt{2t}} \right), \end{aligned} \tag{17.640}$$

(mean sojourn time)

after restoring the initial point a of the Brownian bridge. While it may not be obvious from this expression, when d is in the interval (a, c), the second term does not contribute, and the last term leads to a decreasing, straight-line dependence on d.

17.12.4 Path-Pinning Normalization as a Constrained Integration

In deriving sojourn-time statistics, we made use of the relation (17.592) to remove the delta function that pinned the endpoint of a Wiener path such that $W(T) = c$:

$$\left\langle\!\left\langle \delta[W(T) - c] \, F[W(t)] \right\rangle\!\right\rangle = \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \left\langle\!\left\langle F\big[B_{T(0\to c)}(t)\big] \right\rangle\!\right\rangle. \tag{17.641}$$

We also used this in restricted form in Eqs. (17.561). In both cases we used probabilistic arguments to justify this relation. However, it is also useful to see how this arises as a constrained integration problem, by considering the explicit probability measure of the path. To begin, suppose that we consider N discrete steps $(\Delta W_0, \ldots, \Delta W_{N-1})$ of the Wiener path, each of time step $\Delta T = T/N$, and write out the explicit integration over the multidimensional Gaussian probability density:

$$\begin{aligned}
&\left\langle\!\!\left\langle \delta[W(T) - c]\, F[W(t)] \right\rangle\!\!\right\rangle \\
&\qquad = \int d\Delta W_0 \ldots d\Delta W_{N-1}\, \frac{1}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\big(\Delta W_j\big)^2\right]\, \delta\left[\sum_{j=0}^{N-1}\Delta W_j - c\right]\, F[W(t)].
\end{aligned} \tag{17.642}$$

Note that the path functional $F[W(t)]$ is still written in continuous notation, since its details are not important to this calculation. Now we will define shifted increments that move linearly towards the "target" c,

$$\Delta B_j := \Delta W_j + \frac{c}{N}, \tag{17.643}$$

which will become the increments of the Brownian bridge. Changing variables, the integral becomes

$$\begin{aligned}
&\left\langle\!\!\left\langle \delta[W(T) - c]\, F[W(t)] \right\rangle\!\!\right\rangle \\
&\qquad = \int d\Delta B_0 \ldots d\Delta B_{N-1}\, \frac{1}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\Big(\Delta B_j - \frac{c}{N}\Big)^2\right]\, \delta\left[\sum_{j=0}^{N-1}\Delta B_j\right]\, F\big[B_{T(0\to c)}(t)\big] \\
&\qquad = \int d\Delta B_0 \ldots d\Delta B_{N-1}\, \frac{e^{-c^2/2T}}{(2\pi\Delta T)^{N/2}}\, \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-1}\big(\Delta B_j\big)^2 + \frac{c}{T}\sum_{j=0}^{N-1}\Delta B_j\right] \\
&\qquad\qquad\qquad\qquad \times \delta\left[\sum_{j=0}^{N-1}\Delta B_j\right]\, F\big[B_{T(0\to c)}(t)\big].
\end{aligned} \tag{17.644}$$

Now carrying out the integration over ΔB_{N-1} will remove the delta function and enforce the replacement

$$\Delta B_{N-1} = -\sum_{j=0}^{N-2}\Delta B_j. \tag{17.645}$$

However, some care must be taken in the resulting integral. The delta-function integration formula (17.65) reads

$$\int d^d q\, \delta[h(\mathbf{q})]\, f(\mathbf{q}) = \oint_{h^{-1}(0)} \frac{f(\mathbf{q})}{|\nabla h|}\, dS, \tag{17.646}$$

where the reduced integral involves the Euclidean norm $|\nabla h|$ of the gradient of the constraint function $h(\mathbf{q})$. In Eq. (17.644), the constraint function is a simple sum over all the ΔB_j, and the derivatives are taken with respect to each ΔB_j (i.e., each of the N derivatives in the gradient has unit magnitude). The corresponding norm $|\nabla h|$ is then simply $\sqrt{N}$. Thus, the result of carrying out the ΔB_{N-1} integral in Eq. (17.644) gives

the desired result:

$$\begin{aligned}
&\left\langle\!\!\left\langle \delta[W(T)-c]\,F[W(t)] \right\rangle\!\!\right\rangle \\
&\quad= \int d\Delta B_0 \ldots d\Delta B_{N-2} \frac{e^{-c^2/2T}}{\sqrt{N}(2\pi\Delta T)^{N/2}} \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-2} \big(\Delta B_j\big)^2 - \frac{1}{2\Delta T}\left(\sum_{j=0}^{N-2}\Delta B_j\right)^2 \right] \\
&\qquad\qquad \times F\big[B_{T(0\to c)}(t)\big] \\
&\quad= \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \int d\Delta B_0 \ldots d\Delta B_{N-2} \frac{1}{(2\pi\Delta T)^{(N-1)/2}} \exp\left[-\frac{1}{2\Delta T}\sum_{j=0}^{N-2} \big(\Delta B_j\big)^2 - \frac{1}{2\Delta T}\big(B_{N-1}\big)^2 \right] \\
&\qquad\qquad \times F\big[B_{T(0\to c)}(t)\big]. \\
&\quad= \frac{e^{-c^2/2T}}{\sqrt{2\pi T}} \left\langle\!\!\left\langle F\big[B_{T(0\to c)}(t)\big] \right\rangle\!\!\right\rangle.
\end{aligned} \tag{17.647}$$

In the second-to-last expression here, there are $N-1$ steps of variance ΔT, in addition to a constraint that the variance of B_{N-1} is ΔT (B_{N-1} is close to zero rather than c because we defined the B_j to be ''drifting'' towards c; thus, $B_N = 0$ in these coordinates). Again, the coefficient on the final path average is the probability density for $W(T)$ evaluated at $W(T) = c$, but in this calculation part of this factor came from the path measure, and part came from the integration over a delta function.

17.13 Local Time

We will define the **local time** of a stochastic process $y(t)$ at displacement d as

$$\ell[y(t);d] := \int_0^t dt'\,\delta[y(t')-d]. \tag{17.648}$$
(local time)

The integrand here only ''activates'' when $y(t)$ passes through d, and the local time is a measure of how much time $y(t)$ spends at the displacement d, but normalized so that the answer is not merely zero. Recalling that we defined the sojourn time for $y(t)$ as

$$T_{\mathrm{s}}[y(t);d] := \int_0^t dt'\,\Theta[y(t')-d], \tag{17.649}$$

we can immediately deduce that

$$\ell[y(t);d] = -\partial_d T_{\mathrm{s}}[y(t);d], \tag{17.650}$$
(local time)

so that we may simply adapt our sojourn-time results to obtain local-time statistics.

One useful aspect of the local time arises in calculating functionals of stochastic processes of the form

$$\int_0^t dt'\,F[y(t')] = \int_0^t dt' \int da\,F(a)\,\delta[y(t')-a] = \int da\,F(a)\int_0^t dt'\,\delta[y(t')-a]. \tag{17.651}$$

The definition of the local time then implies

$$\int_0^t dt'\,F[y(t')] = \int_{-\infty}^{\infty} da\,F(a)\,\ell[y(t);a], \tag{17.652}$$
(local-time density formula)

so that the local time acts as an occupation density for $y(t)$. The local time is also commonly thought of as a time-dependent process, here through the time dependence of the process itself. Intuitively, the local time ''accumulates'' as the stochastic process continues, so $\ell[y(t);a]$ is a nondecreasing function of time.

As an alternate representation of the local time, recall the property of the δ function

$$\delta[f(x)] = \sum_{x_0 \in f^{-1}(0)} \frac{\delta(x - x_0)}{|f'(x_0)|}, \tag{17.653}$$

where the sum is over all real roots x_0 of $f(x)$. Then the definition (17.648) becomes

$$\ell[y(t); d] = \int_0^t dt' \sum_{\{t_d : y(t_d) = d\}} \frac{\delta(t' - t_d)}{|y'(t_d)|}, \tag{17.654}$$

or carrying out the integral,

$$\ell[y(t); d] = \sum_{\{t_d : y(t_d) = d\}} \frac{1}{|y'(t_d)|}.$$

(intersection representation of local time) (17.655)

Thus, the local time is given by summing over the intersections of the process $y(t)$ with the boundary at d, where at each intersection the contribution is the reciprocal of the ''speed'' $|y'|$ during the intersection. Intuitively, this makes sense, since the greater the speed, the less the time spent at the level d during the intersection. To get a better feeling for this, consider the case of a Wiener process, $y(t) = W(t)$, in discrete form:

$$\ell[W(t); d] = \lim_{\Delta t \to 0} \sum_{\{j : (W_{j+1} - d)(W_j - d) < 0\}} \frac{\Delta t}{|\Delta W_j|}. \tag{17.656}$$

Note that as $\Delta t \longrightarrow 0$, the contribution from each intersection in the sum decreases as $\Delta t^{1/2}$. As we will show, the local time can converge to a nonzero value; evidently, this means that the smaller step size ''reveals'' extra intersections in the neighborhood of each intersection to compensate for this decrease.

It is also interesting to consider possible generalizations of the local time. For example, consider the functional

$$\ell'[y(t); d] := \int_0^t dt'\, \delta'[y(t') - d], \tag{17.657}$$

which we can see is related to the local time via a $-\partial_d$ derivative, as the local time is related to the sojourn time. Using the composition rule (Problem 17.13)

$$\delta'[f(x)] = \sum_{x_0 \in f^{-1}(0)} \left[\frac{\delta'(x - x_0)}{f'(x_0)|f'(x_0)|} + \frac{f''(x_0)\, \delta(x - x_0)}{|f'(x_0)|^3} \right], \tag{17.658}$$

the local-time derivative becomes

$$\ell'[y(t); d] = \int_0^t dt' \sum_{\{t_d : y(t_d) = d\}} \left[\frac{\delta'(t' - t_d)}{y'(t_d)|y'(t_d)|} + \frac{y''(t_d)\, \delta(t' - t_d)}{|y'(t_d)|^3} \right]. \tag{17.659}$$

The first term under the integral vanishes so long as the velocity in the denominator does not vanish (something that occurs with zero probability, and which we also ignored in the local-time analysis), and we have

$$\ell'[y(t); d] = \sum_{\{t_d : y(t_d) = d\}} \frac{y''(t_d)}{|y'(t_d)|^3}.$$

(local-time derivative) (17.660)

For a Wiener process, $y(t) = W(t)$, in discrete time increments Δt we can count $y'(t)$ as $O(\Delta t^{-1/2})$ and $y''(t)$ as $O(\Delta t^{-3/2})$, so the summand here is of order unity. However, recall from our discussion of local-time intersections in Eq. (17.655) that the number of intersection times t_d grows as $\Delta t^{-1/2}$. Thus, while we can assign an ensemble average to this statistic,

$$\left\langle\!\!\left\langle \ell'[y(t); d] \right\rangle\!\!\right\rangle = -\partial_d \left\langle\!\!\left\langle \ell[y(t); d] \right\rangle\!\!\right\rangle = \partial_d^2 \left\langle\!\!\left\langle T_{\mathrm{s}}[y(t); d] \right\rangle\!\!\right\rangle, \tag{17.661}$$

evidently the *variance* of the derivative statistic is arbitrarily large (see Problem 17.18).

17.13.1 Wiener Process

To compute the probability distribution for the local time of the Wiener process, we will follow closely the procedure of Section 17.12.1 for the sojourn time of the Wiener process. Let $f_\ell(x)$ denote the probability density for the local time $\ell[W(t);d]$ of $W(t)$ at d, with cumulative probability $P(\ell \le x)$, satisfying

$$f_\ell(x) = \partial_x P(\ell \le x). \tag{17.662}$$

Then the Laplace transform of $f_\ell(x)$ is

$$\int_0^\infty dx\, e^{-sx}\, f_\ell(x) = \left\langle\!\!\left\langle \exp\{-s\ell[W(t);d]\} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \exp\left[-s\int_0^t dt'\,\delta[W(t')-d]\right] \right\rangle\!\!\right\rangle. \tag{17.663}$$

Note that $f_\ell(x)$ is not limited in domain to $x < t$ as was the case for the sojourn time, but the domain *is* limited to $x > 0$. Consider then the driven diffusion equation

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x)f - \lambda f + g(x), \tag{17.664}$$

where $V(x)$ is the occupation function, which here is a delta function:

$$V(x) = s\delta(x-d). \tag{17.665}$$

We will also take $g(x) = 1$. The steady-state solution is given by the Feynman–Kac formula (17.440) as

$$\begin{aligned} f(x) &= \left\langle\!\!\left\langle \int_0^\infty dt\, \exp\left(-\lambda t - \int_0^t dt'\, V[x+W(t')]\right) \right\rangle\!\!\right\rangle \\ &= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\delta[x+W(t')-d]\right) \right\rangle\!\!\right\rangle. \end{aligned} \tag{17.666}$$

This is then the solution of the steady-state version of Eq. (17.664):

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\delta(x-d)f(x) + 1. \tag{17.667}$$

For $x \neq d$, the ODE is

$$f'' = 2\lambda f - 2. \tag{17.668}$$

Setting $h = f - 1/\lambda$, we have $h'' = 2\lambda h$, so that for $x \neq d$,

$$h(x) \propto e^{\pm\sqrt{2\lambda}\,x}, \tag{17.669}$$

or choosing the bounded solutions,

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{1}{\lambda} & (x < d) \end{cases} \tag{17.670}$$

for some undetermined constants A and B. Demanding continuity of $f(x)$ at $x = d$ gives

$$B = Ae^{-2\sqrt{2\lambda}\,d}. \tag{17.671}$$

The δ function in the ODE says that the derivative $f'(x)$ should jump by $2sf(d)$ at $x = d$, so that

$$-\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d} - 2s\left(Ae^{-\sqrt{2\lambda}\,d} + \frac{1}{\lambda}\right) = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d} = \sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d}, \tag{17.672}$$

or that is, $f'(d-0^+) = f'(d+0^+) - 2sf(d+0^+)$. The solution of these two equations fixes the coefficients as

$$A = -\frac{se^{\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{17.673}$$

Now, we can equate Eqs. (17.666) and (17.670) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\left(-s\int_0^t dt'\,\delta[W(t')-d]\right)\right\rangle\!\!\right\rangle = B + \frac{1}{\lambda} = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}, \tag{17.674}$$

where we have assumed $d > 0$. Then using Eq. (17.663) on the left-hand side,

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\left\{-s\ell[W(t);d]\right\}\right\rangle\!\!\right\rangle = \frac{1}{\lambda} - \frac{se^{-\sqrt{2\lambda}\,d}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{17.675}$$

Now using the Laplace-transform formulae

$$\mathscr{L}\left[1\right](\lambda) = \frac{1}{\lambda} \tag{17.676}$$

and[63]

$$\mathscr{L}\left[-e^{ak}e^{a^2t}\operatorname{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right)+\operatorname{erfc}\left(\frac{k}{2\sqrt{t}}\right)\right](\lambda) = \frac{ae^{-k\sqrt{\lambda}}}{\lambda\left(\sqrt{\lambda}+a\right)}, \tag{17.677}$$

which becomes with $k = d\sqrt{2}$, $a = s/\sqrt{2}$,

$$\mathscr{L}\left[-e^{sd}e^{s^2t/2}\operatorname{erfc}\left(s\sqrt{t/2}+\frac{d}{\sqrt{2t}}\right)+\operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right)\right](\lambda) = \frac{se^{-d\sqrt{2\lambda}}}{\lambda\left(\sqrt{2\lambda}+s\right)}. \tag{17.678}$$

Thus, Eq. (17.675) becomes

$$\int_0^\infty dx\, e^{-sx}\, f_\ell(x) = 1 - \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) + e^{sd}e^{s^2t/2}\operatorname{erfc}\left(s\sqrt{t/2}+\frac{d}{\sqrt{2t}}\right) \tag{17.679}$$

after using Eq. (17.663) to replace the ensemble average on the left-hand side. Now using

$$\int_0^\infty dx\, e^{-sx}\frac{1}{\sqrt{2\pi t}}e^{-(x+d)^2/2t} = \frac{1}{2}\,e^{sd}\,e^{s^2t/2}\operatorname{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \tag{17.680}$$

we can invert the Laplace transforms on both sides, with the result[64]

$$f_\ell(x) = \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right)\delta(x-0^+) + \sqrt{\frac{2}{\pi t}}\,e^{-(x+|d|)^2/2t}, \tag{17.681}$$

(local-time probability density)

which is normalized in view of

$$\sqrt{\frac{2}{\pi t}}\,e^{-(x+d)^2/2t} = \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) = 1 - \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{17.682}$$

[63] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1027, Eq. (29.3.89).

[64] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.4; A. N. Borodin, "Brownian local time," *Russian Mathematical Surveys* **44**, 1 (1989), p. 5 (doi: 10.1070/RM1989v044n02ABEH002050); Jim Pitman, "The distribution of local times of a Brownian bridge," in *Séminaire de Probabilités XXXIII*, Jacques Azéma, Michel Émery, Michel Ledoux, and Marc Yor, Eds. (Springer, 1999), p. 388, Eq. (1) (doi: 10.1007/BFb0096528).

Again, the δ-function term here gives the boundary-noncrossing probability, with the crossing probability (17.372) appearing, which is the same as the probability to have zero local time at d:

$$P(\ell = 0) = 1 - P_{\text{cross}}(d,t) = 1 - \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) = \operatorname{erf}\left(\frac{d}{\sqrt{2t}}\right). \tag{17.683}$$

We have also inserted absolute-value symbols for d: while we assumed $d > 0$ for this derivation, the result should be exactly symmetric in d owing to the same symmetry in $W(t)$. The cumulative probability is again given by integrating 0 to x:[65]

$$P(\ell \le x) = \int_0^x dx'\, f_\ell(x') = \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right) + \left[\operatorname{erf}\left(\frac{x+|d|}{\sqrt{2t}}\right) - \operatorname{erf}\left(\frac{|d|}{\sqrt{2t}}\right)\right], \tag{17.684}$$

or

$$P(\ell \le x) = \operatorname{erf}\left(\frac{x+|d|}{\sqrt{2t}}\right) \tag{17.685}$$

(local-time cumulative density)

after cancelling terms.

17.13.2 Standard Brownian Bridge

The local-time distributions for the standard Brownian bridge $B(t)$ are special cases of the results in the following section, so we just quote them here.

$$f_\ell(x) = \left[1 - e^{-2|d|^2}\right] \delta(x - 0^+) + \big(x + 2|d|\big)\, e^{-(x+2|d|)^2/2} \tag{17.686}$$

(local-time probability density for $B(t)$)

is the probability density for the standard bridge, while

$$P(\ell \le x) = 1 - e^{-(x+2|d|)^2/2} \tag{17.687}$$

(local-time cumulative density for $B(t)$)

is the cumulative density.

Note that as in Eq. (17.585), it is not difficult to compute moments of the local time using

$$\begin{aligned} \left\langle\!\left\langle \ell^n \right\rangle\!\right\rangle &= n \int_0^\infty dx\, x^{n-1} [1 - P(\ell \le x)] \\ &= n \int_0^\infty dx\, x^{n-1} e^{-(x+2|d|)^2/2}. \end{aligned} \tag{17.688}$$

Thus, for example,

$$\left\langle\!\left\langle \ell \right\rangle\!\right\rangle = \sqrt{\frac{\pi}{2}}\, \operatorname{erfc}\left(\sqrt{2}\,|d|\right) \tag{17.689}$$

for the mean local time of a standard Brownian bridge.

17.13.3 Brownian Bridge

For the local time of the Brownian bridge $B_{0\to c}(t)$ from 0 to c as $t = 0$ to 1, we will merge the procedures of Sections 17.12.2, 17.12.3, and 17.13.1. Here $f_\ell(x)$ for $0 \le x \le \infty$ will denote the probability density for the local time $\ell[B_{0\to c}(t); d]$ of $B_{0\to c}(t)$ at d, with cumulative probability $P(\ell \le x)$. The procedure is the same as in Section 17.13.1 up to Eq. (17.666), where we will leave $g(x)$ in the Feynman–Kac formula:

$$f(x) = \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\left\langle g[x + W(t)] \exp\left(-s \int_0^t dt'\, \delta[x + W(t') - d]\right) \right\rangle\!\right\rangle. \tag{17.690}$$

[65] cf. Lajos Takács, "On the Local Time of the Brownian Motion," *The Annals of Applied Probability* **5**, 741 (1995), Eq. (3) (doi: 10.1214/aoap/1177004703).

In this case, we will take $g(x) = e^{ik(x-c)}$, as we did in Section 17.12.3. This is then the solution of the steady-state diffusion equation

$$\lambda f(x) = \frac{1}{2}\partial_x^2 f(x) - s\delta(x-d)f(x) + e^{ik(x-c)}. \tag{17.691}$$

For $x \neq d$, the ODE is

$$f'' = 2\lambda f - 2e^{ik(x-c)}. \tag{17.692}$$

Then setting $h = f - 2e^{ik(x-c)}/(2\lambda + k^2)$, we have

$$\begin{aligned} h'' &= f'' + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda f - 2e^{ik(x-c)} + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda h + \frac{4\lambda e^{ik(x-c)}}{2\lambda + k^2} - 2e^{ik(x-c)} + \frac{2k^2 e^{ik(x-c)}}{2\lambda + k^2} \\ &= 2\lambda h, \end{aligned} \tag{17.693}$$

so that for $x \neq d$,

$$h(x) \propto e^{\pm\sqrt{2\lambda}\,x}, \tag{17.694}$$

or picking the bounded solutions in each domain,

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{17.695}$$

for undetermined constants A and B. Demanding continuity of $f(x)$ at $x = d$ again gives

$$B = Ae^{-2\sqrt{2\lambda}\,d}. \tag{17.696}$$

The δ function in the ODE causes the derivative $f'(x)$ to jump by $2sf(d)$ at $x = d$, so that

$$-\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d} - 2s\left(Ae^{-\sqrt{2\lambda}\,d} + \frac{e^{ik(d-c)}}{\lambda + k^2/2}\right) = \sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d} = \sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d}, \tag{17.697}$$

or that is, $f'(d-0^+) = f'(d+0^+) - 2sf(d)$. The solution of these two equations fixes the coefficients as

$$A = -\frac{se^{\sqrt{2\lambda}\,d}\,e^{ik(d-c)}}{(\lambda + k^2/2)\left(\sqrt{2\lambda} + s\right)}, \qquad B = -\frac{se^{-\sqrt{2\lambda}\,d}\,e^{ik(d-c)}}{(\lambda + k^2/2)\left(\sqrt{2\lambda} + s\right)}. \tag{17.698}$$

Now we can equate Eqs. (17.666) and (17.695) and set $x = 0$ to obtain

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle e^{ik[W(t)-c]} \exp\left(-s\int_0^t dt'\,\delta[W(t') - d]\right)\right\rangle\!\!\right\rangle = B + \frac{e^{-ikc}}{\lambda + k^2/2}, \tag{17.699}$$

where we have again assumed $d > 0$ [but note that if we assume $d < 0$, the following treatment is the same, but with $d \longrightarrow -d$, so we will simply replace d by $|d|$ in the factor $\exp(-\sqrt{2\lambda}\,d)$]. Integrating with respect to k and dividing through by 2π introduces the δ function that pins the Wiener path to c at time t:

$$\int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle \delta[W(t) - c] \exp\left(-s\int_0^t dt'\,\delta[W(t') - d]\right)\right\rangle\!\!\right\rangle = \frac{1}{2\pi}\int_{-\infty}^\infty dk\,\left(B + \frac{e^{-ikc}}{\lambda + k^2/2}\right). \tag{17.700}$$

Using Eq. (17.592) to change the Wiener-path average into a bridge average,

$$\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle\,\exp\left(-s\int_0^t dt'\,\delta[B_{t(0\to c)}(t')-d]\right)\right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\,\left(B+\frac{e^{-ikc}}{\lambda+k^2/2}\right). \quad (17.701)$$

Then using Eq. (17.663) on the left-hand side and carrying out the k integration,

$$\begin{aligned}\int_0^\infty dt\,\frac{e^{-c^2/2t}}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle\,\exp\left\{-s\ell[B_{t(0\to c)};d]\right\}\right\rangle\!\!\right\rangle &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty dk\left(\frac{e^{-ikc}}{\lambda+k^2/2}-\frac{s\,e^{-\sqrt{2\lambda}|d|}\,e^{ik(d-c)}}{(\lambda+k^2/2)\left(\sqrt{2\lambda}+s\right)}\right)\\ &= \sqrt{\frac{\pi}{\lambda}}\,e^{-\sqrt{2\lambda}|c|}-\frac{\sqrt{\pi}\,s\,e^{-\sqrt{2\lambda}\,(|d|+|c-d|)}}{\sqrt{\lambda}\left(\sqrt{2\lambda}+s\right)}.\end{aligned} \quad (17.702)$$

Now using the Laplace-transform formulae[66]

$$\mathscr{L}\left[\frac{1}{\sqrt{\pi t}}e^{-k^2/4t}\right](\lambda) = \frac{1}{\sqrt{\lambda}}e^{-k\sqrt{\lambda}}, \quad (17.703)$$

and[67]

$$\mathscr{L}\left[e^{ak}e^{a^2t}\,\mathrm{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right)\right](\lambda) = \frac{e^{-k\sqrt{\lambda}}}{\sqrt{\lambda}\left(\sqrt{\lambda}+a\right)}, \quad (17.704)$$

which becomes with $k=(|d|+|c-d|)\sqrt{2}$, $a=s/\sqrt{2}$,

$$\mathscr{L}\left[e^{s(d+|c-d|)}e^{s^2t/2}\,\mathrm{erfc}\left(s\sqrt{t/2}+\frac{(d+|c-d|)}{\sqrt{2t}}\right)\right](\lambda) = \frac{e^{-(d+|c-d|)\sqrt{2\lambda}}}{\sqrt{\lambda}\left(\sqrt{\lambda}+s/\sqrt{2}\right)} = \frac{\sqrt{2}\,e^{-(d+|c-d|)\sqrt{2\lambda}}}{\sqrt{\lambda}\left(\sqrt{2\lambda}+s\right)}. \quad (17.705)$$

Thus, Eq. (17.701) becomes

$$\int_0^\infty dx\,e^{-sx}\,f_\ell(x) = 1-\sqrt{\frac{\pi t}{2}}\,s\,e^{c^2/2t}e^{s(|d|+|c-d|)}e^{s^2t/2}\,\mathrm{erfc}\left(s\sqrt{\frac{t}{2}}+\frac{(|d|+|c-d|)}{\sqrt{2t}}\right) \quad (17.706)$$

after using Eq. (17.663) to replace the ensemble average on the left-hand side. Now again using

$$\int_0^\infty dx\,e^{-sx}\frac{1}{\sqrt{2\pi t}}e^{-(x+d)^2/2t} = \frac{1}{2}\,e^{sd}\,e^{s^2t/2}\,\mathrm{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \quad (17.707)$$

with the derivative rule

$$s\mathscr{L}[f(x)](s) = \mathscr{L}[f'(x)](s)+f(0), \quad (17.708)$$

so that

$$-\int_0^\infty dx\,e^{-sx}\frac{(x+d)}{\sqrt{2\pi t^3}}e^{-(x+d)^2/2t}+\frac{1}{\sqrt{2\pi t}}e^{-d^2/2t} = \frac{s}{2}\,e^{sd}\,e^{s^2t/2}\,\mathrm{erfc}\left[\sqrt{\frac{t}{2}}\left(s+\frac{d}{t}\right)\right], \quad (17.709)$$

we can invert the Laplace transforms on both sides, with the result[68]

$$f_\ell(x) = \left[1-e^{[c^2-(|d|+|c-d|)^2]/2t}\right]\delta(x-0^+)+\frac{1}{t}\left(x+|d|+|c-d|\right)e^{[c^2-(x+|d|+|c-d|)^2]/2t}, \quad (17.710)$$

(local-time probability density for $B_{t(0\to c)}(t')$)

[66] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1026, Eq. (29.3.84).

[67] Milton Abramowitz and Irene A. Stegun, *op. cit.*, p. 1027, Eq. (29.3.90).

[68] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.8; A. N. Borodin, "Brownian local time," *Russian Mathematical Surveys* **44**, 1 (1989), p. 6 (doi: 10.1070/RM1989v044n02ABEH002050).

The δ-function term here gives the boundary-noncrossing probability in the case of $c < d$, where the exponential part reduces to $\exp[-2d(d-c)]$, in agreement with Eq. (17.387).

The corresponding cumulative probability is

$$P(\ell \le x) = \int_0^x dx'\, f_\ell(x') = 1 - e^{[c^2-(|d|+|c-d|)^2]/2t} - \Big[e^{[c^2-(x+|d|+|c-d|)^2]/2t}\Big]_0^x, \tag{17.711}$$

or

$$P(\ell \le x) = 1 - e^{[c^2-(x+|d|+|c-d|)^2]/2t}$$

(local-time cumulative density for $B_{t(0\to c)}$) (17.712)

after cancelling terms.

17.13.3.1 Moments

Note that as in Eqs. (17.585) and (17.688), we can compute moments of the local time via the cumulative as probability function

$$\begin{aligned}\Big\langle\!\Big\langle \ell^n \Big\rangle\!\Big\rangle &= n\int_0^\infty dx\, x^{n-1}[1 - P(\ell \le x)] \\ &= n\int_0^\infty dx\, x^{n-1}\, e^{[c^2-(x+|d|+|c-d|)^2]/2t}\end{aligned} \tag{17.713}$$

for $n > 0$. For example, the mean is

$$\big\langle\!\big\langle \ell \big\rangle\!\big\rangle = \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t}\, \operatorname{erfc}\!\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{17.714}$$

Note that in the case where the endpoints of the bridge straddle the interval (i.e., $0 \le d \le c$ or $c \le d \le 0$) the argument $|d| + |c-d|$ reduces to $|c|$, in which case the mean local time is independent of the boundary location, which seems peculiar.[69] Since the probability (17.712) and density (17.710) depend on c and d in exactly the same way, these probabilities are also invariant to shifts of the interface, provided the shift keeps the interface d between 0 and c.[70]

17.13.3.2 Moment-Generating Function

Since the moments are fairly straightforward to calculate, it shouldn't come as much surprise that the moment-generating function is easy to obtain. Writing this out,

$$\begin{aligned}\Big\langle\!\Big\langle e^{-s\ell} \Big\rangle\!\Big\rangle &= \int_0^\infty dx\, e^{-sx}\, f_\ell(x) \\ &= s\int_0^\infty dx\, e^{-sx} P(\ell \le x) \\ &= 1 - s\int_0^\infty dx\, e^{-sx}\big[1 - P(\ell \le x)\big] \\ &= 1 - s\int_0^\infty dx\, e^{-sx} e^{[c^2-(x+|d|+|c-d|)^2]/2t},\end{aligned} \tag{17.715}$$

where we used Eq. (17.712) for the cumulative density. Note that in the first step, we discarded the boundary terms from integrating by parts; the term at $x = 0$ vanishes if we regard the delta function in Eq. (17.712) to

[69] Zhiyi Chi, Vladimir Pozdnyakov, and Jun Yan, "On expected occupation time of Brownian bridge," *Statistics and Probability Letters* **97**, 83 (2015) (doi: 10.1016/j.spl.2014.11.009).

[70] Peter Howard and Kevin Zumbrun, "Shift invariance of the occupation time of the Brownian bridge process," *Statistics and Probability Letters* **45**, 379 (1999) (doi: 10.1016/S0167-7152(99)00080-2).

be displaced *above* $x = 0$ until the end of the calculation (which is a shortcut for treating the delta-function component separately). Carrying out the final integral, we find[71]

$$\left\langle\!\left\langle e^{-s\ell} \right\rangle\!\right\rangle = 1 - s\sqrt{\frac{\pi t}{2}}\, e^{c^2/2t + s^2 t/2 + s(|d| + |c-d|)} \operatorname{erfc}\left(\frac{|d| + |c-d| + st}{\sqrt{2t}}\right)$$

(local-time moment-generating function for $B_{t(0\to c)}$) (17.716)

for the moment-generating function.

17.13.3.3 Application to the Sojourn Time

The local-time mean can also be used to obtain an expression for the mean of the sojourn time for a general Brownian bridge. By taking the expectation value of Eq. (17.650), we can simply write the mean sojourn time as the integral

$$\left\langle\!\left\langle T_s[y(t); d] \right\rangle\!\right\rangle = \int_d^\infty dx \left\langle\!\left\langle \ell[y(t); x] \right\rangle\!\right\rangle. \tag{17.717}$$

To handle this integral, we can use the integral formula

$$\int dx \operatorname{erfc}(ax - b) = \frac{b}{a} - \frac{1}{a\sqrt{\pi}} + \left(x - \frac{b}{a}\right) \operatorname{erfc}(ax - b), \tag{17.718}$$

and handle the integral in several cases. First, if $d \geq 0$ and $d \geq c$, then we have from Eq. (17.714)

$$\begin{aligned} \left\langle\!\left\langle T_s[y(t); d] \right\rangle\!\right\rangle &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} \int_d^\infty dx \operatorname{erfc}\left(\frac{2x - c}{\sqrt{2t}}\right) \\ &= \frac{t}{2} e^{-2d(d-c)/t} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} (2d - c) \operatorname{erfc}\left(\frac{2d - c}{\sqrt{2t}}\right). \end{aligned} \tag{17.719}$$

In the case $d \geq 0$ and $c \geq d$, we can divide the integrand into the parts with $x \leq c$ and $x \geq c$, the latter of which we just did:

$$\begin{aligned} \left\langle\!\left\langle T_s[y(t); d] \right\rangle\!\right\rangle &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} \int_d^c dx \operatorname{erfc}\left(\frac{|x|}{\sqrt{2t}}\right) + \left\langle\!\left\langle T_s[y(t); c] \right\rangle\!\right\rangle \\ &= \sqrt{\frac{\pi t}{2}}\, e^{c^2/2t} (c - d) \operatorname{erfc}\left(\frac{c}{\sqrt{2t}}\right) + \frac{t}{2} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} c \operatorname{erfc}\left(\frac{c}{\sqrt{2t}}\right) \\ &= \frac{t}{2} - \sqrt{\frac{\pi t}{8}}\, e^{c^2/2t} (2d - c) \operatorname{erfc}\left(\frac{c}{\sqrt{2t}}\right). \end{aligned} \tag{17.720}$$

Combining Eqs. (17.719) and (17.720) into a single case valid for $d \geq 0$,

$$\left\langle\!\left\langle T_s[y(t); d] \right\rangle\!\right\rangle = \frac{t}{2}\, e^{-2[d(d-c)\Theta(d)\Theta(d-c)]/t} - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{17.721}$$

For the case $d \leq 0$, we can change the signs of both c and d to obtain the mirror image, and replace the sojourn time by the sojourn time subtracted from t. The net result for $d \leq 0$ is

$$\left\langle\!\left\langle T_s[y(t); -d] \right\rangle\!\right\rangle = t - \frac{t}{2}\, e^{-2[d(d-c)\Theta(-d)\Theta(c-d)]/t} - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \tag{17.722}$$

Equations (17.721) and (17.722) can then be combined into a single case as

$$\begin{aligned} \left\langle\!\left\langle T_s[y(t); d] \right\rangle\!\right\rangle = \frac{t}{2} + \operatorname{sgn}(d) \frac{t}{2} \Big[& e^{-2[d(d-c)\Theta(d)\Theta(d-c) + d(d-c)\Theta(-d)\Theta(c-d)]/t} - 1 \Big] \\ & - \sqrt{\frac{\pi t}{8}}\, (2d - c)\, e^{c^2/2t} \operatorname{erfc}\left(\frac{|d| + |c-d|}{\sqrt{2t}}\right). \end{aligned} \tag{17.723}$$

[71] cf. Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 155, formula 1.3.7.

Actually, since the second term vanishes as long as $0 \le d \le c$, the $\operatorname{sgn}(d)$ can be shifted, with the result

$$\begin{aligned}\left\langle\!\left\langle T_{\mathrm{s}}[y(t);d]\right\rangle\!\right\rangle = \frac{t}{2} + \operatorname{sgn}(2d-c)\,\frac{t}{2}\left[e^{-2[d(d-c)\Theta(d)\Theta(d-c)+d(d-c)\Theta(-d)\Theta(c-d)]/t} - 1\right] \\ - \sqrt{\frac{\pi t}{8}}\,(2d-c)\,e^{c^2/2t}\operatorname{erfc}\left(\frac{|d|+|c-d|}{\sqrt{2t}}\right),\end{aligned} \tag{17.724}$$

which makes the expression symmetric about $c/2$.

17.13.4 Local Time and Discontinuities in Stochastic Processes

Because of the relation of the local time to the delta function, the local time tends to show up in the theory of stochastic processes especially when discontinuities are involved. We will consider two examples of this here: the absolute-value process, and a diffusion process with a discontinuity in the diffusion rate.

17.13.4.1 Reflected Brownian Motion

Consider the absolute-value diffusion process $|W(t)| = \operatorname{sgn}[W(t)]\,W(t)$. This is an example of **reflected Brownian motion**, in the sense that this is Brownian motion on $(0,\infty)$, but when the process encounters the origin and attempts to cross it, it is ''reflected'' back into the positive real axis.

The Itō differential for this process is, expanding to second order using the Itō chain rule (17.193),

$$\begin{aligned} d|W(t)| &= \frac{d|W|}{dW}dW + \frac{1}{2}\frac{d^2|W|}{dW^2}dW^2 \\ &= \operatorname{sgn}(W)\,dW + \delta(W)\,dt,\end{aligned} \tag{17.725}$$

where we have used

$$\frac{d|x|}{dx} = \operatorname{sgn}(x), \qquad \frac{d|x|^2}{dx^2} = 2\delta(x), \tag{17.726}$$

thinking of the signum function as twice the Heaviside function plus a constant offset. Then integrating this transformed SDE from 0 to t,

$$|W(t)| = \int_0^t \operatorname{sgn}[W(t')]\,dW(t') + \ell[W(t);0], \tag{17.727}$$
(Tanaka formula)

where we used the local-time definition (17.648). This is the **Tanaka formula**.[72] This can also be written as

$$|W(t)-d| - |d| = \int_0^t \operatorname{sgn}[W(t')-d]\,dW(t') + \ell[W(t);d], \tag{17.728}$$
(Tanaka formula)

if $W(t)$ is shifted by d at the start of the derivation. This says that in Itō calculus, the reflecting effect of the absolute value is reflected in a deterministic ''pushing'' term that activates whenever the process hits the origin.

17.13.4.2 Discontinuous Diffusion Rate

Consider the Itō-form SDE

$$dy = \beta(y)\,dW, \tag{17.729}$$

representing state-dependent diffusion without an explicit drift term. We want to consider the case where $\beta(y)$ has a discontinuity (or a number of isolated discontinuities). As a particular model, consider

$$\beta(y) = \begin{cases} \beta_>, & y > 0 \\ \beta_<, & y < 0, \end{cases} \tag{17.730}$$

[72] Andrei N. Borodin and Paavo Salminen, *op. cit.*, p. 43.

with the point $\beta(0)$ being, say, the average of the left- and right-hand values. As we argued before in Section 17.3.5.1, the probability density of particles governed by this SDE evolves by the Fokker–Planck equation

$$\partial_t P(y,t) = \frac{1}{2}\partial_y^2 \beta^2(y) P(y,t), \tag{17.731}$$

so that the diffusion coefficient $D(y) = \beta^2(y)$ has a discontinuity at the origin. Making the Lamperti transform (Section 17.7.4.2), we construct the function

$$z = S(y) = \int_0^y \frac{dy'}{\beta(y')} = \begin{cases} y/\beta_>, & y > 0 \\ -y/\beta_<, & y < 0 \end{cases} \tag{17.732}$$

(which is invertible provided $\beta_{\gtrless}$ have the same sign), we can transform our SDE to the form of Eq. (17.359),

$$dz = -\frac{1}{2}\beta'(y)\,dt + dW. \tag{17.733}$$

Evaluating the derivative of β as the derivative of a step function,

$$dz = -\frac{1}{2}(\beta_> - \beta_<)\,\delta(y)\,dt + dW. \tag{17.734}$$

Now we have

$$\delta(z) = \delta[S(y)] = \frac{\delta(y)}{|S'(y)|} = \frac{2\delta(y)}{\beta_> + \beta_<}, \tag{17.735}$$

where we think of the delta function as a limit of centered distributions, so that the derivative of $S(y)$ is the average of the derivatives on either side of the discontinuity. Thus, we have the fully transformed SDE

$$dz = \left(\frac{\beta_< - \beta_>}{\beta_< + \beta_>}\right)\delta(z)\,dt + dW. \tag{17.736}$$

In this form, where we have scaled away the variance, the discontinuity arises as a deterministic ''kick'' term that acts only at the discontinuity's location. Using the local-time definition (17.648), we can integrate this to find

$$z(t) - z_0 = W(t) + \left(\frac{\beta_< - \beta_>}{\beta_< + \beta_>}\right)\ell[W(t);0], \tag{17.737}$$

(skew Brownian motion)

so that $z(t)$ has the form of a regular Wiener path, with a shift that depends on the time the particle spent at the discontinuity (and thus being kicked by the local gradient). The motion of $z(t)$ is called **skew Brownian motion**.

17.13.4.3 Skew Brownian Motion

Skew Brownian motion[73] is equivalent to ordinary Brownian motion, except that the probability is skewed towards one direction at the origin. Suppose we take a random walk of steps taken every Δt, and probability density

$$f(x) = p\,\delta\left(x - \sqrt{\Delta t}\right) + (1-p)\,\delta\left(x + \sqrt{\Delta t}\right), \tag{17.738}$$

so that the probability of a step of size $x = +\sqrt{\Delta t}$ is p, and the probability is $1-p$ of stepping in the opposite direction. The second moment of each step is

$$\int_{-\infty}^{\infty} dx\, x^2 f(x) = \Delta t, \tag{17.739}$$

[73] Skew Brownian motion was introduced by Kiyosi Itô and Henry P. McKean, Jr., *Diffusion Processes and their Sample Paths* (Springer-Verlag, 1974). See also J. M. Harrison and L. A. Shepp, ''On Skew Brownian Motion,'' *The Annals of Probability* **9**, 309 (1981) (doi: 10.1214/aop/1176994472); and Antoine Lejay, ''On the constructions of the skew Brownian motion,'' *Probability Surveys* **3** 413 (2006) (doi: 10.1214/154957807000000013).

so that the unbiased case $p = 1/2$ leads to the usual Wiener path $W(t)$ as $\Delta t \longrightarrow 0$. The more general case gives a bias of

$$\int_{-\infty}^{\infty} dx\, x f(x) = p\sqrt{\Delta t} - (1-p)\sqrt{\Delta t} = (2p-1)\sqrt{\Delta t} \tag{17.740}$$

in each step (and thus a step variance of $\Delta t - (2p-1)^2 \Delta t = 4p(1-p)\Delta t \leq \Delta t$ for $p \in [0,1]$). If we then consider an inhomogeneous walk where we take a biased stepping probability p in the region $x \in [-\sqrt{\Delta t}/2, \sqrt{\Delta t}/2)$, and an unbiased probability $p = 1/2$ outside this region, then we have the discrete-step equation

$$\Delta x = \Delta W + \text{rect}\left(\frac{x}{\sqrt{\Delta t}}\right)(2p-1)\sqrt{\Delta t}, \tag{17.741}$$

where $\text{rect}(t)$ is a square pulse of unit height and unit width, centered at $t = 0$, and we have replaced the stochastic component of the random walk by an equivalent Gaussian process in view of the limit $\Delta t \longrightarrow 0$. Note that this process has the correct mean

$$\langle\!\langle \Delta x \rangle\!\rangle = \mathbb{1}_{[-\sqrt{\Delta t}/2, \sqrt{\Delta t}/2)}(x), \tag{17.742}$$

where $\mathbb{1}_A(x)$ is the indicator function for $x \in A$, and the process has the correct second moment,

$$\langle\!\langle \Delta x^2 \rangle\!\rangle = \langle\!\langle \Delta W^2 \rangle\!\rangle = \Delta t, \tag{17.743}$$

provided we ignore terms of order $\Delta W\, \Delta t$ and Δt^2. Then as $\Delta t \longrightarrow 0$, we can also note that

$$\frac{1}{\sqrt{\Delta t}}\,\text{rect}\left(\frac{x}{\sqrt{\Delta t}}\right) \longrightarrow \delta(x), \tag{17.744}$$

so that in this limit we obtain the SDE

$$dx = dW + (2p-1)\,\delta(x)\,dt, \tag{17.745}$$
(skew Brownian motion)

with solution

$$x(t) - x_0 = W(t) + (2p-1)\,\ell[W(t); 0]. \tag{17.746}$$
(skew Brownian motion)

These equations are equivalent to Eqs. (17.736) and (17.737) if we identify

$$p = \frac{\beta_<}{\beta_< + \beta_>}, \qquad 1 - p = \frac{\beta_>}{\beta_< + \beta_>}. \tag{17.747}$$

As an example at the extreme end of this range of probabilities, note that the reflected Brownian motion in Eq. (17.725) has a similar form to skew Brownian motion with $p = 1$ (if we modify the process by absorbing the sgn into dW, which is possible due to the symmetry of dW). In this limit, the process is always reflected upward at the origin. In terms of the inhomogeneous diffusion, this corresponds to the limit where $\beta_< \gg \beta_>$, though the diffusion picture does not necessarily show the same repulsion from the $y < 0$ region; this is more an artifact of the transformation (17.732), which "smooshes" the $y < 0$ region up against the origin if $\beta <$ is large.

17.13.4.4 Reflections, Images, and Transition Densities for Skew Brownian Motion

The original construction of skew Brownian motion[74] began with reflected Brownian motion, and then introduced probabilities p and $1-p$ of reflecting $|W(t)|$ upwards or downwards, respectively, on each encounter with the origin.

In this sense,

$$r = 2p - 1 = \frac{\beta_< - \beta_>}{\beta_< + \beta_>} \tag{17.748}$$
("reflection coefficient" for diffusion)

[74] Kiyosi Itô and Henry P. McKean, Jr., *op. cit.*

acts as a ''reflection coefficient'' from above the interface for the Wiener path. The reflection occurs always up or down for $p = 1$ or 0, respectively (corresponding to $r = +1$ or -1), and has no reflection (bias) for $p = 1/2$ ($r = 0$). Indeed, identifying the $\beta_{\lessgtr}$ with the inverse ''refractive indices'' $1/n_{\lessgtr}$ on either side of a refractive interface, the reflection coefficient r has the form of the Fresnel coefficient for the electric field at normal incidence from above a dielectric interface [and $t = 1 + r = 2\beta_</(\beta_> + \beta_<)$ has the form of the Fresnel transmission coefficient from above].

Indeed, the reflection principle can give the transition density for the trajectories,[75] making the connection to a reflection coefficient more explicit. To compute the transition densities for skew Brownian motion, we will use a procedure similar to the calculation for boundary crossings of Brownian bridges in Section 17.8.2. To begin, let's compute the probability density for a skew-Brownian-motion process $z(t)$ starting at $x_0 > 0$ and ending at $x > 0$. Let τ_0 denote the first-crossing time through the boundary $x = 0$. Then we start with

$$P[z(t) = x] = P[z(t) = x \wedge \tau_0 \le t] + P[z(t) = x \wedge \tau_0 > t]. \tag{17.749}$$

That is, the path either hits the origin or it doesn't before time t. We can compute the first term using the Reflection Principle, similar to the calculation in Section 17.8:

$$\begin{aligned} P[z(t) = x \wedge \tau_0 \le t] &= P[z(t) = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= P[z(t) = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x \,|\, \tau_0 \le t]\, P[\tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x \wedge \tau_0 \le t] \\ &= 2pP[W(t) + x_0 = -x] \\ &= 2p\phi(x + x_0)\, dx, \end{aligned} \tag{17.750}$$

where $\phi(x)$ is the transition density for $W(t)$ from 0,

$$\phi(x) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/2t}. \tag{17.751}$$

In changing from skew Brownian to regular Brownian motion, we used the ratio of probabilities for taking an upward step at the point of crossing $z(\tau) = 0$:

$$\frac{P[z(0) = 0 \wedge z(t) = x]}{P[W(t) = x]} = \frac{p}{1/2} = 2p. \tag{17.752}$$

(This ratio is unity for the unbiased case $p = 1/2$.) We then used the Reflection Principle for $W(t)$, given that it crossed $W(\tau) + z_0 = 0$.

We can evaluate second term in Eq. (17.749) using the **method of images**, which works in the same way as the method of images for potentials due to charges in the presence of boundary conditions. Here, we have $P[z(t) = x \wedge \tau_0 > t]$, which is the same as $P[z(t) = x]$, but subject to the boundary condition $P[z(t) = x = 0] = 0$, since no crossings happen here. We can achieve this by considering the density

[75] J. B. Walsh, ''A diffusion with a discontinuous local time,'' *Société Mathématique de France Astérisque* **52-53**, 37 (1978); J. F. Le Gall, ''One-dimensional stochastic differential equations involving the local times of the unknown process,'' in *Stochastic Analysis and Applications: Proceedings of the International Conference held in Swansea, April 11-15, 1983*, A. Truman and D. Williams, Eds. (Springer-Verlag, 1984), p. 51; G. J. M. Uffink, ''A random walk method for the simulation of macrodispersion in a stratified aquifer,'' in *Relation of Groundwater Quantity and Quality (Proceedings of the Hamburg Symposium, August 1983)* (IAHS Publ. no. 146, 1985), p. 103; Antoine Lejay, *op. cit.*, around Eq. (42); Thilanka Appuhamillage, Vrushali Bokil, Enrique Thomann, Edward Waymire, and Brian Wood, ''Occupation and local times for skew Brownian motion with applications to dispersion across an interface,'' *Annals of Applied Probability* **21**, 183 (2011), Eq. (2.2) (doi: 10.1214/10-AAP691); Eric M. LaBolle and Graham E. Fogg, ''Reply [to ''Comment on 'Diffusion theory for transport in porous media: Transition-probability densities of diffusion processes corresponding to advection-dispersion equations' by Eric M. LaBolle et al.''],'' *Water Resources Research* **36**, 823 (2012) (doi: 10.1029/1999WR900325).

diffusing from a distribution initially localized at x_0, balanced by a distribution of *negative* density at $-x_0$, such that perfect cancellation occurs at $x = 0$:

$$P[z(t) = x \wedge \tau_0 > t] = \phi(x - x_0)\,dx - \phi(x + x_0)\,dx. \tag{17.753}$$

Putting this all together in Eq. (17.749), we then have

$$P[z(t) = x] = \phi(x - x_0) + (2p - 1)\,\phi(x + x_0)\,dx, \tag{17.754}$$

again for $x > 0$ and $x_0 > 0$.

For $x_0 > 0$ and $x < 0$, the second term in Eq. (17.749) corresponds to an impossible event, and by a similar Reflection-Principle argument, the second term is $2(1 - p)\,\phi(x + x_0)$:

$$P[z(t) = x] = 2(1 - p)\,\phi(x + x_0)\,dx. \tag{17.755}$$

Collecting both cases together, we have for $z(0) = z_0$, the probability density

$$P[z(t)] = \begin{cases} \phi(z - z_0)\,dz + r\phi(z + z_0)\,dz, & z > 0, z_0 > 0 \\ (1 - r)\,\phi(z + z_0)\,dz, & z < 0, z_0 > 0, \end{cases}$$

(transition density for skew Brownian motion) (17.756)

with $\phi(x)$ defined in Eq. (17.751), and the reflection coefficient r given by Eq. (17.748). The same formula applies for $z_0 < 0$ with minor changes,

$$P[z(t)] = \begin{cases} \phi(z - z_0)\,dz - r\phi(z + z_0)\,dz, & z < 0, z_0 < 0 \\ (1 + r)\,\phi(z + z_0)\,dz, & z > 0, z_0 < 0, \end{cases}$$

(transition density for skew Brownian motion) (17.757)

where both r and z change sign.

If we translate these results back into the language of the discontinuous-diffusion SDE (17.729), we can invert the Lamperti transform (17.732) to obtain

$$y = S^{-1}(z) = \begin{cases} z\beta_>, & z > 0 \\ -z\beta_<, & z < 0. \end{cases} \tag{17.758}$$

Then transforming the probability density, Eq. (17.756) becomes

$$P[y(t)] = \begin{cases} \phi\left(\dfrac{y - y_0}{\beta_>}\right)\dfrac{dy}{\beta_>} + r\phi(x + x_0)\,\dfrac{dy}{\beta_>}, & y > 0, y_0 > 0 \\ (1 - r)\,\phi\left(\dfrac{y - (\beta_</\beta_>)y_0}{\beta_<}\right)\dfrac{dy}{\beta_<}, & y < 0, y_0 > 0, \end{cases}$$

(transition density for discontinuous diffusion) (17.759)

where note that in the $y < 0$ case, the y_0 was determined by x_0 in the Reflection-Principle argument for $x > 0$, and should correspond to the same distance from the origin as $y_0/\beta_>$ to properly match the boundary condition at $y = 0$ (that is, ''equivalent'' trajectories from the effective sources on either side will hit $y = 0$ at the same time only with this rescaling). Again, the $y_0 < 0$ case can be obtained by exchanging $\beta_{\gtrless}$ everywhere, including in the reflection coefficient r.

17.13.4.5 Stratonovich Discontinuous Diffusion

We began the discussion of a discontinuous diffusion rate with the Itō-form diffusion (17.729), and the choice of Itō calculus was important in generating the ''pumping'' effect at the discontinuity. If we instead consider a Stratonovich-form diffusion SDE

$$dy = \beta(y) \circ dW, \tag{17.760}$$

corresponding to the Itō SDE

$$dy = \frac{1}{2}\beta'(y)\beta(y)\,dt + \beta(y)\,dW, \tag{17.761}$$

then things are different; in fact, the Lamperti-transformed version of this SDE is simply

$$dz = dW, \tag{17.762}$$

as we discussed in Section 17.7.4.2. The equivalent Fokker–Planck equation from Eq. (17.130) is

$$\partial_t P(y,t) = -\frac{1}{2}\partial_y \beta'(y)\beta(y)P(y,t) + \frac{1}{2}\partial_y^2 \beta^2(y)P(y,t), \tag{17.763}$$

or commuting a derivative,

$$\partial_t P(y,t) = \frac{1}{2}\partial_y \beta(y)\partial_y \beta(y)P(y,t).$$

(equivalent Fokker–Planck equation) (17.764)

Note that by changing variables to z according to

$$\frac{dz}{dy} = \frac{1}{\beta}, \tag{17.765}$$

which is the same as the transformation (17.732), we can rewrite the Fokker–Planck equation as a standard diffusion equation,

$$\partial_t \rho = \frac{1}{2}\partial_z^2 \rho, \tag{17.766}$$

(equivalent Fokker–Planck equation)

where

$$\rho(y,t) = \beta(y)P(y,t). \tag{17.767}$$

This simplification is the same as the result $dz = dW$ from the Lamperti transformation of the Stratonovich-diffusion SDE.

17.14 Bessel Processes

A **Bessel process** in d dimensions is simply the ''radius'' or Euclidian norm of a vector Wiener process in d dimensions:

$$R_d := \left\|\mathbf{W}(t)\right\|_d = \sqrt{\sum_{\alpha=1}^{d} [W_\alpha(t)]^2}\,. \tag{17.768}$$

(Bessel process)

As we will see, this process is useful for characterizing the motion of a vector Wiener process, particularly in the sense of whether or not it returns to its starting point. Also, note that this is the multidimensional generalization of the reflecting Wiener process of Section 17.13.4.1, to which this reduces for $d = 1$.

To begin, we can try to work out an SDE for R_d. First, writing out the differential for R_d,

$$dR_d = \frac{1}{2R_d}\sum_{\alpha=1}^{d}\left(2W_\alpha\, dW_\alpha + dW_\alpha^2\right) = \frac{d}{2R_d}\,dt + \frac{1}{R_d}\sum_{\alpha=1}^{d} W_\alpha\, dW_\alpha. \tag{17.769}$$

Unfortunately, it is difficult to proceed directly from this point. However, we can also work more simply with the *square* of the Bessel process, and write out the differential:

$$dR_d^2 = \sum_{\alpha=1}^{d}\left(2W_\alpha\, dW_\alpha + dW_\alpha^2\right) = d\,dt + 2\sum_{\alpha=1}^{d} W_\alpha\, dW_\alpha. \tag{17.770}$$

On the other hand, we can write out the differential as

$$dR_d^2 = 2R_d\, dR_d + (dR_d)^2. \tag{17.771}$$

From Eq. (17.769), we can find

$$(dR_{\mathsf{d}})^2 = \frac{1}{R_{\mathsf{d}}^2} \sum_{\alpha=1}^{\mathsf{d}} W_\alpha^2 \, (dW_\alpha)^2 = dt, \tag{17.772}$$

where we used $dW_\alpha \, dW_\beta = 0$ if $\alpha \neq \beta$ for independent Wiener processes. Then using this result and (17.770) in Eq. (17.771), we find

$$dR_{\mathsf{d}} = \frac{\mathsf{d}-1}{2R_{\mathsf{d}}} \, dt + \frac{1}{R_{\mathsf{d}}} \sum_{\alpha=1}^{\mathsf{d}} W_\alpha \, dW_\alpha. \tag{17.773}$$

What remains is to analyze the last term, which turns out to be equivalent to a Wiener process. Defining

$$d\tilde{W}(t) := \frac{1}{R_{\mathsf{d}}} \sum_{\alpha=1}^{\mathsf{d}} W_\alpha \, dW_\alpha, \tag{17.774}$$

it is easy to see that $\langle\!\langle d\tilde{W} \rangle\!\rangle = 0$ (since $\langle\!\langle f(W) \, dW \rangle\!\rangle = 0$ in Itō calculus), and also that

$$\langle\!\langle d\tilde{W}^2 \rangle\!\rangle = \frac{1}{R_{\mathsf{d}}^2} \sum_{\alpha=1}^{\mathsf{d}} W_\alpha^2 \, (dW_\alpha)^2 = dt. \tag{17.775}$$

The higher-order moments of $d\tilde{W}$ vanish also, so we can see that $\tilde{W}$ is another representation of a Wiener process. Then dropping the tilde, we have the SDE

$$dR_{\mathsf{d}} = \frac{\mathsf{d}-1}{2R_{\mathsf{d}}} \, dt + dW \tag{17.776}$$

(Bessel-process SDE)

to directly generate the dynamics of the Bessel process. Note that because of the divergence at $R_{\mathsf{d}} = 0$, we should be careful to specify that the transformation to this equation is only valid provided $R_{\mathsf{d}} \neq 0$. That is, if the Bessel process hits the origin, then this equation becomes invalid afterwards. [For example, in $\mathsf{d} = 1$, we are missing the delta-function term that arises in Eq. (17.725).] We will soon return to the question of whether or not this actually happens.

17.14.1 Generator

The term "Bessel process" comes specifically for the generator. Recall that an SDE of the form $dx = \alpha \, dt + \beta \, dW$ implies a diffusion-type equation, which can be summarized by the generator [see Eq. (17.136)]

$$\mathcal{G} := \alpha(x,t) \, \partial_x + \frac{\beta^2(x,t)}{2} \, \partial_x^2. \tag{17.777}$$

For the Bessel-process SDE (17.776), the generator is

$$\mathcal{G} = \frac{\mathsf{d}-1}{2x} \partial_x + \frac{1}{2} \partial_x^2, \tag{17.778}$$

and the generator-eigenvalue equation $\mathcal{G}u(x) = \gamma u(x)$ has the form

$$\frac{\mathsf{d}-1}{2x} u' + \frac{1}{2} u'' = \gamma u. \tag{17.779}$$

[Note that this can be viewed as a Laplace transform of the backward Kolmogorov equation (17.144).] The idea is that this differential equation is equivalent to a Bessel equation, as we will now see.

First, changing variables according to

$$v = x^m u, \tag{17.780}$$

so that $v' = mx^{m-1}u + x^m u'$ and $v'' = m(m-1)x^{m-2}u + 2mx^{m-1}u' + x^m u''$, we will need to solve these for the derivatives of u to give $u' = x^{-m}v' - mx^{-1}u = x^{-m}v' - mx^{-m-1}v$ and $u'' = x^{-m}v'' - m(m-1)x^{-2}u - 2mx^{-1}u' = x^{-m}v'' - m(m-1)x^{-m-2}v - 2mx^{-m-1}v' + 2m^2x^{-m-2}v$. Thus, Eq. (17.779) becomes

$$\frac{1}{2}x^{-m}v'' + \frac{\mathsf{d}-2m-1}{2}x^{-m-1}v' - \frac{m(\mathsf{d}-m-2)}{2}x^{-m-2}v = \gamma x^{-m}v. \tag{17.781}$$

Then multiplying through by $2x^{m+2}$,

$$x^2v'' + (\mathsf{d}-2m-1)xv' - \big[2\gamma x^2 + m(\mathsf{d}-m-2)\big]v = 0. \tag{17.782}$$

Now we can scale out the eigenvalue by letting $\tilde{x} := \sqrt{2\gamma}\,x$, and choosing m such that $(\mathsf{d}-2m-1) = 1$, leading to $m = (\mathsf{d}-2)/2$ and $m(\mathsf{d}-m-2) = [(\mathsf{d}-2)/2]^2$, we have

$$\tilde{x}^2v'' + \tilde{x}v' - \big[\tilde{x}^2 + m^2\big]v = 0. \tag{17.783}$$

This is the modified Bessel equation, and the solutions are the modified Bessel functions $I_m(\tilde{x})$ and $K_m(\tilde{x})$. Note that for small z, $I_m(z)\tilde{z}^m = z^{(d/2)-1}$, while $K_m(z)\tilde{z}^{-m} = z^{1-(d/2)}$. Also, for large z, $I_m(z) \longrightarrow \infty$ as $e^z/\sqrt{z}$, while $K_m(z) \longrightarrow 0$. So the $K_m(z)$ represent physical solutions in this problem.

17.14.2 Brownian Recurrence

One useful question we can address with Bessel processes is, given a Wiener path in d dimensions starting at the origin and running for arbitrarily large times, how far away does it go, and does it return to the origin?

17.14.2.1 One Dimension

In $\mathsf{d} = 1$, we have already answered this question by analyzing boundary-crossing probabilities. That is, from Eq. (17.372) we know that as $t \longrightarrow \infty$, the probability for crossing a boundary at distance d away converges to unity for any value of d. This means that the path wanders arbitrarily far away from the origin, and once it has wandered any given distance away, it is guaranteed to return to the origin. Thus Brownian motion in $\mathsf{d} = 1$ dimension is said to be **recurrent**.

17.14.2.2 Two Dimensions

The $\mathsf{d} = 2$ case turns out to be *marginally* recurrent, as we will see. We already know that R_2 will wander arbitrarily far from the origin, because the projection onto one dimension already does this, as we know from the $\mathsf{d} = 1$ case. However, the question of returning to the origin is somewhat different.

To analyze this case, it is convenient to define

$$L := \log R_2. \tag{17.784}$$

Then the SDE (17.776) for this case,

$$dR_2 = \frac{1}{2R_2}\,dt + dW, \tag{17.785}$$

will simplify as follows. Expanding the logarithm to second order gives

$$dL = \frac{dR_2}{R_2} - \frac{(dR_2)^2}{2R_2^2}, \tag{17.786}$$

which with Eq. (17.785) becomes

$$dL = \frac{1}{2R_2^2}\,dt + \frac{1}{R_2}dW - \frac{dW^2}{2R_2^2} = \frac{dW}{R_2}. \tag{17.787}$$

Then since $R_2 = e^L$,

$$dL = e^{-L}\,dW. \tag{17.788}$$

This transformation simplifies things by removing the drift, at the expense of multiplicative noise. However, we will use the martingale property of this noise, which means that the ensemble average vanishes:

$$d\langle\langle L\rangle\rangle = 0. \tag{17.789}$$

This means that $\langle\langle L\rangle\rangle = \langle\langle \log R_2\rangle\rangle$ is a constant of the motion. Since we know that L will take on arbitrarily large values, this essentially means that L will also have to take on arbitrarily negative values to keep the mean constant. This means that R_2 will come arbitrarily close to 0.

To make this argument more formal,[76] consider two positive radii $R_<$ and $R_>$ that bound the initial value R_0 of R_2: $R_< \leq R_0 \leq R_>$. Since $\langle\langle L\rangle\rangle$ is constant,

$$\log R_0 = \Big\langle\Big\langle \log R_2(t)\Big\rangle\Big\rangle. \tag{17.790}$$

The reasoning that led to Eq. (17.433) applies in a similar way here. Let τ be the first time at which $R_2(t)$ achieves either $R_<$ or $R_>$. Then we can consider a paths up to its stopping time τ, at which point L takes the value $\log R_2(\tau)$; when averaging over all continuations of the path up to time $t > \tau$, the value $\log R(\tau)$ is unchanged. Now extending the average over all paths up to the stopping time, we see that it is equivalent to write Eq. (17.790) in terms of the hitting time as

$$\log R_0 = \Big\langle\Big\langle \log R_2(\tau)\Big\rangle\Big\rangle \tag{17.791}$$

after taking the limit $t \longrightarrow \infty$. Now writing the expectation value in terms of probabilities for first hitting $R_<$ or $R_>$,

$$\log R_0 = (\log R_<)\, P[R_2(\tau) = R_<] + (\log R_>)\, P[R_2(\tau) = R_>]. \tag{17.792}$$

Solving for $P[R(\tau) = R_>]$ gives

$$P[R_2(\tau) = R_>] = \frac{\log R_0 - (\log R_<)\, P[R_2(\tau) = R_<]}{\log R_>}. \tag{17.793}$$

Now fixing R_0 and $R_< \leq R_0$, we can let $R_> \longrightarrow \infty$ to obtain

$$P[R_2(\tau) = \infty] = 0. \tag{17.794}$$

That is, the probability for $R_2(t)$ to "hit infinity before hitting $R_<$" (which means *never* hitting $R_<$) is zero. Thus, for any inner bound $R_> > 0$, $R_2(t)$ is guaranteed to hit it in finite time. However, the same argument with fixed $R_>$ and $R_< \longrightarrow 0$ gives that $R_2(t)$ will never hit the origin in finite time. Thus, so to speak, the origin is nonrecurrent, but any disc surrounding the origin is recurrent.

17.14.2.3 Three and More Dimensions

In three and more dimensions, it turns out that Brownian motion is nonrecurrent, and in fact the path wanders "far away" in the sense that $R_d(t) \longrightarrow \infty$. It is sufficient to show this in three dimensions, again because any projection of a d-dimensional motion onto three dimensions will satisfy these conditions. However, it isn't much more difficult to treat any $d \geq 3$ directly.

In this case, the simplifying transformation for the SDE (17.776) has the form

$$P := aR_d^b, \tag{17.795}$$

for constants a and b to be determined. Computing the differential of P (and expanding to second order),

$$dP = abR_d^{b-1}\, dR_d + \frac{ab(b-1)}{2} R_d^{b-2}\, (dR_d)^2, \tag{17.796}$$

[76] See Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, 1988), pp. 161-3 (doi: 10.1007/978-1-4612-0949-2).

and then using Eq. (17.776) for dR_d,

$$dP = \frac{ab(b+d-2)}{2} R_d^{b-2}\, dt + abR_d^{b-1}\, dW. \tag{17.797}$$

We can then force the drift term to vanish by setting

$$b = 2 - d, \tag{17.798}$$

in which case

$$dP = a(2-d)R_d^{1-d}\, dW. \tag{17.799}$$

Now since $P = aR_d^{2-d}$, we have $R_d = (P/a)^{1/(2-d)}$, and thus

$$dP = a^{1-c}(2-d)P^c\, dW, \tag{17.800}$$

where

$$c := \frac{1-d}{2-d} = \frac{d-1}{d-2}. \tag{17.801}$$

Finally, if we choose

$$a = (d-2)^{1/(1-c)}, \tag{17.802}$$

we end up with the SDE

$$dP = -P^c\, dW, \tag{17.803}$$

which again has the martingale property $\langle\langle dP \rangle\rangle = 0$. In what follows, it is useful to remember that for $d > 2$, P goes as an *inverse* power of R_d.

As in the previous section, we may take $R_d(0) = R_0$ to be the initial condition, and define inner and outer boundaries satisfying $R_< \le R_0 \le R_>$, with τ the first passage time through either boundary. Then since

$$P_0 = \big\langle\big\langle P(\tau) \big\rangle\big\rangle, \tag{17.804}$$

this becomes

$$R_0^{2-d} = \Big\langle\Big\langle R_d^{2-d}(\tau) \Big\rangle\Big\rangle \tag{17.805}$$

directly in terms of the radial coordinate. We can rewrite the expectation value as before in terms of the probabilities for first crossing the inner and outer boundaries as

$$R_0^{2-d} = R_<^{2-d} P[R_d(\tau) = R_<] + R_>^{2-d} P[R_d(\tau) = R_>]. \tag{17.806}$$

Now if $R_> \longrightarrow \infty$, the last term vanishes, and

$$P[R_d(\tau) = R_<] = \left(\frac{R_<}{R_0}\right)^{d-2}.$$

(probability for achieving inner bound $R_<$) (17.807)

This gives the probability for the path, starting from $R_0 > 0$, to hit $R_< \le R_0$. Note that this probability converges to zero as $R_< \longrightarrow 0$. Thus, the probability for a path starting at $R_0 > 0$ to hit the origin is zero (and thus the origin is nonrecurrent, since once a path moves away from the origin, it will not return). Each path starting at R_0 has a (pathwise) minimum radius, and the probability density for this minimum is given by differentiating (17.807), with the result

$$f(r) = \frac{(d-2)\, r^{d-3}}{R_0^{d-2}}.$$

(probability density for minimum radius) (17.808)

Now to show that $R_d \longrightarrow \infty$ for $d \ge 3$, first fix some upper limit $R_>$. We want to show that after some time $t_>$, $R_d(t) > R_>$ for all $t > t_>$ with probability 1. To demonstrate this, pick a larger boundary $R_{>>} > R_>$. We know that R_d will hit $R_{>>}$ in finite time $t_>$. From Eq. (17.807), the probability that the path crosses below $R_>$ is $(R_>/R_{>>})^{d-2}$. But this probability converges to zero as $R_{>>} \longrightarrow \infty$, in which case we are guaranteed that the path does not return below $R_>$.

17.15 Differentiation of Stochastic Quantities

We have so far studied stochastic processes and ensemble averages over stochastic problems. In cases where we compute quantities like the sojourn time or the local time, analytic expressions are available for the ensemble average, generating function, and so on. But for more complicated statistics, analytic expressions may not be available, and numerical simulations are needed to compute such averages.

Another important class of ensemble-average problems that we have not yet discussed is the *differentiation* of ensemble-averaged statistics. Since Wiener processes are nondifferentiable, the differentation of numeric quantities with underlying Wiener processes can be tricky. Recall, for example, that the sojourn time can be differentiated to yield the local time [see Eq. (17.650)]; this differentiation then carries over to the mean sojourn and local times, but differentiation of the underlying Wiener process causes larger fluctuations in the ensemble-average computation on a pathwise basis [see Eq. (17.656), due to the variation in the number of intersections and the possibility of small values of ΔW_j]. In a further differentiation of the local-time process, the derivative of the ensemble average is well-defined, but the pathwise derivative has arbitrarily large fluctuations in the continuum limit [see Eq. (17.660) and the following discussion].

Thus, it is useful to consider some general approaches to handling derivatives of stochastic processes, which can be handled efficiently in certain cases with a variety of tricks. As a byproduct, we will also briefly study the derivative of a stochastic process with respect to a stochastic quantity, which gives the generalizes variational calculus to stochastic processes. Such derivatives are useful in, for example, computing Casimir forces from energy path integrals (Chapter 21). These techniques are also widely employed in financial mathematics, where ensemble averages over stochastic trajectories are used to price financial derivatives (e.g., financial options). In such pricing, derivatives of the price with respect to parameters such as the volatility or starting price yields the ''sensitivity'' of the price with respect to parameter variations. These techniques also apply in optimization problems, where the ''payoff'' to optimize is estimated via Monte-Carlo simulations.

17.15.1 Parameter Differentiation and Likelihood Estimation

Consider an ensemble average of the form

$$A(\lambda) := \left\langle\!\!\left\langle \Phi(\mathbf{Z}) \right\rangle\!\!\right\rangle_{\mathbf{Z}}, \tag{17.809}$$
(model ensemble average)

where $\mathbf{Z} = (Z_1, \ldots, Z_n)$ is a set of stochastic variables, $\Phi(\mathbf{Z})$ is some scalar ''payoff'' function of the stochastic quantities that we wish to average, and the λ on the left-hand side indicates a parameter dependence of the ensemble average. In some important cases we can arrange things such that the only dependence on the right-hand side on the parameter λ is in the probability distribution for $\mathbf{Z}$. That is, we can write the ensemble average as

$$A(\lambda) = \int d\mathbf{z}\, \Phi(\mathbf{z})\, f_\lambda(\mathbf{z}), \tag{17.810}$$

where $f_\lambda(\mathbf{z})$ is the probability density for $\mathbf{Z}$, and the λ subscript emphasizes the dependence on λ. This approach can work even in some cases where it seems that the payoff function should represent the parameter dependence. For example, suppose we differentiate the sojourn time of Wiener paths with respect to the boundary distance d, which is a part of the functional rather than the paths. However, this quantity can be computed equivalently by differentiating with respect to the starting point of the paths.

Now the derivative can be written simply as

$$\partial_\lambda A(\lambda) = \int d\mathbf{z}\, \Phi(\mathbf{z})\, \partial_\lambda f_\lambda(\mathbf{z}), \tag{17.811}$$

and by multiplying and dividing the integrand by $f_\lambda(\mathbf{z})$, we can rewrite this as an ensemble average:

$$\partial_\lambda A(\lambda) = \left\langle\!\!\left\langle \Phi(\mathbf{Z})\, \frac{\partial_\lambda f_\lambda(\mathbf{Z})}{f_\lambda(\mathbf{Z})} \right\rangle\!\!\right\rangle_{\mathbf{Z}} = \left\langle\!\!\left\langle \Phi(\mathbf{Z})\, \partial_\lambda \log f_\lambda(\mathbf{Z}) \right\rangle\!\!\right\rangle_{\mathbf{Z}}.$$
(likelihood-ratio derivative estimator) (17.812)

Of course, this generalizes readily to higher derivatives,

$$\partial_\lambda^m A(\lambda) = \left\langle\!\!\left\langle \Phi(\mathbf{Z}) \frac{\partial_\lambda^m f_\lambda(\mathbf{Z})}{f_\lambda(\mathbf{Z})} \right\rangle\!\!\right\rangle_{\mathbf{Z}},$$

(likelihood ratio mth derivative estimator) (17.813)

although of course without the logarithmic form. Thus, the derivative here appears simply as a reweighted version of the original ensemble average (17.809). Of course, ''good'' behavior of this ensemble average is not guaranteed, but the hope is that the logarithmic weight will not cause much in the way of fluctuations. In particular, this approach should give an advantage when the probability density is smooth, whereas the payoff function is not (e.g., it may have a discontinuity or singularity as in the sojourn or local time, such that a small change in the boundary location d can produce a large change in the payoff value). One particularly nice property of this expression is that the weight is universal for *any* payoff function $\Phi(\mathbf{Z})$, because the parameter dependence lies entirely with the probability distribution.

17.15.1.1 Likelihood

Now for a brief digression to explain some terminology. The **likelihood** of a parameter λ, given some observed outcome $\mathbf{Z}$, is defined as the probability of the outcome given the particular parameter value:

$$\mathcal{L}(\lambda|\mathbf{Z}) := P(\mathbf{Z}|\lambda). \tag{17.814}$$

In our ensemble average, the likelihood is just the probability density $f_\lambda(\mathbf{Z})$. Then the ratio of two likelihoods $f_{\lambda'}(\mathbf{Z})/f_\lambda(\mathbf{Z})$ is a common quantity in statistics to evaluate the relative plausibility of two models given observed data. Then the likelihood-ratio derivative $\partial_{\lambda'}[f_{\lambda'}(\mathbf{Z})/f_\lambda(\mathbf{Z})]$ appears in the ensemble average (17.812), so a Monte-Carlo evaluation of Eq. (17.812) is known as a **likelihood-ratio estimator** for the derivative.[77]

17.15.1.2 Application: Differentiation of Brownian-Bridge Path Averages

When computing statistics related to sojourn or local times, or also boundary-crossing and escape probabilities, typically path-averages functional of the form

$$A(x_0) := \left\langle\!\!\left\langle \Phi[V(x)] \right\rangle\!\!\right\rangle_{x(t)}, \tag{17.815}$$

(model ensemble average)

where $x(t)$ are stochastic paths over an evolution time t. In this form, for example, $V(x) = \Theta(x - d)$ for the sojourn time, and $\Phi(x) = e^{-sx}$ computes a moment-generating function or $\Phi(x) = x$ computes a simple average. As noted before, derivatives with respect to a boundary position d can be regarded as derivatives with respect to the initial point x_0 of the path. Thus, we will specifically consider Brownian-bridge paths

$$x(t') = x_0 + B_t(t'), \tag{17.816}$$

such that x_0 represents both the source point, $B_t(0) = x_0$, and the path terminus, $B_t(t) = x_0$. In this case, if we consider the time-sliced path in N steps of duration $\Delta t = t/N$, the probability density of the paths in the average (17.815) have the x_0-dependent factor

$$f[x(t)] \propto e^{-(x_1 - x_0)^2/2\Delta t}\, e^{-(x_0 - x_{N-1})^2/2\Delta t}, \tag{17.817}$$

[77] Martin I. Reiman and Alan Weiss, ''Sensitivity analysis via likelihood ratios,'' *Proceedings of the 1986 Winter Simulation Conference*, J. Wilson, J. Henriksen, and S. Roberts, Eds., p. 285 (1986) (doi: 10.1145/318242.318450). P. W. Glynn, ''Stochastic approximation for Monte Carlo optimization,'' *Proceedings of the 1986 Winter Simulation Conference*, J. Wilson, J. Henriksen, and S. Roberts, Eds., p. 356 (1986) (doi: 10.1145/318242.318459). Reuven Y. Rubinstein, ''Sensitivity Analysis and Performance Extrapolation for Computer Simulation Models,'' *Operations Research*, **37**, 72 (1989) (doi: 10.1287/opre.37.1.72). P. W. Glynn, ''Likelihood Ratio Derivative Estimators for Stochastic Systems,'' *Proceedings of the 1989 Winter Simulation Conference*, E. A. MacNair, K. J. Musselman, and P. Heidelberger, Eds., p. 374 (1989) (doi: 10.1109/WSC.1989.718702).

where recall that by convention we are identifying $x_N = x_0$. Then the derivative (17.812) becomes

$$\frac{\partial_{x_0} f[x(t)]}{f[x(t)]} = \frac{(x_1 - x_0)}{\Delta t} - \frac{(x_0 - x_{N-1})}{\Delta t} = \frac{2(\bar{x}_1 - x_0)}{\Delta t}, \tag{17.818}$$

where we are defining the shorthand

$$\bar{x}_1 := \frac{x_1 + x_{N-1}}{2}. \tag{17.819}$$

Thus, the first derivative of the functional (17.815) becomes

$$\partial_{x_0} A(x_0) = \left\langle\!\!\left\langle \Phi[V(x)] \, \frac{2(\bar{x}_1 - x_0)}{\Delta t} \right\rangle\!\!\right\rangle_{x(t)}.$$

(differentiated ensemble average) (17.820)

Note that $(\bar{x}_1 - x_0) = (\Delta x_1 - \Delta x_{N-1})/2 = O(\Delta t^{1/2})$, so overall the average (17.820) includes an overall weight of magnitude $O(\Delta t^{-1/2}) = O(N^{1/2})$ compared to the undifferentiated expression (17.815).

It is also straightforward to derive an expression for high-order derivatives. First, rewriting Eq. (17.817) as

$$f[x(t)] \propto e^{-(x_0 - \bar{x}_1)^2/\Delta t} \, e^{-(x_{N-1} - x_1)^2/4\Delta t} \tag{17.821}$$

by completing the square, we can define $y := (x_0 - \bar{x}_1)/\sqrt{\Delta t}$, and use the Hermite-polynomial definition

$$H_m(y) := (-1)^m e^{y^2} \partial_y^m e^{-y^2} \tag{17.822}$$

we can write out the derivative

$$\frac{\partial_{x_0}^m f[x(t)]}{f[x(t)]} = (\Delta t)^{-m/2} \, \frac{\partial_y^m f(y)}{f(y)} = (-1)^m \, (\Delta t)^{-m/2} \, H_m\!\left(\frac{x_0 - \bar{x}_1}{\sqrt{\Delta t}}\right) = (\Delta t)^{-m/2} \, H_m\!\left(\frac{\bar{x}_1 - x_0}{\sqrt{\Delta t}}\right). \tag{17.823}$$

Then the derivative of the functional (17.815) becomes

$$\partial_{x_0}^m A(x_0) = (\Delta t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)] \, H_m\!\left(\frac{\bar{x}_1 - x_0}{\sqrt{\Delta t}}\right) \right\rangle\!\!\right\rangle_{x(t)}.$$

(differentiated ensemble average) (17.824)

Of course, since $H_1(y) = 2y$, this expression reduces to Eq. (17.820) for $m = 1$. But remarkably, high-order derivatives may be computed merely by reweighting via x_1 and x_{N-1}. Of course, each derivative supplies an additional factor $\Delta t^{-1/2} \propto N^{1/2}$, making the fluctuations larger for each successively higher derivative. Note that the expression (17.824) also applies to ordinary Wiener processes if Δt is replaced by $2\Delta t$, and $\bar{x}_1$ is replaced by x_1.

Although in principle the expression (17.824) applies to arbitrarily high derivatives, it may only practical for relatively small m, depending on the nature of the solution to the path average. The factor of $(\Delta t)^{-m/2}$ is something we can put aside for the moment, since this problem can at least be tamed. Recall that the harmonic-oscillator eigenfunctions have the (normalized) form

$$\psi_m(x) = \frac{1}{\sqrt{2^m m! \sqrt{\pi}}} \, H_m(x) \, e^{-x^2/2}, \tag{17.825}$$

which grow in width as $\sqrt{m}$. Roughly, this means that the weighting factor in Eq. (17.824) favors increasingly large mean steps $|\bar{x}_1|$ with increasing m. This width can match the original probability density quite poorly for large m, in which case it could be advantageous to absorb the Hermite polynomial into the probability measure for the paths—this avoids rare events in the Gaussian tails from making large contributions. However, doing so introduces a normalization factor that grows exponentially with m. It is only in cases where the ensemble-average derivative grows similarly that the relative accuracy in a Monte-Carlo calculation does not suffer (an average that depends on x_0^{-s} for $s > 0$ is one case where this is okay).

17.15.1.3 Variance Reduction

Now to deal with the factor of $(\Delta t)^{-m/2}$ that causes Eq. (17.824) to become numerically inefficient for larger derivatives. Again, this is because, generally speaking, simulations of path integrals and path averages are accurate in the limit $N \longrightarrow \infty$, but the fluctuations of the path average grow as $N^{m/2}$. This obviously complicates accurate simulation of the path averages difficult.

The key idea is that under certain conditions, the ''payoff'' functional $\Phi[V(x)]$ is approximately independent of the path samples near the beginning $(x_1, x_2, \ldots)$ and end $(x_{N-1}, x_{N-2}, \ldots)$ of the path. This can happen, for example, if $\Phi[V(x)]$ represents the sojourn-time functional, where the path at times $t' \ll d^2$ (and $t - t' \ll d^2$, where t is the total running time of the bridge) has a very small probability of crossing the boundary. These parts of the path thus contributes very little to the mean sojourn time. Under these conditions, we can successively integrate over the early coordinates $(x_1, x_2, \ldots)$ and the late coordinates $(x_{N-1}, x_{N-2}, \ldots)$ to perform a ''partial average'' over paths. More specifically, we can hold x_0 and x_2 fixed, while averaging over all possible values of x_1. Then while holding x_3 fixed, average over all possible values of x_2, and so on until the error associated with the averaging is no longer negligible.

To arrive at the appropriate, partially averaged version of Eq. (17.824), consider the integral

$$I := \int dx_j \, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) \frac{e^{-(x_0 - x_{N-k})^2/2k\Delta t}}{\sqrt{2\pi k\Delta t}} \frac{e^{-(x_j - x_0)^2/2j\Delta t}}{\sqrt{2\pi j\Delta t}} \frac{e^{-(x_{j+1} - x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}, \tag{17.826}$$

which consists of the Gaussian and Hermite-polynomial weights for a step of size $j\Delta t$ at the beginning of the path (from x_0 to x_j), and of size $k\Delta t$ (from x_{N-k} to $x_N = x_0$) at the end of the path; a Gaussian weight for the Gaussian step of size Δt from x_j to x_{j+1}; and the partial-averaging integral over x_j. The mean displacement of the first and last steps here is

$$\bar{x}_{jk} := \frac{kx_j + jx_{N-k}}{j+k}. \tag{17.827}$$

Then completing the squares on the first product of Gaussian factors and then the second product, we can rewrite this as

$$\begin{aligned}
I &= \int dx_j \, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) \frac{e^{-(x_0 - \bar{x}_{jk})^2(j+k)/2jk\Delta t}}{\sqrt{2\pi jk\Delta t/(j+k)}} \frac{e^{-(x_{N-k} - x_j)^2/2(j+k)\Delta t}}{\sqrt{2\pi(j+k)\Delta t}} \frac{e^{-(x_{j+1} - x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}} \\
&= \int dx_j \, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) \frac{e^{-(x_0 - \bar{x}_{jk})^2(j+k)/2jk\Delta t}}{\sqrt{2\pi jk\Delta t/(j+k)}} \frac{e^{-(x_j - \check{x}_{jk})^2(j+k+1)/2(j+k)\Delta t}}{\sqrt{2\pi(j+k)\Delta t/(j+k+1)}} \\
&\qquad \times \frac{e^{-(x_{N-k} - x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{2\pi(j+k+1)\Delta t}} \\
&= \frac{e^{-(x_{N-k} - x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{8\pi^3 jk^3\Delta t^3/(j+k)^2}} \int d\bar{x}_{jk} \, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) e^{-(x_0 - \bar{x}_{jk})^2(j+k)/2jk\Delta t} \\
&\qquad \times e^{-(x_j - \check{x}_{jk})^2(j+k+1)/2(j+k)\Delta t} \\
&= \frac{e^{-(x_{N-k} - x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{8\pi^3 jk^3\Delta t^3/(j+k)^2}} \int d\bar{x}_{jk} \, H_m\!\left(\frac{x_0 - \bar{x}_{jk}}{\sqrt{2jk\Delta t/(j+k)}}\right) e^{-(x_0 - \bar{x}_{jk})^2(j+k)/2jk\Delta t} \\
&\qquad \times e^{-(\bar{x}_{jk} - \bar{x}_{(j+1)k})^2(j+k)(j+k+1)/2k^2\Delta t},
\end{aligned} \tag{17.828}$$

where we have defined the offset

$$\check{x}_{jk} := \frac{(j+k)x_{j+1} + x_{N-k}}{j+k+1}, \tag{17.829}$$

and we used

$$\begin{aligned} x_j - \check{x}_{jk} &= \frac{(j+k)\bar{x}_{jk} - jx_{N-k}}{k} - \frac{(j+k)x_{j+1} + x_{N-k}}{j+k+1} \\ &= \frac{(j+k)}{k}\left(\bar{x}_{jk} - \frac{(j+1)}{(j+k+1)}x_{N-k} - \frac{kx_{j+1}}{j+k+1}\right) \\ &= \frac{(j+k)}{k}\left(\bar{x}_{jk} - \bar{x}_{(j+1)k}\right), \end{aligned} \tag{17.830}$$

where now

$$\bar{x}_{(j+1)k} := \frac{kx_{j+1} + (j+1)x_{N-k}}{j+1+k}. \tag{17.831}$$

Then using the convolution formula[78]

$$e^{-x^2/\alpha^2} * \left[H_n\left(\frac{x}{\beta}\right)e^{-x^2/\beta^2}\right] = \frac{\sqrt{\pi}\,\alpha\beta^{n+1}}{(\alpha^2+\beta^2)^{(n+1)/2}}\,H_n\left(\frac{x}{\sqrt{\alpha^2+\beta^2}}\right)e^{-x^2/(\alpha^2+\beta^2)}, \tag{17.832}$$

with

$$\alpha = \sqrt{\frac{2k^2\Delta t}{(j+k)(j+k+1)}}, \qquad \beta = \sqrt{\frac{2jk\Delta t}{(j+k)}} \tag{17.833}$$

such that

$$\alpha^2 + \beta^2 = \frac{2k(j+1)}{(j+k+1)}\Delta t, \qquad \frac{\beta}{\sqrt{\alpha^2+\beta^2}} = \sqrt{\frac{j(j+k+1)}{(j+1)(j+k)}}, \tag{17.834}$$

we find

$$\begin{aligned} I = \frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{4\pi^2(j+1)k\Delta t^2}}&\left[\frac{j(j+k+1)}{(j+1)(j+k)}\right]^{m/2} \\ &\times H_m\left(\frac{x_0 - \bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right)e^{-(x_0-\bar{x}_{(j+1)k})^2(j+1+k)/2(j+1)k\Delta t}. \end{aligned} \tag{17.835}$$

Rearranging to make the recursion more clear, we have

$$\begin{aligned} \left[\frac{j+k}{jk}\right]^{m/2} I &= \left[\frac{j+k+1}{(j+1)k}\right]^{m/2} H_m\left(\frac{x_0 - \bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right) \\ &\qquad \times \frac{e^{-(x_0-\bar{x}_{(j+1)k})^2(j+1+k)/2(j+1)k\Delta t}}{\sqrt{2\pi(j+1)k\Delta t/(j+1+k)}}\,\frac{e^{-(x_{N-k}-x_{j+1})^2/2(j+k+1)\Delta t}}{\sqrt{2\pi(j+1+k)\Delta t}} \\ &= \left[\frac{j+k+1}{(j+1)k}\right]^{m/2} H_m\left(\frac{x_0 - \bar{x}_{(j+1)k}}{\sqrt{2(j+1)k\Delta t/(j+1+k)}}\right) \\ &\qquad \times \frac{e^{-(x_0-x_{N-k})^2/2k\Delta t}}{\sqrt{2\pi k\Delta t}}\,\frac{e^{-(x_{j+1}-x_0)^2/2(j+1)\Delta t}}{\sqrt{2\pi(j+1)\Delta t}} \end{aligned} \tag{17.836}$$

where we are basically ''unravelling'' the first square that we completed in Eq. (17.828). The same argument may be made for integrating over x_{N-k} by the time-symmetry of the Brownian bridge. Thus the expression that satisfies this recursion and reduces to Eq. (17.824) for $j = k = 1$ is

$$\begin{aligned} \partial_{x_0}^m A(x_0) &\approx \left[\frac{j+k}{2jk\Delta t}\right]^{m/2} \left\langle\!\!\left\langle \Phi[V(x)]\,H_m\left(\frac{\bar{x}_{jk} - x_0}{\sqrt{2jk\Delta t/(j+k)}}\right)\right\rangle\!\!\right\rangle_{x(t)} \\ &= \left[\frac{t_1+t-t_2}{2t_1(t-t_2)}\right]^{m/2} \left\langle\!\!\left\langle \Phi[V(x)]\,H_m\left(\frac{\bar{x}_{jk} - x_0}{\sqrt{2t_1(t-t_2)/(t_1+t-t_2)}}\right)\right\rangle\!\!\right\rangle_{x(t)}, \end{aligned}$$

(differentiated ensemble average, with partial average) (17.837)

[78] Daniel A. Steck, *Classical and Modern Optics* (2006), available online at `http://steck.us/teaching`.

as we have just proven by induction, where again

$$\bar{x}_{jk} := \frac{kx_j + jx_{N-k}}{j+k} = \frac{(t-t_2)x(t_1) + t_1 x(t-t_2)}{t_1 + t - t_2} \tag{17.838}$$

in both discrete and continuous notation, where $t_1 = j\Delta t$ and $t - t_2 = k\Delta t$, with a total running time t of the Brownian bridge. Again, the expression here is approximate because we are assuming that $\partial_{x_i}\Phi[V(x)] \approx 0$ for all $0 < i < j$ and $N - k < j < N$. In the simpler, symmetric case where $j = k$, we have

$$\begin{aligned}\partial_{x_0}^m A(x_0) &\approx (j\Delta t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jj} - x_0}{\sqrt{j\Delta t}}\right)\right\rangle\!\!\right\rangle_{x(t)} \\ &= (t_1/t)^{-m/2} \left\langle\!\!\left\langle \Phi[V(x)]\, H_m\!\left(\frac{\bar{x}_{jj} - x_0}{\sqrt{t_1/t}}\right)\right\rangle\!\!\right\rangle_{x(t)},\end{aligned}$$

(differentiated ensemble average, with symmetric partial average) (17.839)

where

$$\bar{x}_{jj} = \frac{x_j + x_{N-j}}{2} = \frac{x(t_1) + x(t-t_1)}{2}. \tag{17.840}$$

The advantage here is clear, especially in the last expression. In Eq. (17.824), the statistical fluctuations in the path average grew as $N^{m/2}$. But since typically j and k are chosen for some fixed times j and k (e.g., as in the sojourn-time example), the fluctuations here are *independent* of N. Of course, since $t_1 < t/2$, the fluctuations still increase by a factor of $(t/t_1)^{1/2}$ for each derivative, but this is a huge improvement when working with large N.

One remaining detail is to examine the statistics of $\bar{x}_{jk}$ in the Hermite polynomial, to make sure it does not cause any problem. For a discrete Brownian bridge B_n of unit running time, the covariance is given by [Problem 17.10]

$$\langle\!\langle B_n\, B_m \rangle\!\rangle = \frac{[N - \max(m,n)]\min(m,n)}{N^2}, \tag{17.841}$$

which gives

$$\mathrm{Var}(x_j) = \mathrm{Var}(x_{N-j}) = \frac{j(N-j)\Delta t}{N} \tag{17.842}$$

and

$$\mathrm{Cov}(x_j, x_{N-j}) = \frac{j^2 \Delta t}{N} \tag{17.843}$$

for the symmetric case. Then setting $x_0 = 0$, we thus have

$$\begin{aligned}\mathrm{Var}[\bar{x}_{jj}] &= \frac{1}{4}\Big[\mathrm{Var}(x_j) + \mathrm{Var}(x_{N-j}) + 2\,\mathrm{Cov}(x_j, x_{N-j})\Big] \\ &= \frac{\Delta t}{2N}\Big[j(N-j) + j^2\Big] \\ &= \frac{j\Delta t}{2}.\end{aligned} \tag{17.844}$$

Thus the typical $\bar{x}_{jj}$ is of the order $\sqrt{j\Delta t}$, which is precisely what we see in the denominator of the Hermite polynomial.

Again, to use this partially averaged path integral as a numerical technique, it can be helpful to incorporate the Hermite polynomial into the sampling distribution for the path itself. To simplify the math somewhat, we will stick to the symmetric expression (17.839) for the path integral. We know from our discussion of the Brownian bridge above that we may choose the path coordinates at times $j\Delta t = t_1$ and $(N-j)\Delta t = t - t_1$, since the path at either time is Gaussian with variance $j(N-j)\Delta t^2 = t_1(t-t_1)$. However,

they are correlated because they are part of a bridge, with covariance $j^2\Delta t^2 = t_1^2$. We can decouple this correlation by working with $\bar{x}_{jj}$ and

$$\delta x_{jj} := \frac{x_j - x_{N-j}}{2} = \frac{x(t_1) - x(t-t_1)}{2}, \tag{17.845}$$

such that $x_j = \bar{x}_{jj} + \delta x_{jj}$ and $x_{N-j} = \bar{x}_{jj} - \delta x_{jj}$. We already computed $\mathrm{Var}[\bar{x}_{jj}]$, and we also need

$$\begin{aligned}\mathrm{Var}[\delta x_{jj}] &= \frac{1}{4}\Big[\mathrm{Var}(x_j) + \mathrm{Var}(x_{N-j}) - 2\,\mathrm{Cov}(x_j, x_{N-j})\Big] \\ &= \frac{\Delta t}{2N}\Big[j(N-j) - j^2\Big] \\ &= \frac{j(N-2j)\Delta t}{2N}.\end{aligned} \tag{17.846}$$

Of course, the simplification comes because of the vanishing covariance:

$$\mathrm{Cov}(\bar{x}_{jj}, \delta x_{jj}) = \frac{1}{4}\Big[\mathrm{Var}(x_j) - \mathrm{Var}(x_{N-j})\Big] = 0. \tag{17.847}$$

Thus the ensemble average in the path integral (17.839) implies a probability measure

$$P[x(t)] = \left[\frac{e^{-(\bar{x}_{jj}-x_0)^2/j\Delta t}}{\sqrt{\pi j\Delta t}}\,\frac{e^{-(\delta x_{jj})^2/j(1-2j/N)\Delta t}}{\sqrt{\pi j(1-2j/N)\Delta t}}\right]\frac{e^{-(x_{j+1}-x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}\cdots\frac{e^{-(x_{N-j}-x_{N-j-1})^2/2\Delta t}}{\sqrt{2\pi\Delta t}}, \tag{17.848}$$

where the bracketed factor gives the partially averaged first and last steps, and the rest of the factors give the small, in-between steps. Note that there is a factor of 1/2 that is now omitted from the bracketed factor, owing to the choice of $\bar{x}_{jj}$ and δx_{jj} as path variables (before transforming to these variables the corresponding normalization factors had a factor of 2 associated with each π); this factor of 1/2 is absorbed into the integration measure via the Jacobian derivative in the variable transformation. Then the Hermite polynomial may be grouped with the first Gaussian factor

$$\begin{aligned}P_{H_m}[x(t)] = \left\{\left[\frac{\eta_m}{\sqrt{\pi j\Delta t}}\left|H_m\!\left(\frac{\bar{x}_{jj}-x_0}{\sqrt{j\Delta t}}\right)\right|e^{-(\bar{x}_{jj}-x_0)^2/j\Delta t}\right]\frac{e^{-(\delta x_{jj})^2/j(1-2j/N)\Delta t}}{\sqrt{\pi j(1-2j/N)\Delta t}}\right\} \\ \times\,\frac{e^{-(x_{j+1}-x_j)^2/2\Delta t}}{\sqrt{2\pi\Delta t}}\cdots\frac{e^{-(x_{N-j}-x_{N-j-1})^2/2\Delta t}}{\sqrt{2\pi\Delta t}},\end{aligned} \tag{17.849}$$

where the factor

$$\eta_m^{-1} := \frac{1}{\sqrt{\pi j\Delta t}}\int_{-\infty}^{\infty}dx\,\left|H_m\!\left(\frac{x}{\sqrt{j\Delta t}}\right)\right|e^{-x^2/j\Delta t} = \frac{2}{\sqrt{\pi}}\int_0^{\infty}dx\;|H_m(x)|\;e^{-x^2} \tag{17.850}$$

normalizes the Hermite–Gaussian probability density for $\bar{x}_{jj}$. Then to summarize, we should take

$$\bar{x}_{jj} = x_0 + \sqrt{j\Delta t}\,\bar{z}, \qquad \delta x_{jj} = \sqrt{\frac{j(1-2j/N)\Delta t}{2}}\,z, \tag{17.851}$$

where z is standard normal, and $\bar{z}$ is chosen from the density

$$f(\bar{z}) = \frac{\eta_m}{\sqrt{\pi}}\;|H_m(x)|\;e^{-x^2}. \tag{17.852}$$

Then with the path measure (17.849), the path average (17.839) becomes

$$\partial_{x_0}^m A(x_0) \approx \eta_m^{-1}(j\Delta t)^{-m/2}\left\langle\!\!\left\langle\,\Phi[V(x)]\,\mathrm{sgn}\big[H_m(\bar{z})\big]\,\right\rangle\!\!\right\rangle_{H_m},$$

(differentiated ensemble average, with Hermite–Gauss path) (17.853)

where note that we kept the sign of the Hermite polynomial in the path-average argument rather than the measure, and we have included the factor of 1/2 that we discussed after Eq. (17.848). Then with the new path measure, x_j and x_{N-j} can be calculated via $x_j = \bar{x}_{jj} + \delta x_{jj}$ and $x_{N-j} = \bar{x}_{jj} - \delta x_{jj}$, while the rest of the path can be constructed as a Brownian bridge connecting x_j to x_{N-j} in time $t - 2t_1$.

17.15.2 Stochastic Differentiation

For a stochastic process $y(t)$ satisfying the SDE

$$dy = \alpha[y(t)]\,dt + \beta[y(t)]\,dW(t), \tag{17.854}$$

the notion of a **stochastic derivative** arises by considering the effect on $y(t)$ of a perturbation to $dW(t_0)$ at some earlier time t_0, where $0 \le t_0 \le t$. The corresponding derivative is typically written $D_{t_0}y(t)$, which for our purposes can be written as the partial derivative[79]

$$D_{t_0}y(t) = \frac{\partial y(t)}{\partial[dW(t_0)]}. \tag{17.855}$$
(stochastic derivative)

Another common, related notation is Dy, which is essentially the collection of all stochastic derivatives $D_{t_0}y(t)$. That is, considering t to be fixed, Dy is a function of t_0, with $0 \le t_0 \le t$.

This derivative generalizes the notion of a functional derivative to stochastic processes. Thus, Dy is essentially the functional derivative $\delta y/\delta(dW)$, regarding $\delta(dW)(t)$ as a perturbation to $dW(t)$. And while Dy acts as a gradient, $D_{t_0}y$ acts as a directional derivative in the "direction" $\delta(t - t_0)$, in the same sense as the functional derivative via a perturbation only at time t_0: $\delta y/\delta(dW)(t_0)$.

17.15.2.1 Example: Stochastic Derivative of the Wiener Process

As a first example of stochastic differentiation, we will begin by computing $D_{t_0}W(t)$. Since

$$W(t) = \int_0^t dW(t'), \tag{17.856}$$

the function to differentiate is a simple sum of increments $dW(t')$ and we are differentiating with respect to one of them. Then we can use

$$\frac{\partial[dW(t)]}{\partial[dW(t')]} = dt\,\delta(t - t'), \tag{17.857}$$

as a consequence of the independence of $dW(t)$ at different times: that is, the derivative is unity if $t = t'$, and zero otherwise. Thus,

$$D_{t_0}W(t) = \int_0^t D_{t_0}dW(t') = \int_0^t dt'\,\delta(t_0 - t'), \tag{17.858}$$

or simply

$$D_{t_0}W(t) = 1, \tag{17.859}$$

provided, of course, $0 \le t_0 \le t$. Otherwise this derivative vanishes.

17.15.2.2 Example: Stochastic Derivative of Additive Diffusion

As a slightly more sophisticated example, let's now compute Dy, where

$$dy(t) = \beta(t)\,dW(t). \tag{17.860}$$

[79] See Arturo Kohatsu–Higa and Miquel Montero, "Malliavin Calculus in Finance," in *Handbook of Computational and Numerical Methods in Finance* Svetlozar T. Rachev, Ed. (Springer, 2004), pp. 111-74, especially p. 130. This derivative is more generally known as the **Malliavin derivative**. For a more rigorous and sophisticated development, see David Nualart, *The Malliavin Calculus and Related Topics*, 2nd ed. (Springer, 2006), Section 1.2, p. 24.

We can do this by generalizing the solution from the previous section, which leads straightforwardly to

$$D_{t_0} y(t) = \beta(t_0). \tag{17.861}$$

Thus, a simple relabeling gives the result[80] $Dy = \beta$.

17.15.2.3 Variation Process

Going back to the stochastic derivative $D_{t_0} y(t)$, where $y(t)$ satisfies the prototype SDE (17.854), we can think of the derivative *itself* as a stochastic process (considering t to be the time variable, with fixed t_0). We can refer to this as the **derivative process** or the **variation process**. This describes the modified evolution of $y(t)$ as a linearized displacement, given a perturbation to $dW(t_0)$. That is, the perturbed evolution is given as $y(t) = y_0(t) + [D_{t_0} y_0(t)]\, \delta(dW)(t_0)$, where $y_0(t)$ is the unperturbed solution, and $\delta(dW)(t_0)$ is the perturbation to $dW(t_0)$.

Taking the stochastic derivative of Eq. (17.854) by applying the D_{t_0} operator, we find

$$d[D_{t_0} y(t)] = \alpha'[y(t)]\, [D_{t_0} y(t)]\, dt + \beta'[y(t)]\, [D_{t_0} y(t)]\, dW(t) + \beta[y(t)]\, D_{t_0} dW(t). \tag{17.862}$$

Using Eq. (17.857) again in the last term, and dividing through by $D_{t_0} y(t)$, we arrive at the SDE

$$\frac{d[D_{t_0} y(t)]}{D_{t_0} y(t)} = \alpha'[y(t)]\, dt + \beta'[y(t)]\, dW(t) + \frac{\beta[y(t_0)]}{D_{t_0} y(t)}\, \delta(t - t_0)\, dt,$$

(SDE for variation process) (17.863)

which yields the evolution of the variation processs.

To write down an integral solution for this SDE, the idea is now to integrate Eq. (17.863) from $t' = 0$ to t, which gives the perturbation to $y(t)$ due to the perturbation to $dW(t_0)$ (where $0 \le t_0 \le t$). Of course, there is no effect for $t' \le t_0$, so we may take $D_{t_0} y(t') = 0$ for all $t' < t_0$. The delta function induces a jump discontinuity in $D_{t_0} y(t')$ at $t' = t_0$ of height $\beta[y(t_0)]$. Thus, the task is to integrate the homogeneous differential equation

$$\frac{d[D_{t_0} y(t)]}{D_{t_0} y(t)} = \alpha'[y(t)]\, dt + \beta'[y(t)]\, dW(t) \tag{17.864}$$

from $t' = t_0$ to t, with initial condition $D_{t_0} y(t' = t_0) = \beta[y(t_0)]$. Rewriting the left-hand side of the homogeneous SDE,

$$d \log[D_{t_0} y(t)] + \frac{\{d[D_{t_0} y(t)]\}^2}{2[D_{t_0} y(t)]^2} = \alpha'[y(t)]\, dt + \beta'[y(t)]\, dW(t), \tag{17.865}$$

where from the second-order expansion $d \log x = \log(x + dx) - \log x = dx/x - (dx)^2/2x^2$, it follows that $dx/x = d \log x + (dx)^2/2x^2$. Using the SDE (17.864) to evaluate the second-order term, we find

$$d \log[D_{t_0} y(t)] = \alpha'[y(t)]\, dt - \frac{\beta'^2[y(t)]}{2}\, dt + \beta'[y(t)]\, dW(t). \tag{17.866}$$

Integrating from t_0 to t,

$$\log[D_{t_0} y(t)] - \log \beta[y(t_0)] = \int_{t_0}^{t} \left(\alpha'[y(t')] - \frac{\beta'^2[y(t')]}{2} \right) dt' + \int_{t_0}^{t} \beta'[y(t')]\, dW(t'). \tag{17.867}$$

Exponentiating this equation leads to the expression

$$D_{t_0} y(t) = \beta[y(t_0)]\, \exp\left[\int_{t_0}^{t} \left(\alpha'[y(t')] - \frac{\beta'^2[y(t')]}{2} \right) dt' + \int_{t_0}^{t} \beta'[y(t')]\, dW(t') \right],$$

(solution for variation process) (17.868)

which solves the SDE (17.863).

[80] Compare to the equivalent statement $DW(h) = h$ in the cryptic notation of Nualart, *op. cit.*, p. 25, just after Eq. (1.29), where it is not obvious that $W(h)$ can be interpreted as the integral of $h\, dW$, but see Example 1.1 on p. 3 in the corresponding lecture notes at `http://www.math.wisc.edu/~kurtz/NualartLectureNotes.pdf`; for a more direct comparison, see Giulia Di Nunno, Bernt Øksendal, and Frank Proske, *Malliavin Calculus for Lévy Processes with Applications to Finance* (Springer, 2009), p. 29, Eq. (3.8).

17.15.3 Integration by Parts

The notion of a stochastic derivative also gives a kind of integration by parts, which can be helpful in transforming certain path averages. As an example, consider the model path average

$$A(c) = \Big\langle\!\Big\langle \delta[W(t) - c] \Big\rangle\!\Big\rangle_{W(t)}. \tag{17.869}$$
(model ensemble average)

That is, we are averaging over Wiener paths $W(t)$, which only "score" when the endpoint $W(t)$ matches the desired ending point c. Since we know the explicit probability density for $W(t)$, we can compute this explicitly as

$$A(c) = \int_{-\infty}^{\infty} dx\, \frac{e^{-x^2/2t}}{\sqrt{2\pi t}}\, \delta(x - c) = \frac{e^{-c^2/2t}}{\sqrt{2\pi t}}. \tag{17.870}$$

However, as we will see, the general idea here will apply to more general stochastic processes, where we may not have an analytic expression for the density of paths.

From the standpoint of numerical simulation, Eq. (17.869) is very difficult, because only a subset of paths of zero measure contribute at all to the average, each with an infinite "score." In a numerical simulation, the delts function could, for example, be changed into a function of finite height and width, but the variance of a sample path average would still be very large for any function narrow enough to give an accurate result. Of course, one can also change explicitly to Brownian bridges that run to $W(t) = c$ as in Eq. (17.592), but this presumes that we know the distribution of $W(t)$. In this case we do, but we would like to explore more general methods for cases where the distribution is not known.

The approach we will try here is to rewrite the path functional as

$$A(c) = -\frac{1}{2}\partial_c \Big\langle\!\Big\langle \operatorname{sgn}[W(t) - c] \Big\rangle\!\Big\rangle_{W(t)}, \tag{17.871}$$

which removes the delta function at the expense of introducing a derivative. Now switching to shifted Brownian bridges

$$x(t) = x_0 + W(t) \tag{17.872}$$

with source point x_0, we can write the path functional as

$$A(c) = -\frac{1}{2}\partial_c \Big\langle\!\Big\langle \operatorname{sgn}[x(t) - c] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0} = \frac{1}{2}\partial_{x_0} \Big\langle\!\Big\langle \operatorname{sgn}[x(t) - c] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0}, \tag{17.873}$$

Then using the likelihood-ratio estimator (17.812),

$$A(c) = \frac{1}{2}\Big\langle\!\Big\langle \operatorname{sgn}[x(t) - c]\, \partial_{x_0} \log f[x(t)] \Big\rangle\!\Big\rangle_{x(t)}\Big|_{x_0=0}, \tag{17.874}$$

where $f[x(t)]$ is the probability density of the paths $x(t)$, which we regard as containing all the dependence on x_0. For Wiener paths,

$$\log f[x(t)] = -\frac{[x(t) - x_0]^2}{2t} - \log\sqrt{2\pi t}, \tag{17.875}$$

and thus

$$\partial_{x_0} \log f[x(t)] = -\frac{x(t) - x_0}{t} = -\frac{W(t)}{t}. \tag{17.876}$$

Then Eq. (17.874) becomes

$$A(c) = \frac{1}{2t}\Big\langle\!\Big\langle \operatorname{sgn}[W(t) - c]\, W(t) \Big\rangle\!\Big\rangle_{W(t)}. \tag{17.877}$$
(alternative path average)

We can verify by direct integration with respect to the probability desnity that this gives the correct answer. However, unlike the original path functional (17.869), *every* path $W(t)$ contributes value to the average. The "cost" is a weight of $W(t)/2t$ in addition to the integral of the delta function, but everything is well-behaved

here from a numerical standpoint. Note that other antiderivatives of the delta function are possible: the Heaviside function is an obvious choice, but then half of the paths contribution no ''information'' to the path average.

However, in deriving Eq. (17.877), we again used the explicit probability density for $W(t)$, as an illustration of the method. To perform a more general calculation, we cannot do this. Consider now the model path integral with more general stochastic processes:

$$\begin{aligned} A(c) &= \left\langle\!\!\left\langle \delta[y(t)-c] \right\rangle\!\!\right\rangle_{y(t)} \\ dy &= \alpha[y(t)]\,dt + \beta[y(t)]\,dW(t). \end{aligned} \qquad \begin{array}{r}(17.878)\\ \text{(model ensemble average)}\end{array}$$

To replace the delta function here, first we will compute the chain-rule derivative of the step function:

$$\frac{1}{2} D_{t_0} \mathrm{sgn}[y(t)-c] = \delta[y(t)-c]\, D_{t_0} y(t). \tag{17.879}$$

Solving for the delta function and integrating over time gives

$$\delta[y(t)-c] = \frac{1}{2t} \int_0^t dt_0\, \frac{D_{t_0}\mathrm{sgn}[y(t)-c]}{D_{t_0} y(t)}. \tag{17.880}$$

Now if we include an explicit probability measure for $dW(t_0)$, we can integrate by parts with respect to $dW(t_0)$ to write

$$\int dW(t_0)\, \frac{e^{-dW^2(t_0)/2dt}}{\sqrt{2\pi dt}} D_{t_0}\mathrm{sgn}[y(t)-c] = \int dW(t_0)\, \frac{e^{-dW^2(t_0)/2dt}}{\sqrt{2\pi dt}}\, \frac{dW(t_0)}{dt} \mathrm{sgn}[y(t)-c]. \tag{17.881}$$

Thus, *under an ensemble average* integration by parts amounts to writing

$$D_{t_0}\mathrm{sgn}[y(t)-c] = \frac{dW(t_0)}{dt}\, \mathrm{sgn}[y(t)-c]. \tag{17.882}$$

If we then introduce an expectation value in Eq. (17.880) and perform this integration by parts, we find

$$\left\langle\!\!\left\langle \delta[y(t)-c] \right\rangle\!\!\right\rangle = \frac{1}{2t} \int_0^t dt_0 \left\langle\!\!\left\langle \frac{D_{t_0}\mathrm{sgn}[y(t)-c]}{D_{t_0} y(t)} \right\rangle\!\!\right\rangle = \frac{1}{2t} \int_0^t dt_0 \left\langle\!\!\left\langle \frac{dW(t_0)}{dt}\, \frac{\mathrm{sgn}[y(t)-c]}{D_{t_0} y(t)} \right\rangle\!\!\right\rangle, \tag{17.883}$$

noting that $D_{t_0} y(t)$ is independent of $dW(t_0)$. Simplifying, we have[81]

$$A(c) = \frac{1}{2t} \left\langle\!\!\left\langle \mathrm{sgn}[y(t)-c] \int_0^t \frac{dW(t_0)}{D_{t_0} y(t)} \right\rangle\!\!\right\rangle, \qquad \begin{array}{r}(17.884)\\ \text{(alternate path functional)}\end{array}$$

where the derivative in the denominator is given by Eq. (17.868). Note that in the limit $\alpha = 0$ and $\beta = 1$, the derivative reduces to $D_{t_0} y(t) = D_{t_0} W(t) = 1$, and the path integral here reduces to the simpler one in Eq. (17.877). This expression also simplifies somewhat for the case of constant diffusion ($\beta = 1$), in the sense that the derivative (17.868) in the denominator of the integral reduces to the simpler form

$$D_{t_0} y(t) = \exp\left[\int_{t_0}^t dt'\, \alpha'[y(t')] \right], \tag{17.885}$$

so that Eq. (17.884) becomes

$$A(c) = \frac{1}{2t} \left\langle\!\!\left\langle \mathrm{sgn}[y(t)-c] \int_0^t dW(t_0)\, \exp\left[-\int_{t_0}^t dt'\, \alpha'[y(t')] \right] \right\rangle\!\!\right\rangle, \tag{17.886}$$

such that a drifting path still introduces a nontrivial weighting function that measures the drift-induced focusing or divergence of paths.

[81] A more general version of this argument is given in Peter K. Friz, ''Malliavin Calculus in Finance,'' Section 7.1, available at `http://www.math.nyu.edu/phd_students/frizpete/finance/my_case_lecture/malliavin_lecture.pdf`. See also Eric Fournié, Jean-Michel Lasry, Jérôme Lebuchoux, and Pierre-Louis Lions, ''Applications of Malliavin calculus to Monte-Carlo methods in finance. II,'' *Finance and Stochastics* **5**, 201 (2001), Section 4.1 (doi: 10.1007/PL00013529).

17.15.3.1 Digression: Faddeev–Popov Approach to Conditional Averages

An alternative approach, based on the method of Faddeev and Popov, can work to give regularized expressions for the conditional path averages in the previous section. To start with the simpler case of Eq. (17.869),

$$A(c) = \left\langle\!\!\left\langle \delta[W(t) - c] \right\rangle\!\!\right\rangle_{W(t)}, \qquad \begin{matrix}(17.887)\\ \text{(model ensemble average)}\end{matrix}$$

as a first step, let's write out the probability measure explicitly in discrete form with N path steps:

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j\right)^2\right] \delta\big[W(t) - c\big]. \tag{17.888}$$

Note that we can regard $W(t)$ here as an explicit sum:

$$W(t) = \sum_{j=0}^{N-1} \Delta W_j. \tag{17.889}$$

Now the key is to note that because we are integrating over all possible values of ΔW_j, the integral $A(c)$ is invariant if we shift the variable ΔW_j. In particular, suppose that we introduce a scalar quantity ω, and shift each coordinate according to

$$\Delta W_j \longrightarrow \Delta W_j - \frac{\omega}{N}. \tag{17.890}$$

Then $A(c)$ is independent of ω, and we can regard this shift as a kind of *gauge transformation* with gauge parameter ω. Implementing this gauge transformation, Eq. (17.888) becomes

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{\omega}{N}\right)^2\right] \delta\big[W(t) - c - \omega\big]. \tag{17.891}$$

Now suppose that we introduce some function $g(\omega)$, with the only requirement that it be normalized such that

$$\int d\omega \, g(\omega) = 1. \tag{17.892}$$

Then we don't change the value of Eq. (17.891) by multiplying by $g(\omega)$ and integrating, thus taking a linear combination of the same value. The delta function disappears in the result:

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1} \, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2\right] g\big[W(t) - c\big]. \tag{17.893}$$

In the exponentiated summand, we can multiply out the quadratic factor to obtain

$$\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2 = \left(\Delta W_j\right)^2 - 2\Delta W_j \frac{\big[W(t) - c\big]}{N} + \frac{\big[W(t) - c\big]^2}{N^2}, \tag{17.894}$$

and thus when summed over j, we find

$$\sum_{j=0}^{N-1}\left(\Delta W_j - \frac{W(t) - c}{N}\right)^2 = \sum_{j=0}^{N-1}\left(\Delta W_j\right)^2 + \frac{c^2}{N} - \frac{\big[W(t)\big]^2}{N}. \tag{17.895}$$

In this case the path integral (17.893) becomes

$$A(c) = \left\langle\!\!\left\langle \exp\left[\frac{\big[W(t)\big]^2 - c^2}{2t}\right] g\big[W(t) - c\big] \right\rangle\!\!\right\rangle_{W(t)} \tag{17.896}$$

after again hiding the path measure in the ensemble average. At this point we are free to choose $g(\omega)$ in order to simplify or improve the statistical behavior of this ensemble average. For example, if we choose $g(\omega)$ to be a Gaussian of variance t,

$$g(\omega) = \frac{1}{\sqrt{2\pi t}}\, e^{-\omega^2/2t}, \tag{17.897}$$

then the path average (17.896) becomes

$$A(c) = \frac{e^{-c^2/t}}{\sqrt{2\pi t}} \left\langle\!\!\left\langle e^{cW(t)/t} \right\rangle\!\!\right\rangle_{W(t)}, \tag{17.898}$$

(alternate path integral)

which attains the correct value of $e^{-c^2/2t}/\sqrt{2\pi t}$. On the other hand, if we take $g(\omega)$ to be centered at $-c$ instead of at 0,

$$g(\omega) = \frac{1}{\sqrt{2\pi t}}\, e^{-(\omega+c)^2/2t}, \tag{17.899}$$

the path integral (17.896) takes on the average value directly, with zero variance. This illustrates the advantage of adapting $g(\omega)$ to the problem as well as can possibly be done.

Of course, this approach becomes more complicated for the more general diffusion problem (17.878):

$$\begin{aligned} A(c) &= \left\langle\!\!\left\langle \delta[y(t) - c] \right\rangle\!\!\right\rangle_{y(t)} \\ dy &= \alpha[y(t)]\, dt + \beta[y(t)]\, dW(t). \end{aligned} \tag{17.900}$$

(model ensemble average)

The general approach, however, will be the same. Writing out the details of the path integral, we have

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1}\, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1} \big(\Delta W_j\big)^2\right] \delta\big[y(t) - c\big] \tag{17.901}$$

when integrating in terms of the unshifted path variables, where

$$y(t) = \int_0^t dt'\, \alpha\big[y(t')\big] + \int_0^t dW(t')\, \beta\big[y(t')\big]. \tag{17.902}$$

Implementing the same gauge transformation (17.890), the result is

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1}\, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1} \Big(\Delta W_j - \frac{\omega}{N}\Big)^2\right] \delta\left[y(t) - c - \frac{\omega}{N}\sum_{j=0}^{N-1} \frac{\partial y(t)}{\partial \Delta W_j}\right], \tag{17.903}$$

where in the delta function we have expanded to $O(N^{-1})$. Then we can define the shorthand for the last term in the argument of the delta function,

$$\mathcal{A} := \frac{1}{N}\sum_{j=0}^{N-1} \frac{\partial y(t)}{\partial \Delta W_j} = \frac{1}{t}\int_0^t dt_0\, D_{t_0} y(t), \tag{17.904}$$

in both discrete and continuous notation. This quantity is written more explicitly using the expression (17.868) for the Malliavin derivative, which in the case of constant diffusion ($\beta = 1$), for example, simplifies via Eq. (17.885) to give

$$\mathcal{A} = \frac{1}{t}\int_0^t dt_0\, \exp\left[\int_{t_0}^t dt'\, \alpha'[y(t')]\right]. \tag{17.905}$$

Then we can proceed by introducing the function $g(\omega)$ and integrating over all ω, with the result

$$A(c) = \left(\frac{N}{2\pi t}\right)^{N/2} \int d\Delta W_0 \cdots d\Delta W_{N-1}\, \exp\left[-\frac{N}{2t}\sum_{j=0}^{N-1} \left(\Delta W_j - \frac{y(t) - c}{N|\mathcal{A}|}\right)^2\right] \frac{1}{|\mathcal{A}|}\, g\left[\frac{y(t) - c}{|\mathcal{A}|}\right], \tag{17.906}$$

where we have used $\delta(ax) = \delta(x)/|a|$. Multiplying out the exponentiated factor and hiding the path measure again gives

$$A(c) = \left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \exp\left[\frac{[y(t) - c]}{2t|\mathcal{A}|} \left(2W(t) - \frac{[y(t) - c]}{|\mathcal{A}|} \right) \right] g\left[\frac{y(t) - c}{|\mathcal{A}|} \right] \right\rangle\!\!\right\rangle_{y(t)} . \tag{17.907}$$

Again, we can choose the form of $g(\omega)$ to simplify the problem, but for a general drift function $\alpha[y]$, we cannot choose it to arrive directly at the answer. Choosing the gaussian form (17.897) again gives the result

$$A(c) = \frac{1}{\sqrt{2\pi t}} \left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \exp\left[\frac{[y(t) - c]}{t|\mathcal{A}|} \left(W(t) - \frac{[y(t) - c]}{|\mathcal{A}|} \right) \right] \right\rangle\!\!\right\rangle_{y(t)} , \tag{17.908}$$

which does not make the path integral particularly simple, but at least it casts the path contribution in terms of a direct difference between $W(t)$ and path $y(t)/|A|$ in the case $c = 0$. Note that in the case $\alpha = 0$, then $y(t)$ reduces to $W(t)$ and $A = 1$, so that Eq. (17.908) reduces to Eq. (17.898). Of course, we can also make the probably better choice (17.899), which leads to

$$A(c) = \frac{e^{-c^2/2t}}{\sqrt{2\pi t}} \left\langle\!\!\left\langle \frac{1}{|\mathcal{A}|} \exp\left[\frac{[y(t) - c]}{t|\mathcal{A}|} \left([W(t) - c] - \frac{[y(t) - c]}{|\mathcal{A}|} \right) \right] \right\rangle\!\!\right\rangle_{y(t)} ,$$

(alternate path integral) (17.909)

which again reduces directly to the mean value (with no variance among the paths) in the case $\alpha = 0$. Note in the path integrals how $y(t)$ is accompanied by a factor of $\mathcal{A}$; intuitively, from the discussion in Section 17.15.2.3, this factor gives a time-averaged measure of the divergence of trajectories in the vicinity of $y(t)$, and thus accounts for any focusing or defocusing effects due to gradients in $\alpha(y)$.

17.16 Exercises

Problem 17.1

Geometric Brownian motion describes the time-dependence price S of a stock according to the SDE

$$dS = S\left(\mu + \frac{\sigma^2}{2}\right)dt + S\sigma\, dW, \tag{17.910}$$

where μ represents the steady growth of the stock value, and σ represents the stock volatility (which is assumed to be constant within this model). Show that

$$S(t) = S_0 e^{\mu t + \sigma W(t)} \tag{17.911}$$

satisfies the above SDE.

Problem 17.2

Show that the vector SDE (17.145)

$$dx_i = \alpha_i(\mathbf{x}, t)\, dt + \beta_{ij}(\mathbf{x}, t)\, dW_j \tag{17.912}$$

is equivalent to the Fokker–Planck equation (17.146),

$$\partial_t f(\mathbf{x}, t) = -\partial_i \alpha_i(\mathbf{x}, t) f(\mathbf{x}, t) + \frac{1}{2}\partial_i \partial_j D_{ij}(\mathbf{x}, t) f(\mathbf{x}, t), \tag{17.913}$$

where

$$D_{ij} := \beta_{ik}\beta_{jk} = \beta_{ik}\beta^{\mathsf{T}}_{kj} = (\boldsymbol{\beta}\boldsymbol{\beta}^{\mathsf{T}})_{ij}. \tag{17.914}$$

Problem 17.3

By formally resumming the Taylor series expansion, compute $\exp(dN)$.

Problem 17.4

Recall the Poisson distribution in terms of the single parameter λ has the form

$$P(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \tag{17.915}$$

where n is a nonnegative integer. Also, suppose that N is a Poisson-distributed random variable.

(a) Show that $P(n)$ is normalized.

(b) Show that $\langle N \rangle = \lambda$.

(c) Show that $\mathrm{Var}[N] = \lambda$.

Problem 17.5

Consider a monochromatic field with added, white (Gaussian) frequency noise,

$$E^{(+)}(t) = E_0^{(+)} e^{-i\omega_{\mathrm{L}} t - i\phi(t)}, \tag{17.916}$$

so that the instantaneous frequency is

$$\frac{d\phi_{\mathrm{total}}}{dt} = \omega_{\mathrm{L}} + \frac{d\phi}{dt}, \tag{17.917}$$

where $\phi_{\text{total}} = \omega_{\text{L}} t + \phi(t)$ is the total temporal phase, and the noisy phase is given in terms of a Wiener process by

$$d\phi(t) = \sqrt{\gamma}\, dW(t). \tag{17.918}$$

Use the Wiener–Khinchin theorem to show that the spectrum is Lorentzian with full width at half maximum γ.

Problem 17.6

Let $y(t)$ denote a Brownian bridge (i.e., $y(0) = y(1) = 0$), and let $y_0, y_1, \ldots, y_N$ denote samples of $y(t)$, taken at times $t_j = j\Delta t = j/N$. (Note that $y_0 = y_N = 0$.) Consider the Gaussian path integral

$$I := \int dy_1 \ldots dy_{N-1} \exp\left[-\frac{N}{2}\left(\sum_{j=1}^{N}(y_j - y_{j-1})^2\right)\right]. \tag{17.919}$$

(a) Evaluate this integral by using the recurrence (17.305)

$$y_n = \frac{y'_n}{\sqrt{c_n}} + \frac{y_{n+1}}{c_n}. \tag{17.920}$$

to decouple the integral into a product of Gaussian integrals.

(b) Use the $y(t) = W(t) - tW(1)$ construction of the Brownian bridge, along with the independence of $y(t)$ and $W(1)$, to evaluate this integral, to show the consistency of these approaches.

Problem 17.7

Show that the integral expression (17.335)

$$B(t) = (1-t)\int_0^t \frac{dW(t')}{1-t'} \tag{17.921}$$

solves the SDE (17.334)

$$dB = -\left(\frac{B}{1-t}\right)dt + dW. \tag{17.922}$$

Problem 17.8

Derive the correlation function

$$\langle\!\langle B(t)\, B(t')\rangle\!\rangle = \min(t,t') - tt' \tag{17.923}$$

(a) using the definition (17.287)

$$B(t) := W(t) - tW(1). \tag{17.924}$$

(b) using the definition (17.335)

$$B(t) := (1-t)\int_0^t \frac{dW(t')}{1-t'}. \tag{17.925}$$

Problem 17.9

Show that if $B(t)$ is a standard Brownian bridge, the covariance of two Brownian-bridge increments $\text{Cov}[dB(t), dB(t')] = 0$ provided $t \neq t'$.

Problem 17.10

(a) Using the recurrence for the finite Brownian bridge [Eq. (17.310)],

$$\begin{aligned} B_0 &= 0 \\ B_n &= z_n \sqrt{\frac{N-n}{N(N-n+1)}} + \left(\frac{N-n}{N-n+1}\right) B_{n-1}, \qquad n = 1, \ldots, N-1 \\ B_N &= 0, \end{aligned} \tag{17.926}$$

derive an explicit formula for the Brownian-bridge samples B_n in terms of (only) the standard-normal deviates z_n.

(b) Show that

$$\langle\langle B_n \, B_m \rangle\rangle = \frac{[N - \max(m,n)] \min(m,n)}{N^2}, \tag{17.927}$$

and show that this is consistent with what you know for the continuous-time limit.

(c) Show that

$$\langle\langle \Delta B_n \, \Delta B_m \rangle\rangle = \delta_{nm} \frac{(N-n-1)}{N(N-n)} + \frac{\min(n,m)}{N^2[N - \min(n,m)]} + \frac{(\delta_{nm} - 1)}{\sqrt{N}(N-n)(N-m)}. \tag{17.928}$$

and show that this is consistent with what you know for the continuous-time limit, if we take

$$\Delta B_n := B_{n+1} - B_n. \tag{17.929}$$

Problem 17.11

For the standard Brownian bridge $B(t)$:

(a) Use the statistics derived in Section 17.7 derived for the Brownian bridge to write down the probability density for $B(t)$ at any time $0 \le t \le 1$.

(b) Compute the *time-averaged* density for $B(t)$, where the average is taken over $0 \le t \le 1$.

Problem 17.12

Given a state-dependent diffusion function of the form

$$\sigma(y) = \sigma_0 + (\sigma_1 - \sigma_0)\, \Theta(y - d), \tag{17.930}$$

where the diffusion rate is σ_0 for $y < d$ and σ_1 for $y > d$, show that an explicit solution (17.349) for the SDE (17.338)

$$dy(t) = \sigma(y) \circ dW(t) \tag{17.931}$$

with the conditions $y(1) = y(0) = 0$, is given by

$$y(t) = \sigma_0 B(t) + \left[(\sigma_1 - \sigma_0)B(t) + d\left(1 - \frac{\sigma_1}{\sigma_0}\right)\right] \Theta[B(t) - d], \tag{17.932}$$

where $\Theta(x)$ is the Heaviside function and $B(t)$ is a standard Brownian bridge.

Problem 17.13

(a) Show that

$$\delta[f(x)] = \sum_{x_0 \in f^{-1}(0)} \frac{\delta(x - x_0)}{|f'(x_0)|}, \tag{17.933}$$

where the sum is over the (isolated) zeros of $f(x)$. Good ways to proceed are: to write the delta function as a derivative of the step function $\Theta[f(x)]$, or to treat the delta function as a limit of a more "reasonable" function, such as a box function of width a and height $1/a$.

(b) Show that

$$\delta'[f(x)] = \sum_{x_0 \in f^{-1}(0)} \left[\frac{\delta'(x-x_0)}{f'(x_0)|f'(x_0)|} + \frac{f''(x_0)\,\delta(x-x_0)}{|f'(x_0)|^3} \right]. \tag{17.934}$$

The same suggested approaches apply here.

Problem 17.14

Recall that for a generalized Brownian bridge $B_{0\to c}(t)$ that connects $B_{0\to c}(0) = 0$ to final point $B_{0\to c}(1) = c$, the probability of crossing a boundary at d is [Eq. (17.387)]

$$P_{\text{cross}}[B_{0\to c}(t); d, c] = e^{-2d(d-c)} \qquad (d \geq 0, c). \tag{17.935}$$

Using *only* this expression (along with basic probability theory and your own cunning) to obtain the analogous expression [Eq. (17.372)]

$$P_{\text{cross}}[W(t); d, t] = \operatorname{erfc}\left(\frac{d}{\sqrt{2t}}\right) \tag{17.936}$$

for a standard Wiener path $W(t)$ to cross the boundary at $d \geq 0$ in the time interval $[0, t]$.

Hint: rescale the bridge probability to account for the alternate time interval, and interpret this as a conditional probability on a Wiener path $W(t)$, given that $W(t) = c$.

Problem 17.15

Consider a Wiener process $W(t)$. Show that for any d, compute the probability that $W(t') = d$ for some time $0 \leq t' \leq T$, and show that it approaches unity as $T \longrightarrow \infty$.

Problem 17.16

Rederive the probability density 17.710

$$f_\ell(x) = \left[1 - e^{[c^2 - (|d|+|c-d|)^2]/2t}\right] \delta(x - 0^+) + \frac{1}{t}\left(x + |d| + |c-d|\right) e^{[c^2-(x+|d|+|c-d|)^2]/2t}, \tag{17.937}$$

for the local time of a Brownian bridge $B_{t(0\to c)}(t)$ pinned to c at time t. Use an alternate derivation: begin with the same Feynman–Kac-based approach as in Section 17.13.3, but take $g(x) = \Theta(c - x)$ instead of $g(x) = e^{ik(x-c)}$, and then after working out the solution of the appropriate ODE, differentiate with respect to c to obtain the pinning delta function.

Note that the Wiener path can be pinned even more directly by taking $g(x) = \delta(x - c)$. The advantage of the method using $g(x) = e^{ik(x-c)}$ is that it replaces the book-keeping of several different cases with a single extra integral over k. Also, note that by avoiding the c-derivative and completing the calculation, it is possible to obtain the joint probability of $W(t)$ occupying $[0, \infty)$ and the local time $\ell[W(t); d]$ taking on a particular value.

Problem 17.17

(a) Using the calculation of the probability density for the local time $\ell[W(t); d]$ as a template, derive

the following formula for the expectation value[82]

$$\lambda\int_0^\infty dt\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-s\ell[W(t);d]-s'\ell[W(t);d']\Big)\right\rangle\!\!\right\rangle_{W(t)} = 1-\frac{s\left[\sqrt{2\lambda}+s'\left(1-e^{-\sqrt{2\lambda}|d-d'|}\right)\right]e^{-\sqrt{2\lambda}|d|}+s'\left[\sqrt{2\lambda}+s\left(1-e^{-\sqrt{2\lambda}|d-d'|}\right)\right]e^{-\sqrt{2\lambda}|d'|}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}, \tag{17.938}$$

giving the dual moment-generating function for the two local times $\ell[W(t);d]$ and $\ell[W(t);d']$, given that the process $W(t)$ is stopped exponentially at rate λ.

(b) Repeat for a Brownian bridge $B_{t(0\to c)}(t)$, pinned at the exponentially distributed stopping time t to $B(t)=c$, to find the analogous result[83]

$$\begin{aligned}\lambda\int_0^\infty dt\,e^{-\lambda t}\,\frac{e^{-c^2/2t}}{\sqrt{2\pi t}}\left\langle\!\!\left\langle \exp\Big(-s\ell[B_{t(0\to c)}(t');d]-s'\ell[B_{t(0\to c)}(t');d']\Big)\right\rangle\!\!\right\rangle_{B_{t(0\to c)}(t')} \\ = \sqrt{\frac{\lambda}{2}}\,e^{-\sqrt{2\lambda}|c|}\left(1-\frac{s\left[\left(\sqrt{2\lambda}+s'\right)e^{-\sqrt{2\lambda}|d|}-s'e^{-\sqrt{2\lambda}(|d'|+|d-d'|)}\right]e^{-\sqrt{2\lambda}(|c-d|-|c|)}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}\right. \\ \left.-\frac{s'\left[\left(\sqrt{2\lambda}+s\right)e^{-\sqrt{2\lambda}|d'|}-se^{-\sqrt{2\lambda}(|d|+|d-d'|)}\right]e^{-\sqrt{2\lambda}(|c-d'|-|c|)}}{\left(\sqrt{2\lambda}+s\right)\left(\sqrt{2\lambda}+s'\right)-ss'e^{-2\sqrt{2\lambda}|d-d'|}}\right),\end{aligned} \tag{17.939}$$

where the Gaussian factor on the left-hand side indicates this expectation value was taken with respect to a joint distribution for the local times and for the stopping point of $B(t)$.

Problem 17.18

(a) Using the representation

$$\delta'(x)=\lim_{a\to 0^+}\frac{\delta(x+a/2)-\delta(x-a/2)}{a} \tag{17.940}$$

and the results of Problem 17.17, derive the following formula for the expectation value

$$\lambda\int_0^\infty dt\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-s\ell'[W(t);d]\Big)\right\rangle\!\!\right\rangle_{W(t)}=1-e^{-\sqrt{2\lambda}|d|}, \tag{17.941}$$

giving the moment-generating function for the local-time derivative $\ell'[W(t);d]$ and $\ell[W(t);d']$, given that the process $W(t)$ is stopped exponentially at rate λ.

(b) Invert the λ and s Laplace transforms in this result to obtain an expression for the probability density $f_{\ell'}(x)$ of $\ell'[W(t);d]$. You should obtain something a bit funny; what gives?

Problem 17.19

Consider closed Brownian bridges $B_T(t)$ [i.e., such that $B_T(T)=0$].

(a) Using the probability method of Eqs. (17.592), show that

$$\left\langle\!\!\left\langle \delta[B_T(\tau)-c]\,F[B_T(t)]\right\rangle\!\!\right\rangle=\Phi[c,\tau(1-\tau/T)]\left\langle\!\!\left\langle F\big[B_T(t)\big]\right\rangle\!\!\right\rangle_{B_T(\tau)=c}, \tag{17.942}$$

[82] Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhäuser, 2002), p. 177, Eq. (1.18.1).

[83] This is similar to, but not quite the same as, Borodin and Salminen, *op. cit.*, p. 177, Eq. (1.18.5).

where the conditional ensemble average on the right-hand side refers to Brownian bridges pinned to c at time τ ($0 < \tau < T$), and

$$\Phi(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-x^2/2\sigma^2} \tag{17.943}$$

is the centered Gaussian distribution with variance σ^2.

(b) Prove the same result by adapting the explicit-integration method of Section 17.12.4.

Problem 17.20

Work out an analytic expression for the path integral (17.900) in the special case $\alpha(y) = -ay$,

$$\begin{aligned} A(c) &= \Big\langle\!\Big\langle\, \delta[y(t) - c] \,\Big\rangle\!\Big\rangle_{y(t)} \\ dy &= -a\, y(t)\, dt + dW(t), \end{aligned} \tag{17.944}$$

under the initial condition $y(0) = 0$, and verify numerically that the Faddeev–Popov form (17.909) of the path integral converges to the same value.

Chapter 18

Quantum Trajectories for Photodetection

18.1 Quantum Jumps

In deriving the **unconditioned** master equation for spontaneous emission in Section 11.5,

$$\partial_t \rho = -\frac{i}{\hbar}[H,\rho] + \Gamma \mathcal{D}[\sigma]\rho,$$

(unconditioned master equation for spontaneous emission) (18.1)

where the Lindblad superoperator $\mathcal{D}[\sigma]\rho$ is again given by

$$\mathcal{D}[c]\rho := c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right), \tag{18.2}$$

(Lindblad superoperator)

we have explicitly ignored the state of the field by computing the partial trace. Now we will consider what happens when we measure it. In particular, we will assume that we make *projective* measurements of the field photon number in every mode, not distinguishing between photons in different modes. It is this extra interaction that will yield the continuous measurement of the atomic state.

From the relation

$$\partial_t \rho_{\mathrm{ee}} = -\Gamma \rho_{\mathrm{ee}} \tag{18.3}$$

that we derived in the Weisskopf–Wigner treatment of spontaneous emission [Eq. (11.42)], the transition probability in a time interval of length dt is $\Gamma\rho_{\mathrm{ee}}\,dt = \Gamma\langle\sigma^\dagger\sigma\rangle\,dt$, where we recall that $\sigma^\dagger\sigma = |\mathrm{e}\rangle\langle\mathrm{e}|$ is the excited-state projection operator. Then assuming an ideal detector that detects photons at all frequencies, polarizations, and angles, there are two possibilities during this time interval:

1. **No photon detected.** The detector does not ''click'' in this case, and this possibility happens with probability $1-\Gamma\langle\sigma^\dagger\sigma\rangle\,dt$. The same construction as above for the master equation carries through, so we keep the equations of motion for ρ_{ee}, ρ_{eg}, and ρ_{ge}, where the last two are given by

$$\partial_t \rho_{\mathrm{ge}} = \left(i\omega_0 - \frac{\Gamma}{2}\right)\rho_{\mathrm{ge}}. \tag{18.4}$$

   However, we do not keep the same equation for ρ_{gg}: no photodetection implies that the atom does not return to the ground state. To see this, recall that the atom–field interaction Hamiltonian H_{AF} in the rotating-wave approximation contains a term of the form $\sigma a^\dagger$. The only way for the atom to end up in the ground state (the action of σ) is for a photon to be created (the action of $a^\dagger$). So if no photon is detected, then σ could not have acted on the atom. Thus, $\partial_t\rho_{\mathrm{gg}} = 0$. This case is thus generated by the master equation

$$\partial_t \rho = -\frac{i}{\hbar}[H,\rho] - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+. \tag{18.5}$$

This equation is the same as the unconditioned equation (18.1), except for the $\sigma\rho\sigma^\dagger$ term, which is precisely what we argued should be omitted. This evolution is unnormalized since $\mathrm{Tr}[\rho]$ decays to zero at long times. We can remedy this by explicitly renormalizing the state $\rho(t+dt)$, which gives

$$\partial_t \rho = -\frac{i}{\hbar}[H,\rho] - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+ + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho. \tag{18.6}$$

This follows from adding an extra term to the master equation to satisfy the normalization $\mathrm{Tr}[\rho]=1$, which implies $d\mathrm{Tr}[\rho] = \mathrm{Tr}[d\rho] = 0$.

2. **Photon detected.** A click on the photodetector occurs with probability $\Gamma\langle\sigma^\dagger\sigma\rangle\,dt$. Again, the interaction Hamiltonian H_{AF} in the rotating-wave approximation contains a term of the form $\sigma a^\dagger$, which tells us that photon creation (and subsequent detection) is accompanied by lowering of the atomic state. Thus, the evolution for this time interval is given by the reduction

$$\rho(t+dt) = \frac{\sigma\rho(t)\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle}. \tag{18.7}$$

We can write this in differential form as

$$d\rho = \frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho, \tag{18.8}$$

so that ρ is subtracted from itself, and is thus replaced by the first term, which is the lowered and renormalized density operator.

The overall evolution is stochastic, with either case occurring during a time interval dt with the stated probabilities.

We can explicitly combine these two probabilities by using the counting process $N(t)$ (see Sections 17.5.1 and 17.5.2). In any given time interval dt, we define dN such that it is unity with probability $\Gamma\langle\sigma^\dagger\sigma\rangle\,dt$ and zero otherwise. Thus, we can write the average over all possible photodetection histories as

$$\langle\!\langle dN\rangle\!\rangle = \Gamma\langle\sigma^\dagger\sigma\rangle\,dt. \tag{18.9}$$
(instantaneous photodetection probability)

Now we can add the two above possible cases together, with a weighting factor of dN for the second case, to obtain a **stochastic master equation** (SME):

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+ dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right) dN.$$
(SME for direct photodetection) (18.10)

It is unnecessary to include a weighting factor of $(1-dN)$ for the first term, since $dN\,dt = 0$. The evolution here is smooth, except punctuated by individual events where the density operator changes discontinuously to the ground state. We will refer to such events as **quantum jumps**.[1] Note that while we are using the notation $dN(t)$ that suggests a Poisson process, because of the state dependence of the mean, $N(t)$ is not Poisson-distributed in general (photon antibunching is a good example), and it is more properly called a counting process (Section 17.5.2).

This SME gives the evolution of the quantum state *conditioned* on the results of the photodetection measurement. To see this explicitly, note that we can write the current from the idealized photodetector as

$$I_{\mathrm{det}}(t) = Q_{\mathrm{ph}}\frac{dN(t)}{dt}, \tag{18.11}$$

where Q_{ph} is the total charge conducted by the detector for each photon detection event (which we assume to be deterministic and perfectly repeatable). The current thus appears ideally as a sequence of random

[1] E. Schrödinger, "Are There Quantum Jumps? Part I," *The British Journal for the Philosophy of Science* **3**, 109 (1952); E. Schrödinger, "Are There Quantum Jumps? Part II," *The British Journal for the Philosophy of Science* **3**, 233 (1952).

delta functions representing the detected photons, and dN is the same quantity that appears in the SME. We can thus write the SME directly in terms of the photocurrent as

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H,\rho] - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+ + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)\tilde{I}_{\rm det}(t). \tag{18.12}$$

(SME with measurement record)

where $\tilde{I}_{\rm det}(t) := I_{\rm det}(t)/Q_{\rm ph}$ is the scaled detector current. We can also refer to $\tilde{I}(t)$ as the **measurement record**, since it contains all the information from the measurement of the resonance fluorescence. The SME (18.12) thus shows how to incorporate the measurement record into the evolution of the quantum state—the observer's state of knowledge of the atomic system.

18.1.1 Ensemble Average

To work out the ensemble average of the SME (18.10), we will make use of the ensemble average $\langle\!\langle dN\rangle\!\rangle$ from Eq. (18.9). It is tempting to assume the statistical independence of $\rho(t)$ and $dN(t)$,

$$\langle\!\langle \rho\, dN\rangle\!\rangle = \langle\!\langle\rho\rangle\!\rangle\langle\!\langle dN\rangle\!\rangle \qquad \text{(wrong!)}, \tag{18.13}$$

similar to the statement $\langle\!\langle \rho\, dW\rangle\!\rangle = 0$ from Itō calculus. Although this assumption does in fact lead to the correct ensemble-averaged master equation, it does so for the wrong reasons: as we can see from Eq. (18.9) that $dN(t)$ depends on $\rho(t)$ [though $\rho(t)$ depends only on $dN(t')$ for $t' < t$, not on $dN(t)$].

However, we will still need to work out the expectation values like $\langle\!\langle\rho\, dN\rangle\!\rangle$, that appear in the last term of the ensemble average of Eq. (18.10). Treating the explicit possible outcomes n of $dN(t)$ in terms of the corresponding probablities $P[dN(t)=n]$,

$$\begin{aligned}\langle\!\langle\rho(t)\,dN(t)\rangle\!\rangle &= \left\langle\!\!\left\langle \rho(t)\sum_{n\in\{0,1\}} nP[dN(t)=n]\right\rangle\!\!\right\rangle \\ &= \left\langle\!\!\left\langle \rho(t)P[dN(t)=1]\right\rangle\!\!\right\rangle \\ &= \Gamma\left\langle\!\!\left\langle\rho(t)\langle\sigma^\dagger\sigma\rangle\right\rangle\!\!\right\rangle dt.\end{aligned} \tag{18.14}$$

Similarly, it follows that

$$\left\langle\!\!\left\langle \frac{\rho(t)\,dN(t)}{\langle\sigma^\dagger\sigma\rangle}\right\rangle\!\!\right\rangle = \Gamma\langle\!\langle\rho(t)\rangle\!\rangle\, dt. \tag{18.15}$$

Using these two relations, the ensemble average of Eq. (18.10) becomes

$$\begin{aligned} d\langle\!\langle\rho\rangle\!\rangle &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \Gamma\left\langle\!\!\left\langle\langle\sigma^\dagger\sigma\rangle\,\rho\right\rangle\!\!\right\rangle dt + \left\langle\!\!\left\langle\left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle}-\rho\right)dN\right\rangle\!\!\right\rangle \\ &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \Gamma\left\langle\!\!\left\langle\langle\sigma^\dagger\sigma\rangle\,\rho\right\rangle\!\!\right\rangle dt + \Gamma\left(\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger - \left\langle\!\!\left\langle\rho\langle\sigma^\dagger\sigma\rangle\right\rangle\!\!\right\rangle\right)dt \\ &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \Gamma\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger dt. \end{aligned} \tag{18.16}$$

The resulting ensemble-averaged master equation,

$$\frac{d\langle\!\langle\rho\rangle\!\rangle}{dt} = -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle] + \Gamma\mathcal{D}[\sigma]\langle\!\langle\rho\rangle\!\rangle, \tag{18.17}$$

(ensemble-averaged SME)

has precisely the same form as the *unconditioned* master equation (18.1) that we derived earlier. The unconditioned master equation thus follows from a measurement process, where the information from the measurement is discarded, and the observer averages (traces) over all possible measurement outcomes.

18.1.2 Quantum Trajectories and the Stochastic Schrödinger Equation

It is a fairly easy exercise to verify that this master equation is equivalent to the **stochastic Schrödinger equation** (SSE)

$$d|\psi\rangle = -\frac{i}{\hbar}H|\psi\rangle dt + \frac{\Gamma}{2}\left(\langle\sigma^\dagger\sigma\rangle - \sigma^\dagger\sigma\right)|\psi\rangle dt + \left(\frac{\sigma}{\sqrt{\langle\sigma^\dagger\sigma\rangle}} - 1\right)|\psi\rangle\, dN, \quad \text{(SSE for direct detection)} \tag{18.18}$$

by keeping terms to second order,

$$d\rho = d\Big(|\psi\rangle\langle\psi|\Big) = \Big(d|\psi\rangle\Big)\langle\psi| + |\psi\rangle\Big(d\langle\psi|\Big) + \Big(d|\psi\rangle\Big)\Big(d\langle\psi|\Big), \tag{18.19}$$

and using $dN^2 = dN$. This SSE is called an **unravelling**[2] of the unconditioned master equation (18.1), since the dissipation in the master equation—which can't be written as a *deterministic* Schrödinger equation—can be written as a *stochastic* Schrödinger equation. Note that Markovian master equations can generally be unravelled in an infinite number of ways; we have only indicated one way so far here. A solution to the SSE is called a **quantum trajectory**,[3] and the solution of the unconditioned master equation represents the average over all trajectories. (Solutions of the SME are also called quantum trajectories.)

18.1.3 Information Gain

Clearly, the observer gains information about the atom when the photodetector clicks, but what is perhaps less obvious is that the observer *also* gains information about the atom even when a photon is *not* detected. We can see this from either the SME (18.10) or the SSE (18.18), since even during times when photons are not detected ($dN(t) = 0$), there are still measurement-related terms (proportional to Γ) that are still "active." Let's look at the evolution in two cases to see how this works. In particular, suppose the atom starts in a pure state, so that we can work with the SSE. Suppose also that the atom is not driven by an external field, so there are no Rabi oscillations. In this case, we can ignore *all* Hamiltonian evolution by trasforming to a suitable rotating frame (where the two levels are degenerate). Then the state vector of the general form

$$|\psi\rangle = c_{\rm g}|{\rm g}\rangle + c_{\rm e}|{\rm e}\rangle \tag{18.20}$$

obeys the equation, *assuming no photon is detected*,

$$\partial_t|\psi\rangle = -\frac{\Gamma}{2}c_{\rm e}|{\rm e}\rangle + \frac{\Gamma}{2}|c_{\rm e}|^2|\psi\rangle, \tag{18.21}$$

which in terms of the coefficients becomes

$$\begin{aligned}\partial_t c_{\rm e} &= -\frac{\Gamma}{2}\left(1 - |c_{\rm e}|^2\right)c_{\rm e}\\ \partial_t c_{\rm g} &= \frac{\Gamma}{2}|c_{\rm e}|^2 c_{\rm g}.\end{aligned} \tag{18.22}$$

Note that unlike the equations of motion for the unconditioned case, these equations are nonlinear in the coefficients due to the measurement process.

Now suppose the atom starts in the superposition state

$$|\psi(t=0)\rangle = \frac{1}{\sqrt{2}}\Big(|{\rm g}\rangle + |{\rm e}\rangle\Big), \tag{18.23}$$

[2]This term was coined by Howard Carmichael, *An Open Systems Approach to Quantum Optics* (Springer, 1993). For earlier work on photon counting statistics, see H. J. Carmichael, Surendra Singh, Reeta Vyas, and P. R. Rice, "Photoelectron waiting times and atomic state reduction in resonance fluorescence," *Physical Review A* **39**, 1200 (1989).

[3]Howard Carmichael, *op. cit.*

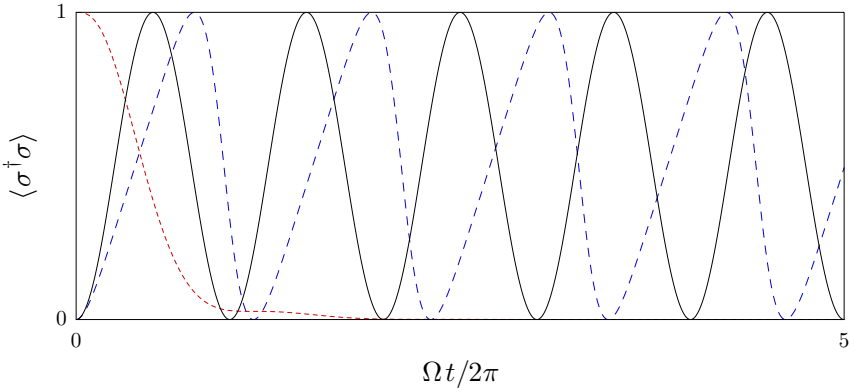
and suppose that no photon is detected until some long time t. Since $|c_e|^2 < 1$, the excited state decays (somewhat more quickly than exponentially) away asymptotically to zero, and the population is thus transferred to the ground state. The interpretation is this: if the observer knows the atom to be in either state with equal probability, and does not see an emitted photon after a very long time (compared to $1/\Gamma$), then the observer concludes that the atom was in fact in the ground state. By *not* observing a photon, the observer still gains information about the state of the atom and "collapses" it to the ground state.

On the other hand, if the initial atomic state is the excited state,

$$|\psi(t=0)\rangle = |e\rangle, \tag{18.24}$$

then the observer knows that at *some* point, the atom *must* decay: the observer knows the atom to be in the excited state with certainty until the photon is detected. This is reflected by the above equations, since if $|c_e|^2 = 1$, the excited state does not decay: $\partial_t c_e = 0$.

The measurement terms have interpretations that are less clear if the atom also undergoes Rabi oscillations. However, the measurement terms must act in the same way in a small time dt, regardless of other processes that influence the atom. The combined evolution of a monitored atom, conditioned on not detecting a photon ($dN = 0$), with Rabi oscillations is illustrated here for the case $\Omega = \Gamma$.



The solid line is what we expect of ordinary, resonant Rabi oscillation without any spontaneous emission ($\Gamma = 0$). The dashed line is the nonlinear, conditioned evolution with $dN = 0$. The Rabi oscillations are severely distorted by the measurement, which tends to cause decay of the excited-state amplitude; thus the rise in excited-state population is slowed, and the fall is accelerated, leading to oscillations closer in shape to a sawtooth wave. The dotted line shows the probability of the atom not having decayed by that time. This is significant since we only see the distorted Rabi oscillations by keeping experiments where the photodetector does not click; any experiments where the detector clicks must be discarded, so that we post-select on the $dN = 0$ cases. Note that this nonlinear evolution is thus difficult to see past the first oscillation, since the nondecay probability becomes very small.

18.1.4 Monte Carlo Trajectories

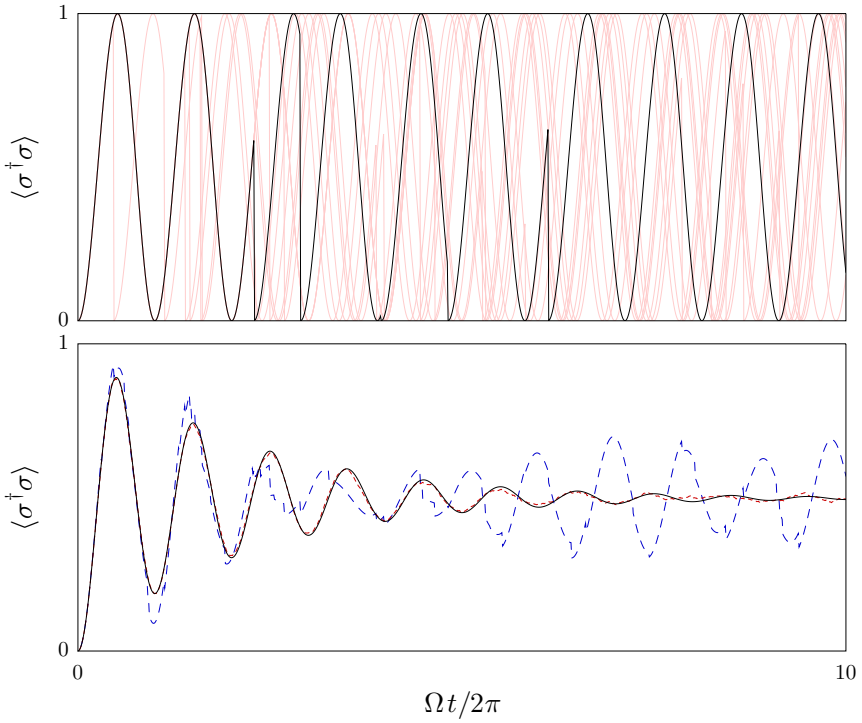
Stochastic Schrödinger equations of this form are popular for simulating master equations,[4] since if the state vector has $O(n)$ components, the density matrix will have $O(n^2)$ components, and thus is much more computationally expensive to solve.[5] If s solutions (quantum trajectories) of the stochastic Schrödinger

[4] R. Dum, P. Zoller, and H. Ritsch, "Monte Carlo simulation of the atomic master equation for spontaneous emission," *Physical Review A* **45** 4879 (1992) (doi: 10.1103/PhysRevA.45.4879); C. W. Gardiner, A. S. Parkins, and P. Zoller, "Wavefunction quantum stochastic differential equations and quantum-jump simulation methods," *Physical Review A* **46** 4363 (1992) (doi: 10.1103/PhysRevA.46.4363); Yvan Castin, Jean Dalibard, and Klaus Mølmer, "A Wave Function Approach to Dissipative Processes," *AIP Conference Proceedings* **275**, 143 (1993) (Thirteenth International Conference on Atomic Physics, ICAP-13, H. Walther, T. W. Hänsch, and B. Neizert, Eds.) (doi: 10.1063/1.43795); P. Marte, R. Dum, R. Taïeb, P. D. Lett, and P. Zoller, "Quantum wave function simulation of the resonance fluorescence spectrum from one-dimensional optical molasses," *Physical Review Letters* **71** 1335 (1993) (doi: 10.1103/PhysRevLett.71.1335); Klaus Mølmer and Yvan Castin, "Monte Carlo wavefunctions in quantum optics," *Quantum and Semiclassical Optics* **8**, 49 (1996) (doi: 10.1088/1355-5111/8/1/007).

[5] If you want to be fastidious, a state that can be represented in terms of n basis states will have a state vector that can be represented by $2n - 2$ real numbers, while the density-matrix representation will require $n^2 - 1$ real numbers. The count for

equation can be averaged together to obtain a sufficiently accurate solution to the master equation and $s \ll n$, then this Monte-Carlo-type method is computationally efficient for solving the master equation. However, note that the average of the simulated ensemble converges slowly, typically as $1/\sqrt{s}$, so this Monte-Carlo method is best used where n is very large (such as when simulating the center-of-mass motion of a quantum system), and very high accuracy is not required.

This idea is illustrated here, where the first plot shows the evolution of the excited-state probability for a single atom (quantum trajectory) with jumps to the ground state, corresponding to a detected photon. Nine other trajectories are included to illustrate the dephasing due to the random nature of the jumps. The usual Rabi oscillations are visible here, since the atom is driven by a classical field of Rabi frequency Ω, but the oscillations are ''interrupted'' by the spontaneous emission events (here with $\Gamma = \Omega/10$).



The second plot shows the ensemble-averaged excited-state probability computed from the unconditioned master equation (solid line), an average of 20 trajectories (dashed line), and an average of 2000 trajectories (dotted line). As many trajectories are averaged together, the average converges to the master-equation solution for the ensemble average. (About 20,000 trajectories are necessary for the Monte-Carlo average to be visually indistinguishable from the master-equation solution on the time scale plotted here.) Note that the ''Rabi oscillations'' apparent here are distorted slightly by the nonlinear renormalization term in Eq. (18.18) from the usual sinusoidal oscillations in the absence of spontaneous emission. However, the damping rate in the above plots is small, so the distortion is not visually apparent. Unravellings of this form are much easier to solve computationally than ''quantum-state diffusion'' unravellings involving dW that we will study later. Of course, it is important for more than just a numerical method, since this gives us a powerful formalism for handling the evolution of a quantum state, accounting for photodetection.

Although quantum trajectories represent a useful simulation method for the master equation, they are interesting in their own right, since they model the measurement process itself and the resulting conditioned dynamics. In fact, some of the original work[6] that motivated quantum trajectories was to understand

$|\psi\rangle$ follows from having n complex numbers, one of which can be discarded due to fixed normalization and an arbitrary overall phase. The count for ρ follows from having n real numbers on the diagonal, and then counting the complex numbers above the diagonal (the elements below the diagonal are redundant). The number of complex diagonal numbers is the sum of all integers from 1 to $(n-1)$, or $n(n-1)/2$. Multiply by 2 to get the real numbers, and then add n to get $n^2 - n + n = n^2$. Remove one real number for $\mathrm{Tr}[\rho] = 1$, and you get $n^2 - 1$.

[6] C. Cohen–Tannoudji and J. Dalibard, ''Single-Atom Laser Spectroscopy. Looking for Dark Periods in Fluorecence Light.''

experiments[7] on **quantum jumps** in vee atoms (a slightly different usage of the term from our usage above), where the fluorescence on a fast transition depends on—blinks on and off to indicate—the state of the atom with respect to the other (slow) transition. The quantum-jump results are understood as "single-shot" (single-trajectory) phenomena, not as ensemble averages.

18.1.5 Detector Efficiency

To handle the case of photodetectors with less than ideal efficiency η, we simply combine the conditioned master equation (18.10) and unconditioned master equation (18.1), with weights η and $1-\eta$, respectively:

$$d\rho = -\frac{i}{\hbar}[H_{\rm A},\rho]dt + \eta\frac{\Gamma}{2}\left[\langle\sigma^\dagger\sigma\rangle - \sigma^\dagger\sigma,\rho\right]_+ dt + (1-\eta)\Gamma\mathcal{D}[\sigma]\rho\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN_\eta. \tag{18.25}$$

(SME for inefficient detection)

The weighting for the counting-process term is absorbed into the counting process itself,

$$\langle\!\langle dN_\eta\rangle\!\rangle = \eta\Gamma\langle\sigma^\dagger\sigma\rangle\,dt, \tag{18.26}$$

(detection probability for inefficient detection)

to account for the fact that fewer photons are detected. We can also write this master equation in the form

$$d\rho = -\frac{i}{\hbar}[H_{\rm A},\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \eta\Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt - \eta\Gamma\sigma\rho\sigma^\dagger\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN_\eta, \tag{18.27}$$

(SME for inefficient detection)

where it is clear from the Lindblad-superoperator term that the *total* disturbance is equivalent to the unconditioned case, and the subsequent terms, whose effects are proportional to η, represent the influence of detecting the fraction η of the photons.

That sounds reasonable, but let's do that more carefully. First, divide dN up into two parts, dN_1 and dN_2, such that

$$\begin{aligned}\langle\!\langle dN_1\rangle\!\rangle &= \eta\Gamma\langle\sigma^\dagger\sigma\rangle\,dt\\ \langle\!\langle dN_2\rangle\!\rangle &= (1-\eta)\Gamma\langle\sigma^\dagger\sigma\rangle\,dt.\end{aligned} \tag{18.28}$$

The sum of two counting processes is still a counting process, and each counting process is characterized by its intensity (mean transition rate) in the same way that a Poisson process is fully characterized by its mean. Thus, we can write

$$dN_1 + dN_2 = dN, \tag{18.29}$$

noting that while this is always true for standard Poisson processes, we are justified in writing this here only because the state dependence enters each of the counting processes in the same way here. We can thus write the SME (18.10) as

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN_1 + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN_2. \tag{18.30}$$

Europhysics Letters **1**, 441 (1986); P. Zoller, M. Marte, and D. F. Walls, "Quantum jumps in atomic system," *Physical Review A* **35**, 198 (1987) (doi: 10.1103/PhysRevA.35.198).

[7] Warren Nagourney, Jon Sandberg, and Hans Dehmelt, "Shelved Optical Electron Amplifier: Observation of Quantum Jumps," *Physical Review Letters* **56**, 2797 (1986) (doi: 10.1103/PhysRevLett.56.2797); Th. Sauter, W. Neuhauser, R. Blatt, and P. E. Toschek, "Observation of Quantum Jumps," *Physical Review Letters* **57**, 1696 (1986) (doi: 10.1103/PhysRevLett.57.1696); J. C. Bergquist, R. G. Hulet, W. M. Itano, and D. J. Wineland, "Observation of Quantum Jumps in a Single Atom," *Physical Review Letters* **57**, 1699 (1986) (doi: 10.1103/PhysRevLett.57.1699); W. M. Itano, J. C. Bergquist, and D. J. Wineland, "Photon Antibunching and Sub-Poissonian Statistics from Quantum Jumps in One and Two Atoms," *Physical Review A* **38**, 559 (1988) (doi: 10.1103/PhysRevA.38.559); R. G. Hulet, D. J. Wineland, J. C. Bergquist, and W. M. Itano, "Precise Test of Quantum Jump Theory," *Phys. Rev. A* **37**, 4544 (1988) (doi: 10.1103/PhysRevA.37.4544) W. M. Itano, D. J. Heinzen, J. J. Bollinger, and D. J. Wineland, "Quantum Zeno effect," *Physical Review A* **41**, 2295 (1990) (doi: 10.1103/PhysRevA.41.2295); D. J. Berkeland, D. A. Raymondson, and V. M. Tassin, "Tests for non-randomness in quantum jumps," *Physical Review A* **69**, 052103 (2004) (doi: 10.1103/PhysRevA.69.052103).

If we detect photons with efficiency η, then *not* detecting a fraction $1-\eta$ of the photons is equivalent to taking an ensemble average over dN_2, because we are discarding the information provided by dN_2. Taking this average, we let double angle brackets denote this ensemble average with respect to dN_2, and again note that $\rho(t)$ and $dN_2(t)$ are statistically independent:

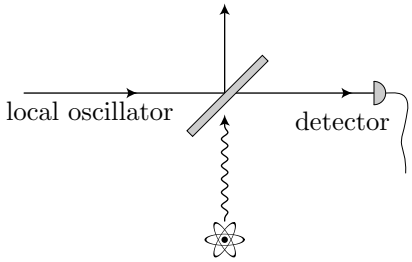
$$\begin{aligned} d\langle\!\langle\rho\rangle\!\rangle &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \Gamma\langle\sigma^\dagger\sigma\rangle\langle\!\langle\rho\rangle\!\rangle dt + \left(\frac{\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \langle\!\langle\rho\rangle\!\rangle\right)dN_1 + \left(\frac{\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \langle\!\langle\rho\rangle\!\rangle\right)\langle\!\langle dN_2\rangle\!\rangle \\ &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \Gamma\langle\sigma^\dagger\sigma\rangle\langle\!\langle\rho\rangle\!\rangle dt + \left(\frac{\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \langle\!\langle\rho\rangle\!\rangle\right)dN_1 \\ &\qquad + (1-\eta)\Gamma\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger\,dt - (1-\eta)\Gamma\langle\sigma^\dagger\sigma\rangle\langle\!\langle\rho\rangle\!\rangle dt \\ &= -\frac{i}{\hbar}[H,\langle\!\langle\rho\rangle\!\rangle]dt - \eta\frac{\Gamma}{2}[\sigma^\dagger\sigma,\langle\!\langle\rho\rangle\!\rangle]_+dt + \eta\Gamma\langle\sigma^\dagger\sigma\rangle\langle\!\langle\rho\rangle\!\rangle dt + (1-\eta)\Gamma\mathcal{D}[\sigma]\langle\!\langle\rho\rangle\!\rangle + \left(\frac{\sigma\langle\!\langle\rho\rangle\!\rangle\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \langle\!\langle\rho\rangle\!\rangle\right)dN_1. \end{aligned} \tag{18.31}$$

This is equivalent to Eqs. (18.25) if we understand ρ to be the ensemble average $\langle\!\langle\rho\rangle\!\rangle$ and we relabel $dN_1 \longrightarrow dN_\eta$.

In the case $\eta < 1$, it is not possible to unravel the SME into an equivalent SSE. It is only in the case $\eta = 1$ that this is possible, because only in this case does an initially pure state remain pure under the master-equation evolution—otherwise, the observer must trace over all possibilities for the undetected photons and thus necessarily ends up with a mixed state. Of course, the resulting density operator can be computed by simulating many trajectories of the SSE, and then computing the appropriate (partial) ensemble average.

18.2 Homodyne Detection

Now we will see how we can get a completely different unravelling of the master equation by modifying the atomic field measurement. In particular, we will obtain a white-noise limit of a jump process in the master equation. To set up this measurement, suppose that all of the light radiated from an atom is collimated ''somehow'' (e.g., by exotic ''4π'' optics) into a directed beam. An alternative to simply feeding it into a photon-counting detector is to mix it on a beam splitter with a **local oscillator**. What exactly we mean by a local oscillator is a monochromatic field with an intensity much larger than the atomic field, but otherwise is somewhat context-dependent: if the atom is driven by a monochromatic field, then we would use part of the driving field as the local oscillator, while if the atom is undriven, we would simply use a laser field at the atomic resonance frequency. The point is, the atom will naturally be oscillating with a spectrum centered at *some* frequency, and we would choose the local oscillator to have the same frequency. Once the radiation field and local oscillator are mixed together by the beam splitter, the light is then detected. This measurement is called **homodyne detection**,[8] and here we will consider the simplest case of homodyne detection, where only one output of the beam splitter is monitored.



Even though this setup detects some ''irrelevant'' light due to the local oscillator, the amount of information that we acquire via this setup (within the idealizations below) is equivalent to that of direct detection. If the

[8]Homodyne detection was treated in terms of quantum trajectories first by Howard Carmichael, *op. cit.* Our treatment here more closely follows that of Howard Mark Wiseman, *Quantum Trajectories and Feedback*, Ph.D. thesis (University of Queensland, 1994), and H. M. Wiseman and G. J. Milburn, ''Quantum theory of field-quadrature measurements,'' *Physical Review A* **47**, 642 (1993) (doi: 10.1103/PhysRevA.47.642).

radiated field from the atom is detected directly, the operator that we associate with the field is proportional to

$$C = \sqrt{\Gamma}\sigma, \tag{18.32}$$

so that the average number of photons detected in a time interval dt is

$$\langle C^\dagger C\rangle\, dt = \Gamma\langle\sigma^\dagger\sigma\rangle\, dt. \tag{18.33}$$

We can similarly associate the operator

$$C_{\rm loc} = \sqrt{\Gamma}a, \tag{18.34}$$

with the local-oscillator field, where a is the annihilation operator for the local-oscillator field mode (which is comparable to the atomic lowering operator σ—recall from Section 5.7 that the dipole radiation field is written in terms of the atomic dipole operator σ). The normalization here assures that photons in either the dipole or local-oscillator fields are detected in the same way by the detector. The combined field operator, after the beam splitter (on the detector side), is then

$$C_r = \sqrt{\Gamma}\left(r\sigma + \sqrt{1-r^2}\,a\right), \tag{18.35}$$

where $r \in [0,1]$ is the field reflection coefficient of the beam splitter (as seen by the atomic field). Here, we are assuming that the fields of the collimated atomic radiation and the local oscillator are perfectly ''mode matched.''

We will model the local oscillator as a coherent state $|\alpha\rangle$ of light—the quantum model for a classical, monochromatic field—with photon flux $\Gamma|\alpha|^2$ (whose detection is also described by a counting process). Recalling that $|\alpha\rangle$ is an eigenstate of the field annihilation operator a, we can write $a|\alpha\rangle = \alpha|\alpha\rangle$, and thus the operator for the total field at the detector effectively becomes

$$C_r = \sqrt{\Gamma}\left(r\sigma + \sqrt{1-r^2}\,\alpha\right) \tag{18.36}$$

whenever applied to the field state. Note also that α is in general a complex number, representing the phase of the local oscillator field.

In this homodyne scheme, any atomic radiation that transmits through the beam splitter is not detected, and the information associated with that light is ''wasted.'' Thus we will want to consider the limit $r \longrightarrow 1$, but we will then also take the limit $|\alpha| \longrightarrow \infty$ such that the transmitted field amplitude

$$\beta := \alpha\sqrt{1-r^2} \tag{18.37}$$

remains nonzero. (In fact we will eventually take the limit $\beta \longrightarrow \infty$.) Thus, the detected field operator is

$$C_\beta = \sqrt{\Gamma}\,(\sigma + \beta)\,, \tag{18.38}$$

and so we see that the effect of adding the local-oscillator field is to add a scalar constant to the atomic lowering operator. The average photodetection rate due to the combined field is

$$\langle\langle dN\rangle\rangle = \left\langle C_\beta^\dagger C_\beta\right\rangle dt = \Gamma\left\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\right\rangle dt = \Gamma\left[\left\langle\sigma^\dagger\sigma\right\rangle + \left\langle\beta^*\sigma + \beta\sigma^\dagger\right\rangle + |\beta|^2\right]dt. \tag{18.39}$$

As we expect, the photodetection rate is the sum of the individual photodetection rates for the local oscillator and for the atomic radiation, plus an interference term. Herein lies the advantage of homodyne detection: the atomic signal due to the interference terms is effect ''boosted'' by the local oscillator by a factor of $|\beta|$. This is an enormous advantage if the detector suffers from a low level of background noise (''dark currents''), since the homodyne scheme can raise the signal to a level much larger than background noise. We will also see that homodyne detection measures *different aspects* of the atom, as compared with direct detection.

18.2.1 State Collapse

When a photon is detected directly from the atom, recall that the state vector is reduced according to

$$|\psi\rangle \longrightarrow \frac{C|\psi\rangle}{\sqrt{\langle C^\dagger C\rangle}} = \frac{\sigma|\psi\rangle}{\sqrt{\langle\sigma^\dagger\sigma\rangle}}, \tag{18.40}$$

or equivalently, the density operator is reduced according to

$$\rho \longrightarrow \frac{C\rho C^\dagger}{\mathrm{Tr}[C\rho C^\dagger]} = \frac{C\rho C^\dagger}{\langle C^\dagger C\rangle} = \frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle}. \tag{18.41}$$

The *same* reduction occurs when a homodyne photon is detected, but now using the operator C_β, for the state vector,

$$|\psi\rangle \longrightarrow \frac{C_\beta|\psi\rangle}{\sqrt{\left\langle C_\beta^\dagger C_\beta\right\rangle}} = \frac{(\sigma+\beta)|\psi\rangle}{\sqrt{\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}}, \tag{18.42}$$

and of course the density operator,

$$\rho \longrightarrow \frac{C_\beta\rho C_\beta^\dagger}{\left\langle C_\beta^\dagger C_\beta\right\rangle} = \frac{(\sigma+\beta)\rho(\sigma^\dagger+\beta^*)}{\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}. \tag{18.43}$$

The reduction here is partial: when a photon is detected, it is not possible to distinguish whether the photon came from the atom or the local oscillator, and thus the atom is projected into a *coherent superposition* of being reduced as in the direct-detection case (if the photon came from the atom) and of being unaffected (if the photon came from the local oscillator). The two cases are weighted by the amplitudes of the atomic and local oscillator fields, respectively.

18.2.2 Quantum-State Diffusion

Evidently, we obtain the master equation for homodyne detection from the master equation for direct detection by the replacement

$$\sigma \longrightarrow \sigma + \beta. \tag{18.44}$$

However, note that adding the local oscillator field *cannot* change the form of the *unconditioned* master equation

$$\partial_t\rho = -\frac{i}{\hbar}[H,\rho] + \Gamma\mathcal{D}[\sigma]\rho, \tag{18.45}$$

since in the unconditioned case the local oscillator influences quantum information that we are discarding anyway. Noting that

$$\begin{aligned}\mathcal{D}[\sigma+\beta]\rho &= (\sigma+\beta)\rho(\sigma^\dagger+\beta^*) - \frac{1}{2}\left[(\sigma^\dagger+\beta^*)(\sigma+\beta),\rho\right]_+ \\ &= \mathcal{D}[\sigma]\rho + \frac{1}{2}\left[\beta^*\sigma - \beta\sigma^\dagger,\rho\right] \\ &= \mathcal{D}[\sigma]\rho - \frac{i}{\hbar}\left[\frac{i\hbar}{2}\left(\beta^*\sigma-\beta\sigma^\dagger\right),\rho\right].\end{aligned} \tag{18.46}$$

Thus, the unconditioned master equation is invariant under the *simultaneous* replacement

$$\sigma \longrightarrow \sigma+\beta, \qquad H \longrightarrow H - \frac{i\hbar\Gamma}{2}\left(\beta^*\sigma-\beta\sigma^\dagger\right). \tag{18.47}$$

(transformation from jump to homodyne detection)

Thus, this is the total transformation we should make to go from direct to homodyne detection.

Recalling from Eq. (18.10) that the master equation for direct detection is

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \mathcal{J}[\sigma]\rho\,dN, \tag{18.48}$$

where we have defined the jump superoperator

$$\mathcal{J}[c]\rho := \left(\frac{c\rho c^\dagger}{\mathrm{Tr}[c\rho c^\dagger]} - \rho\right) = \left(\frac{c\rho c^\dagger}{\langle c^\dagger c\rangle} - \rho\right). \tag{18.49}$$
(jump superoperator)

Thus, the master equation for homodyne detection is

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\left[\left(\beta^*\sigma - \beta\sigma^\dagger\right),\rho\right]dt - \frac{\Gamma}{2}[(\sigma^\dagger+\beta^*)(\sigma+\beta),\rho]_+dt + \Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle\,\rho\,dt + \mathcal{J}[\sigma+\beta]\rho\,dN. \tag{18.50}$$

Expanding out the middle three terms, we see that the $|\beta|^2$ terms cancel, so that

$$\begin{aligned} d\rho &= -\frac{i}{\hbar}[H,\rho]dt - \Gamma\mathcal{H}[\beta^*\sigma]\rho\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \mathcal{J}[\sigma+\beta]\rho\,dN \\ &= -\frac{i}{\hbar}[H,\rho]dt - \Gamma\mathcal{H}[\beta^*\sigma]\rho\,dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\sigma+\beta]\rho\,dN, \end{aligned} \tag{18.51}$$

where we have defined the measurement superoperator

$$\begin{aligned} \mathcal{H}[c]\rho &:= c\rho + \rho c^\dagger - \mathrm{Tr}[c\rho + \rho c^\dagger]\rho \\ &= c\rho + \rho c^\dagger - \langle c + c^\dagger\rangle\,\rho. \end{aligned} \tag{18.52}$$
(homodyne superoperator)

The transition rate of the counting process is correspondingly modified to read

$$\langle\!\langle dN\rangle\!\rangle = \Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle\,dt, \tag{18.53}$$

which matches the photodetection rate we computed above for the homodyne case.

Now we will consider the limit $|\beta| \longrightarrow \infty$ of a strong local oscillator, so that most of the detected photons come from the local oscillator. The rate of information gain—and thus the rate at which we disturb the system—remains constant, but the rate at which photons are detected becomes arbitrarily large. Thus, the white-noise approximation for the counting process $N(t)$ is appropriate (here the approximation has same form as for the Poisson process), and we can make the replacement

$$dN \longrightarrow \left\langle\!\!\left\langle \frac{dN}{dt} \right\rangle\!\!\right\rangle dt + \sqrt{\left\langle\!\!\left\langle \frac{dN}{dt} \right\rangle\!\!\right\rangle}\,dW, \tag{18.54}$$

which in the homodyne case here becomes

$$dN \longrightarrow \Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle\,dt + \sqrt{\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}\,dW \tag{18.55}$$

to obtain an Itō SDE for the state evolution. First, let's work out the part proportional to dt:

$$\begin{aligned} \mathcal{J}[\sigma+\beta]\rho\langle\!\langle dN\rangle\!\rangle &= \left(\frac{(\sigma+\beta)\rho(\sigma^\dagger+\beta^*)}{\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle} - \rho\right)\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle\,dt \\ &= \Gamma(\sigma+\beta)\rho(\sigma^\dagger+\beta^*)\,dt - \Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle\,\rho\,dt \\ &= \Gamma\sigma\rho\sigma^\dagger\,dt - \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \Gamma\left(\beta^*\sigma\rho + \rho\beta\sigma^\dagger\right)dt - \Gamma\langle\beta^*\sigma + \beta\sigma^\dagger\rangle\,\rho\,dt \\ &= \Gamma\langle\sigma^\dagger\sigma\rangle\,\mathcal{J}[\sigma]\rho\,dt + \Gamma\mathcal{H}[\beta^*\sigma]\rho\,dt. \end{aligned} \tag{18.56}$$

Notice that we have dropped the ensemble-average symbols on the right-hand side, since the ensemble average is only for convenience of notation. Thus, the $\mathcal{H}[\beta^*\sigma]\rho$ terms cancel in the master equation (18.51), which thus becomes

$$\begin{aligned} d\rho &= -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\mathcal{J}[\sigma]\rho\,dt + \sqrt{\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}\,\mathcal{J}[\sigma+\beta]\rho\,dW \\ &= -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \sqrt{\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}\,\mathcal{J}[\sigma+\beta]\rho\,dW. \end{aligned} \tag{18.57}$$

In the part proportional to dt, we see that all the β-dependent parts cancel, so taking the limit of large $|\beta|$ is trivial. We can now expand out the part proportional to dW, keeping only the lowest-order terms in $|\beta|^{-1}$:

$$\begin{aligned} \sqrt{\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle}\,\mathcal{J}[\sigma+\beta]\rho &= \left(\frac{(\sigma+\beta)\rho(\sigma^\dagger+\beta^*)}{\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle} - \rho\right)\sqrt{\Gamma\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\rangle} \\ &= \left(\frac{|\beta|^2\rho + (\beta^*\sigma\rho + \rho\beta\sigma^\dagger) - \langle\beta^*\sigma+\beta\sigma^\dagger\rangle\,\rho}{|\beta|^2} - \rho\right)\sqrt{\Gamma|\beta|^2} \\ &= \sqrt{\Gamma}\frac{\left[(\beta^*\sigma\rho + \rho\beta\sigma^\dagger) - \langle\beta^*\sigma+\beta\sigma^\dagger\rangle\,\rho\right]}{|\beta|} \\ &= \sqrt{\Gamma}\mathcal{H}\left[\sigma\frac{\beta^*}{|\beta|}\right]\rho \\ &= \sqrt{\Gamma}\mathcal{H}\left[\sigma e^{i\phi}\right]\rho. \end{aligned} \tag{18.58}$$

Here, ϕ is the phase of the local oscillator, defined by

$$\beta =: |\beta|e^{-i\phi}. \tag{18.59}$$
(local-oscillator phase)

Thus, the master equation becomes

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}\left[\sigma e^{i\phi}\right]\rho\,dt + \sqrt{\Gamma}\mathcal{H}\left[\sigma e^{i\phi}\right]\rho\,dW,$$
(SME for homodyne detection) (18.60)

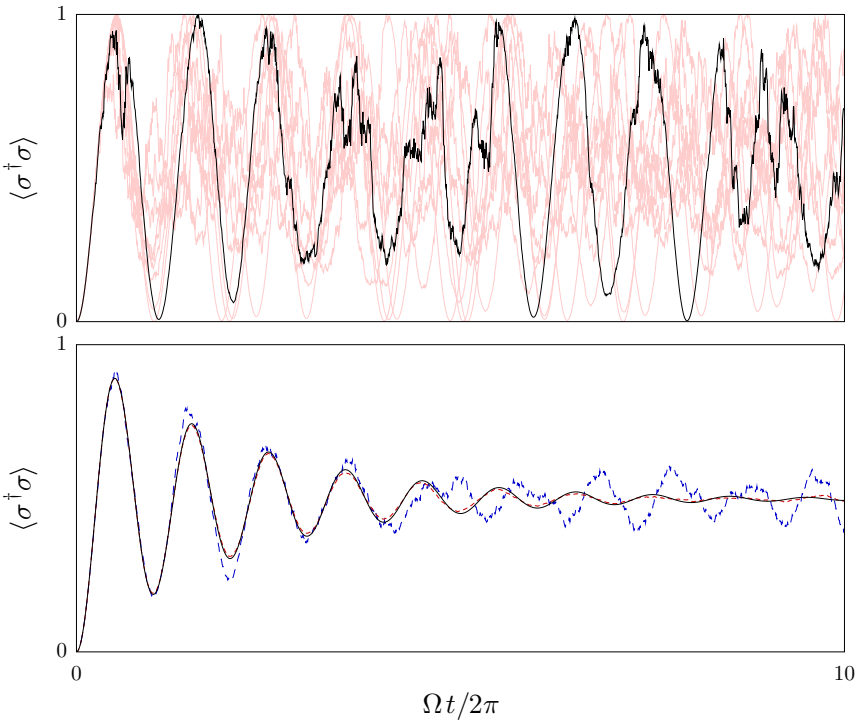
where we note that due to the quadratic nature of the Lindblad superoperator, we have used

$$\mathcal{D}[\sigma]\rho = \mathcal{D}\left[\sigma e^{i\phi}\right]\rho. \tag{18.61}$$

The form (18.60) for the stochastic master equation is our main result here: adding the local oscillator field prior to detection completely changes the form of the SME. The form here, in terms of the Wiener process $dW(t)$, is called the **quantum-state diffusion** form[9] for the SME (as opposed to the quantum-jump form). In this form it is easy to see that the ensemble average recovers the unconditioned master equation (18.1), since the ensemble average in Itō calculus amounts to setting $dW = 0$.

[9] N. Gisin and I. C. Percival, "The quantum-state diffusion model applied to open systems," *Journal of Physics A: Mathematical and General* **25**, 5677 (1992) (doi: 10.1088/0305-4470/25/21/023). For other related early work, see A. Barchielli, L. Lanz, and G. M. Prosperi, "A Model for the Macroscopic Description and Continual Observations in Quantum Mechanics," *Nuovo Cimento B* **72**, 79 (1982); A. Barchielli and G. Lupieri, "Quantum stochastic calculus, operation valued stochastic processes, and continual measurements in quantum mechanics," *Journal of Mathematical Physics* **26**, 2222 (1985) (doi: 10.1063/1.526851); V. P. Belavkin, in *Information Complexity and Control in Quantum Physics*, A. Blaquiere, S. Diner, and G. Lochak, Eds. (Springer, 1987); V. P. Belavkin, "A new wave equation for a continuous nondemolition measurement," *Physics Letters A* **140**, 355 (1989) (doi: 10.1016/0375-9601(89)90066-2); V. P. Belavkin, "A posterior Schrödinger equation for continuous nondemolition measurement," *Journal of Mathematical Physics* **31**, 2930 (1990) (doi: 10.1063/1.528946); L. Diósi, "Stochastic Pure State Representation for Open Quantum Systems," *Physics Letters A* **114**, 451 (1986) (doi: 10.1016/0375-9601(86)90692-4); and L. Diósi, "Continuous Quantum Measurement and Itô Formalism," *Physics Letters A* **129**, 419 (1988) (doi: 10.1016/0375-9601(88)90309-X).

These trajectories are illustrated here, for the same paramters as for the quantum-jump unravelling above ($\Gamma = \Omega/10$). The first plot shows the evolution of the excited-state probability for a single atom (quantum trajectory) with quantum-state diffusion, with each infinitesimal jump corresponding to a detected photon in the homodyne setup. Nine other trajectories are included to illustrate the dephasing due to the stochastic nature of the evolution, and the qualitative difference with respect to the quantum-jump evolution. The usual Rabi oscillations are still visible here, since the atom is driven by a classical field of Rabi frequency Ω, the the oscillations are distorted by influence of the quantum noise. The influence is visually greatest when the excited-state population is greatest, which makes sense intuitively: if the atom is in the ground state, then the observer knows that any detected photon is due only to the local oscillator.



The second plot shows the ensemble-averaged excited-state probability computed from the unconditioned master equation (solid line), an average of 20 trajectories (dashed line), and an average of 2000 trajectories (dotted line). As many trajectories are averaged together, the average converges to the master-equation solution for the ensemble average. (About 20,000 trajectories are necessary for the Monte-Carlo average to be visually indistinguishable from the master-equation solution on the time scale plotted here.) In all cases, the trajectories are plotted in discrete time increments of $\Delta t = 0.005$, but the trajectories were calculated using increments of $\Delta t = 0.0025$ (using a stochastic Runge-Kutta-type numerical method of order 1.5).

18.2.3 Measurement Record

Now we can ask, what exactly are we measuring here? Recall from Eq. (18.11) that the photodetector current is

$$I_{\rm det}(t) = Q_{\rm ph}\frac{dN(t)}{dt}, \tag{18.62}$$

where we are in the limit where dN is given by the replacement (18.55). Thus,

$$I_{\rm det}(t) = Q_{\rm ph}\Gamma\big\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\big\rangle + Q_{\rm ph}\sqrt{\Gamma\big\langle(\sigma^\dagger+\beta^*)(\sigma+\beta)\big\rangle}\,\xi(t), \tag{18.63}$$

where again $\xi(t) = dW(t)/dt$, and the second term represents the shot noise (Section 17.5.3.1) associated with the detector photocurrent. In general, we will want to subtract off the large constant photocurrent due

to the local oscillator field, and then retain only the lowest-order terms in $|\beta|^{-1}$ in the dt and dW parts:

$$I_{\rm det}(t) - Q_{\rm ph}\Gamma|\beta|^2 = Q_{\rm ph}\Gamma\langle\beta^*\sigma + \beta\sigma^\dagger\rangle + Q_{\rm ph}\sqrt{\Gamma|\beta|^2}\,\xi(t). \tag{18.64}$$

We can then define the normalized photocurrent by

$$\tilde{I}_{\rm det}(t) := \frac{I_{\rm det}(t) - Q_{\rm ph}\Gamma|\beta|^2}{Q_{\rm ph}|\beta|} = \Gamma\langle\sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle + \sqrt{\Gamma}\,\xi(t).$$

(normalized photocurrent) (18.65)

In the standard form of an Itō SDE, we can write

$$dr(t) := \tilde{I}_{\rm det}(t)\,dt = \Gamma\langle\sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle\,dt + \sqrt{\Gamma}\,dW.$$

(homodyne measurement record) (18.66)

We can regard $dr(t)$ as the (scaled) measurement record for homodyne detection, with $r(t)$ proportional to the total accumulated charge conducted by the photodetector (total photon count).

We thus see that on average, we gain information about either the real or imaginary part of the radiated field, or equivalently σ, depending on the local-oscillator phase:

$$\langle\!\langle dr(t)\rangle\!\rangle = \Gamma\langle\sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle\,dt. \tag{18.67}$$

For example, if we choose $\phi = 0$ (such that $\beta \in \mathbb{R}$), we measure the "X_1 quadrature"

$$\langle\sigma + \sigma^\dagger\rangle = \langle\sigma_x\rangle, \tag{18.68}$$

whereas if we choose $\phi = \pi/2$ (such that $-i\beta \in \mathbb{R}$), we measure the "X_2 quadrature"

$$\langle i\sigma - i\sigma^\dagger\rangle = \langle\sigma_y\rangle. \tag{18.69}$$

In either case, clearly the homodyne measurement provides information about the mean atomic dipole moment, or the phase of the atomic dipole, as opposed to the direct measurement, which was more closely related to the atomic excitation $\langle\sigma_z\rangle$. Note that with the proper choice of local-oscillator frequency, the dipole field and local oscillator time dependences cancel, these expectation values are measured in the rotating frame of the local oscillator. Thus, the expectation values are (adiabatic) constants. Of course, due to the dW term in the measurement record (18.66), the information in the mean is masked by quantum noise, and so the information must be extracted, e.g., by signal averaging. Of course, the best (i.e., correct in the Bayesian sense) method for obtaining information about the system is to evolve the master equation, *conditioned* on the measurement record. Since the dW in (18.66) is the same as the dW in the SME (18.60), we can eliminate it and write the SME directly in terms of the measurement record:

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}\big[\sigma e^{i\phi}\big]\rho\,dt + \sqrt{\Gamma}\mathcal{H}\big[\sigma e^{i\phi}\big]\rho\left(\frac{dr(t) - \Gamma\langle\sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle\,dt}{\sqrt{\Gamma}}\right).$$

(homodyne SME with measurement signal) (18.70)

This equation is still an Itō SDE, even though dW does not explicitly appear here.

18.2.4 Information Gain from the Measurement Record

Now let's make a statement that is somewhat subtle, but nevertheless helps to gain insight here. Suppose that you and someone else have access to the same measurement record $dr(t)$, which begins at $t = 0$, but for some reason the two of you *disagree* on the initial state $\rho(t = 0)$ of the system (i.e., the Bayesian "prior"). That is, you two have *different density operators* at $t = 0$. We are thus regarding the density operator as being *subjective* information about the quantum state. We will return to this topic in depth later, but let's just go with it for now. For technicalities of consistency, we will assume that both you and the other observer either *both* assign a particular measurement outcome (based on the density operator) to consistenly have

either zero or nonzero probability (i.e., in any basis, both observers agree on whether or not any particular diagonal density-matrix element is zero). In this view, either state is as good as any other, so let's view the measurement record as being related to the expectation value of the *other* observer B:

$$dr(t) = \Gamma\langle \sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle_B \, dt + \sqrt{\Gamma}\, dW'. \tag{18.71}$$

Here, the B subscript denotes that the expectation value is taken with respect to the density operator ρ_B of the second observer. Clearly, dW' is not the same as the original dW if the expectation values differ with respect to the two quantum states. Then, we may rewrite the SME (18.70) as

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\, dt + \sqrt{\Gamma}\mathcal{H}[\sigma e^{i\phi}]\rho\left[\sqrt{\Gamma}\Big(\langle \sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle_B - \langle \sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle\Big)dt + dW'\right]. \tag{18.72}$$

Notice that the other expectation value here is taken with respect to your (observer A's) state ρ. Thus, we see that the measurement term is sensitive to the *difference* between the estimates of the quantity

$$\langle \sigma e^{i\phi} + \sigma^\dagger e^{-i\phi}\rangle \tag{18.73}$$

according to you and to observer B. In fact, this difference will be *suppressed* during the evolution, so that as you and observer B gain more and more consistent information, the difference in your density operators will tend to vanish at long times: in light of new information, the density operator ''forgets'' its initial condition. It is also possible to insist on an *objective* view of this same situation, where your initial density operator is ''wrong,'' and observer B is an omniscient observer that is ''right.'' In this case, with more measurement information, your density operator will tend to converge to the ''true'' density operator. This is sometimes a useful way to think about things, but you must be careful since it will also get you into trouble (e.g., it leads to problems associated with wave-function collapse being a physical process). Note that the convergence of states here must be viewed with caution, since complete convergence can only occur if the measurement information in fact resolves the differences between the different states. For example, if you and I disagree about the uncertainty in the X_1 quadrature, but the measurement only provides information about the X_2 quadrature, then clearly our states need not converge to the same state as a result of the continuous measurement: the measurement won't affect the X_1 uncertainty, so we won't agree on that aspect of the state.

The convergence of states is in some sense more intuitive for the jump process *without* the local oscillator. Recall that a detected photon lowers the atom to the ground state $|\mathrm{g}\rangle$. Thus, with this measurement, *any* initial state is mapped to the same final state via a single jump, and so any observers must subsequently agree on the final state.

18.2.5 Diffusion Form of the Stochastic Schrödinger Equation

The SME (18.60) is equivalent to the SSE

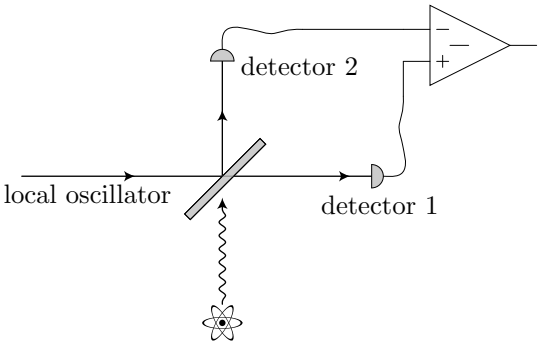
$$d|\psi\rangle = -\frac{i}{\hbar}H|\psi\rangle\, dt - \frac{\Gamma}{2}\left[\sigma^\dagger\sigma - \langle\sigma+\sigma^\dagger\rangle\,\sigma + \frac{1}{4}\langle\sigma+\sigma^\dagger\rangle^2\right]|\psi\rangle\, dt + \sqrt{\Gamma}\left[\sigma - \frac{1}{2}\langle\sigma+\sigma^\dagger\rangle\right]|\psi\rangle\, dW, \tag{18.74}$$

(SSE for homodyne detection)

as we can see again by expanding $d\rho$ to second order in $d|\psi\rangle$ and using the Itō rule $dW^2 = dt$. Again, this ''diffusion unravelling'' of the SME is only valid for unit detection efficiency. Otherwise, we must use a modified SME, which we will derive below.

18.2.6 Balanced Homodyne Detection

The homodyne technique above, while relatively easy to analyze, has some practical disadvantages. First is the large dc offset due to the local oscillator that must be subtracted to obtain the desired signal. Second is the beam splitter, which must have close to unit reflection, and the corresponding requirement that the local oscillator field be very strong. Both of these problems are solved in practice by **balanced homodyne detection**, which involves a 50/50 beam splitter, detecting *both* output ports of the beam splitter, and then subtracting the two photocurrents.



Suppose that the beam splitter is lossless with reflection coefficient $r \in [0,1]$. Then proceeding in the same way as in the last section, the operator associated with the field impinging on detector 1 is

$$C_1 = \sqrt{\Gamma}\left(r\sigma + \sqrt{1-r^2}\,\alpha\right), \tag{18.75}$$

where $t = \sqrt{1-r^2}$ is the transmission coefficient of the (lossless) beam splitter, and again α is the amplitude of the coherent state $|\alpha\rangle$ of the local-oscillator field. The operator associated with the field impinging on detector 2 is

$$C_2 = \sqrt{\Gamma}\left(\sqrt{1-r^2}\,\sigma - r\alpha\right), \tag{18.76}$$

which follows from the Stokes relation $r' = -r$ relating the reflection coefficient of the beam splitter from the two sides, or alternately that we may regard the beam splitter as inducing a *unitary* transformation on the input fields of the form

$$\begin{bmatrix} t & r \\ -r & t \end{bmatrix}, \tag{18.77}$$

with $r^2 + t^2 = 1$, which in this case is an *orthogonal* transformation since we have assumed the coefficients to be real.

We can now associate two counting processes dN_1 and dN_2 with detectors 1 and 2, respectively, to account for ''clicks'' on each detector. The average photodetection rate of detector 1 is

$$\begin{aligned} \langle\langle dN_1 \rangle\rangle &= \left\langle C_1^\dagger C_1 \right\rangle dt \\ &= \Gamma\left[r^2\langle\sigma^\dagger\sigma\rangle + r\sqrt{1-r^2}\langle\alpha^*\sigma + \alpha\sigma^\dagger\rangle + \left(1-r^2\right)|\alpha|^2\right] dt, \end{aligned} \tag{18.78}$$

while the mean photodetection rate of detector 2 is

$$\begin{aligned} \langle\langle dN_2 \rangle\rangle &= \left\langle C_2^\dagger C_2 \right\rangle dt \\ &= \Gamma\left[\left(1-r^2\right)\langle\sigma^\dagger\sigma\rangle - r\sqrt{1-r^2}\langle\alpha^*\sigma + \alpha\sigma^\dagger\rangle + r^2|\alpha|^2\right] dt. \end{aligned} \tag{18.79}$$

Again, the contributions of the two fields, as well as the interference effects, are apparent in these expressions.

Now the two detectors generate photocurrents as before according to

$$\begin{aligned} I_{\mathrm{det},1} &= Q_{\mathrm{ph},1}\frac{dN_1(t)}{dt} \\ I_{\mathrm{det},2} &= Q_{\mathrm{ph},2}\frac{dN_2(t)}{dt}, \end{aligned} \tag{18.80}$$

where for the moment, we will assume the detectors have generally different ''gains'' $Q_{\mathrm{det},1}$ and $Q_{\mathrm{det},2}$ per photon. The subtracted photocurrent is then

$$\begin{aligned} I_- &= I_{\mathrm{det},1} - I_{\mathrm{det},2} \\ &= Q_{\mathrm{ph},1}\frac{dN_1(t)}{dt} - Q_{\mathrm{ph},2}\frac{dN_2(t)}{dt}, \end{aligned} \tag{18.81}$$

where the mean is given by

$$\begin{aligned}\langle\langle I_- \rangle\rangle = \Gamma\Big[&\left[Q_{\mathrm{ph},1} r^2 - Q_{\mathrm{ph},2}\left(1-r^2\right)\right]\langle\sigma^\dagger\sigma\rangle \\ &+ \left(Q_{\mathrm{ph},1} + Q_{\mathrm{ph},2}\right) r\sqrt{1-r^2}\langle\alpha^*\sigma + \alpha\sigma^\dagger\rangle \\ &+ \left[Q_{\mathrm{ph},1}\left(1-r^2\right) - Q_{\mathrm{ph},2} r^2\right]|\alpha|^2\Big].\end{aligned} \tag{18.82}$$

Note that with the condition

$$Q_{\mathrm{ph},1}\left(1-r^2\right) = Q_{\mathrm{ph},2} r^2, \tag{18.83}$$

the $|\alpha|^2$ term vanishes; that is, the gains of the photodetectors can be adjusted to null out the large dc term. However, the small $\langle\sigma^\dagger\sigma\rangle$ term vanishes only if

$$Q_{\mathrm{ph},1} r^2 = Q_{\mathrm{ph},2}\left(1-r^2\right). \tag{18.84}$$

The only way to satisfy both conditions is to take a completely symmetric, or *balanced*, setup with $Q_{\mathrm{ph},1} = Q_{\mathrm{ph},2} = Q_{\mathrm{ph}}$ and $r^2 = 1/2$. In this case, the mean subtracted photocurrent takes on the simpler form

$$\langle\langle I_- \rangle\rangle = Q_{\mathrm{ph}}\Gamma\langle\alpha^*\sigma + \alpha\sigma^\dagger\rangle, \tag{18.85}$$

and thus in the balanced setup, only the interference terms contribute without further subtraction or approximation.

18.2.6.1 Master Equation

Now we will derive the master equation for balanced homodyne detection, and see that in the limit of a strong local oscillator, the result is the same as for simple homodyne detection. Recalling again from Eq. (18.10) that the master equation for direct detection is

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\sigma]\rho\,dN, \tag{18.86}$$

where $\langle\langle dN\rangle\rangle = \Gamma\langle\sigma^\dagger\sigma\rangle\,dt$, and the relevant superoperators are once again

$$\begin{aligned}\mathcal{J}[c]\rho &:= \left(\frac{c\rho c^\dagger}{\langle c^\dagger c\rangle} - \rho\right) \\ \mathcal{H}[c]\rho &:= c\rho + \rho c^\dagger + \langle c + c^\dagger\rangle\,\rho.\end{aligned} \tag{18.87}$$

We can decompose the counting process dN into two parts dN_1 and dN_2, as we did in Section 18.1.5, and correspondingly split the other damping term in the master equation, so that the split counting processes are determined by

$$\begin{aligned}\langle\langle dN_1\rangle\rangle &= r^2\Gamma\langle\sigma^\dagger\sigma\rangle\,dt \\ \langle\langle dN_2\rangle\rangle &= \left(1-r^2\right)\Gamma\langle\sigma^\dagger\sigma\rangle\,dt,\end{aligned} \tag{18.88}$$

and the master equation is

$$\begin{aligned}d\rho = &-\frac{i}{\hbar}[H,\rho]\,dt \\ &- r^2\frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\sigma]\rho\,dN_1 \\ &- \left(1-r^2\right)\frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\sigma]\rho\,dN_2.\end{aligned} \tag{18.89}$$

This equation is equivalent to the original direct-detection master equation (18.86), but now we can interpret this form as direct detection of the atomic fluorescence after a beam splitter, where terms on the second line represent detector 1, and the terms on the third line represent detector 2.

Now note that we currently have collapse operators

$$\begin{aligned} C_1 &= \sqrt{\Gamma}\, r\sigma \\ C_2 &= \sqrt{\Gamma}\sqrt{1-r^2}\,\sigma \end{aligned} \tag{18.90}$$

associated with the two detectors, which give the expectation values (18.88). To incorporate the local-oscillator field, and obtain the proper homodyne collapse operators (18.75) and (18.76), we must make the replacement

$$\sigma \longrightarrow \sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha \tag{18.91}$$

in the parts of the master equation (18.89) associated with detector 1, and we must also make the replacement

$$\sigma \longrightarrow \sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha \tag{18.92}$$

in the parts of the master equation (18.89) associated with detector 2. In analogy with the calculation of Eq. (18.47), upon making these replacements, to keep the *unconditioned* master equation unchanged, we must also transform the Hamiltonian according to

$$H \longrightarrow H - \frac{i\hbar\Gamma}{2} r\sqrt{1-r^2}\left(\alpha^*\sigma - \alpha\sigma^\dagger\right) + \frac{i\hbar\Gamma}{2} r\sqrt{1-r^2}\left(\alpha^*\sigma - \alpha\sigma^\dagger\right) = H, \tag{18.93}$$

and thus the Hamiltonian needs no modification under the above replacements. That is, the effects of the two replacements on the unconditioned master equation exactly cancel, unlike the case of simple homodyne detection. Thus, implementing the replacements in Eq. (18.89), we have

$$\begin{aligned} d\rho = &-\frac{i}{\hbar}[H,\rho]dt \\ &- r^2\frac{\Gamma}{2}\mathcal{H}\left[\left(\sigma^\dagger + \frac{\sqrt{1-r^2}}{r}\,\alpha^*\right)\left(\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right)\right]\rho\,dt + \mathcal{J}\left[\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right]\rho\,dN_1 \\ &- \left(1-r^2\right)\frac{\Gamma}{2}\mathcal{H}\left[\left(\sigma^\dagger - \frac{r}{\sqrt{1-r^2}}\,\alpha^*\right)\left(\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right)\right]\rho\,dt + \mathcal{J}\left[\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right]\rho\,dN_2. \end{aligned} \tag{18.94}$$

Expanding out and simplifying the $\mathcal{H}[c]\rho$ terms, we see that all terms involving α cancel, leaving the simpler expression

$$\begin{aligned} d\rho = &-\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}\left[\sigma^\dagger\sigma\right]\rho\,dt \\ &+ \mathcal{J}\left[\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right]\rho\,dN_1 + \mathcal{J}\left[\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right]\rho\,dN_2. \end{aligned} \tag{18.95}$$

Now, to take the white-noise limit, where the amplitude of the local oscillator is large ($|\alpha| \longrightarrow \infty$). In this case, we again make replacements of the form

$$dN \longrightarrow \langle\!\langle \frac{dN}{dt} \rangle\!\rangle dt + \sqrt{\langle\!\langle \frac{dN}{dt} \rangle\!\rangle}\, dW, \tag{18.96}$$

In particular, the contribution to the master equation proportional to dt from the dN_1 term is

$$\mathcal{J}\left[\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right]\rho\langle\!\langle dN_1\rangle\!\rangle = r^2\Gamma\left\langle\sigma^\dagger\sigma\right\rangle\mathcal{J}[\sigma]\rho\,dt + r\sqrt{1-r^2}\,\Gamma\mathcal{H}[\alpha^*\sigma]\rho\,dt, \tag{18.97}$$

where the details are exactly as in the previous calculation of Eqs. (18.56). Similarly, the contribution to the master equation proportional to dt from the dN_2 term is

$$\mathcal{J}\left[\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right]\rho\langle\!\langle dN_2\rangle\!\rangle = \left(1-r^2\right)\Gamma\left\langle\sigma^\dagger\sigma\right\rangle\mathcal{J}[\sigma]\rho\,dt - r\sqrt{1-r^2}\,\Gamma\mathcal{H}[\alpha^*\sigma]\rho\,dt. \tag{18.98}$$

When these parts are included in the master equation, the first terms combine to form the usual dissipation term from the unconditioned equation, while the second terms cancel each other, and the result is

$$\begin{aligned} d\rho = &-\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt \\ &+\sqrt{r^2\Gamma\left\langle\left(\sigma^\dagger + \frac{\sqrt{1-r^2}}{r}\,\alpha^*\right)\left(\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right)\right\rangle}\;\mathcal{J}\left[\sigma + \frac{\sqrt{1-r^2}}{r}\,\alpha\right]\rho\,dW_1 \\ &+\sqrt{(1-r^2)\,\Gamma\left\langle\left(\sigma^\dagger - \frac{r}{\sqrt{1-r^2}}\,\alpha^*\right)\left(\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right)\right\rangle}\;\mathcal{J}\left[\sigma - \frac{r}{\sqrt{1-r^2}}\,\alpha\right]\rho\,dW_2, \end{aligned} \tag{18.99}$$

where now dW_1 and dW_2 are the Wiener processes corresponding to dN_1 and dN_2, respectively. Unfortunately, the stochastic terms do not simplify nearly as well, and so we will now consider the limit of a strong local oscillator, $|\alpha| \longrightarrow \infty$, and keep only the lowest order terms in $|\alpha|^{-1}$. The calculation is the same as in Eqs. (18.58),

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \sqrt{\Gamma}\mathcal{H}\left[\sigma e^{i\phi}\right]\rho\,dW, \tag{18.100}$$

(balanced homodyne master equation)

where as before ϕ is the phase of the local-oscillator field,

$$\alpha =: |\alpha| e^{-i\phi}, \tag{18.101}$$

(local-oscillator phase)

and where the Wiener process dW is given in terms of its independent constituents by

$$dW = r\,dW_1 + \sqrt{1-r^2}\,dW_2, \tag{18.102}$$

(composite noise process)

and is thus itself a standard Wiener process. The form of the master equation is thus independent of r, but the two Wiener processes contribute to dW equally only in the balanced case $r = 1/\sqrt{2}$.

18.2.6.2 Measurement Record

When we take the white-noise limit of the subtracted photocurrent (i.e., the measurement record), and again keep only lowest-order terms in $|\alpha|^{-1}$, we find

$$I_- = (Q_{\mathrm{ph},1} + Q_{\mathrm{ph},2})\,r\sqrt{1-r^2}\left\langle\alpha^*\sigma + \alpha\sigma^\dagger\right\rangle dt + Q_{\mathrm{ph},1}\sqrt{\Gamma(1-r^2)|\alpha|^2}\,dW_1 - Q_{\mathrm{ph},2}\sqrt{\Gamma r^2|\alpha|^2}\,dW_2, \tag{18.103}$$

if we assume condition (18.83) is fulfilled so that the large $|\alpha|^2$ dc offset term in the mean photocurrent vanishes. However, note that even with Eq. (18.83), there is no simple way to combine the two quantum-noise terms. If the two detectors have equal response, $Q_{\mathrm{ph},1} = Q_{\mathrm{ph},2} = Q_{\mathrm{ph}}$, then

$$I_- = Q_{\mathrm{ph}}\Gamma\left\langle\alpha^*\sigma + \alpha\sigma^\dagger\right\rangle dt + Q_{\mathrm{ph}}\sqrt{\Gamma}|\alpha|\,dW', \tag{18.104}$$

where the composite Wiener process is

$$dW' = \sqrt{1-r^2}\,dW_1 + r\,dW_2. \tag{18.105}$$

However, note that dW' is *not* equivalent to the composite Wiener process dW that appeared in the master equation (18.102). These two noise processes are only simply identified in the case of perfect balancing, $r = 1/\sqrt{2}$, in which case $dW' = dW$,

$$I_- = Q_{\mathrm{ph}}\Gamma\left\langle\alpha^*\sigma + \alpha\sigma^\dagger\right\rangle dt + Q_{\mathrm{ph}}\sqrt{\Gamma}|\alpha|\,dW. \tag{18.106}$$

(balanced homodyne photocurrent)

This is of the same form as for simple homodyne detection, but no extra subtraction is required to eliminate the dc offset term. The measurement photocurrent can then be rescaled as for simple homodyne detection.

18.2.7 Heterodyne Detection

In general, the local-oscillator field need not have the same nominal frequency as the atomic radition (i.e., it need not be the same field that drives the atom), and in such case the measurement setup corresponds to **heterodyne detection**. In the above analysis of homodyne detection, we have made no particular assumptions about the phase ϕ of the local oscillator, and we can thus treat heterodyne detection by simply setting $\phi = \Delta t$, where $\Delta = \omega_{\rm lo} - \omega$ is the detuning between the local oscillator frequency $\omega_{\rm lo}$ and the driving field ω of the atom. Thus, from the homodyne SME (18.89), the heterodyne-detection master equation becomes

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \sqrt{\Gamma}\mathcal{H}\big[\sigma e^{i\Delta t}\big]\rho\,dW, \tag{18.107}$$
(heterodyne SME)

and from Eq. (18.66), the measurement record (scaled photocurrent) is

$$dr(t) = \Gamma\big\langle \sigma e^{i\Delta t} + \sigma^\dagger e^{-i\Delta t}\big\rangle\,dt + \sqrt{\Gamma}\,dW. \tag{18.108}$$
(heterodyne measurement record)

Heterodyne detection is often used in the case where the detuning Δ is large (compared to some relevant frequency scale, such as the inverse of a measurement averaging time), in which case the information in the measurement record is "encoded" at the high frequency Δ. This is often pragmatically useful in the laboratory, since technical noise tends to drop off as $1/\omega$, and thus the encoding at high frequency is generally less susceptible to technical noise. After detection it is most useful to demodulate the photocurrent signal by shifting the useful information to near-dc. In post-processing this is done by multiplying by the harmonic function $e^{-i\Delta t}$ to obtain

$$d\tilde{r}(t) := dr(t)e^{-i\Delta t} = \Gamma\big\langle \sigma + \sigma^\dagger e^{-2i\Delta t}\big\rangle\,dt + \sqrt{\Gamma}\,dV, \tag{18.109}$$

where

$$dV := e^{-i\Delta t}dW \tag{18.110}$$
(rotating Wiener process)

is the frequency-shifted noise process. For large detunings, the rapidly rotating term at frequency 2Δ is ignorable over any reasonable averaging time and is thus negligible, so we may write[10]

$$d\tilde{r}(t) \approx \Gamma\langle\sigma\rangle\,dt + \sqrt{\Gamma}\,dV, \tag{18.111}$$
(heterodyne measurement record)

Thus, the measurement record now contains information about $\langle\sigma\rangle$, and thus about *both* quadratures $\langle\sigma_x\rangle$ and $\langle\sigma_y\rangle$. The reason that this works is that the detuning separates the components $\langle\sigma\rangle$ and $\langle\sigma^\dagger\rangle$ to different frequencies $\pm\Delta$, and thus the information about only one of these expectation values may be detected. By contrast, in homodyne detection the information about the two operators is encoded at the same frequency, and thus we only obtain information in the quadrature combinations $\langle\sigma \pm \sigma^\dagger\rangle$. Note that demodulating the signal, as in analog electronics, by multiplying dr by the *real* harmonic function $\cos\Delta t$ does *not* work in the same way, since it shifts the information for *both* $\langle\sigma\rangle$ and $\langle\sigma^\dagger\rangle$ to zero frequency, and thus amounts simply to homodyne detection.

In terms of the noise process dV, we may write the master equation as

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \sqrt{\Gamma}\mathcal{H}\big[\sigma e^{i\Delta t}\big]\rho\,e^{i\Delta t}dV. \tag{18.112}$$
(heterodyne SME)

In the noise term there are thus contributions that go as $\sigma e^{2i\Delta t}dV$ and $\sigma^\dagger dV$. However, we may *not* make a rotating wave approximation here and neglect the former compared to the latter, because dW contains all

[10] For explicit solutions of the dynamics for homodyne and heterodyne detection, see Howard Wiseman, "Complementarity in Spontaneous Emission: Quantum Jumps, Staggers, and Slides," in *Directions in Quantum Optics*, H. J. Carmichael, R. J. Glauber, and M. O. Scully, Eds. (Springer, 2001), p. 347.

frequencies, and thus both terms are ''fast'' in the same sense. In treating this equation it is useful to note the relations (valid in the limit of large Δ)

$$\begin{aligned} \langle\langle dV \rangle\rangle &= e^{i\Delta t}\langle\langle dW \rangle\rangle = 0 \\ dV^2 &= e^{-2i\Delta t} dt \approx 0 \\ (dV)^* dV &= dW^2 = dt. \end{aligned} \tag{18.113}$$

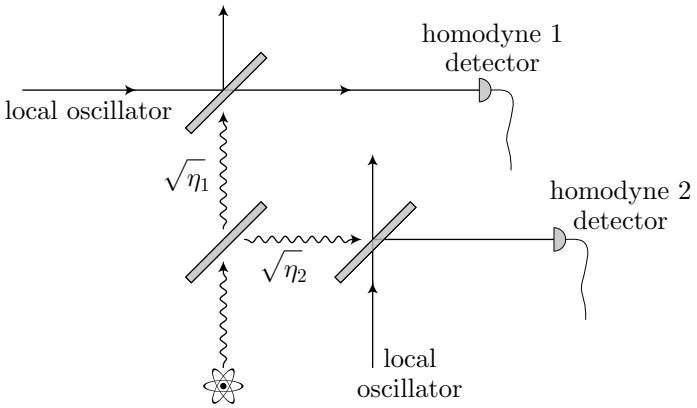
Otherwise, there is no fundamental simplification to the master equation in this case, and even though the measurement information is different, it is simplest to think of the heterodyne and homodyne master equations as equivalent except for the time dependence of the phase ϕ.

18.2.8 Detector Efficiency and Multiple Observers

The SME (18.60) for homodyne detection,

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\, dt + \sqrt{\Gamma}\mathcal{H}[\sigma]\rho\, dW, \tag{18.114}$$

where for simplicity we have taken the local-oscillator phase $\phi = 0$ (though all of our conclusions apply to any setup above), applies when the detector catches *all* of the photons emitted by the atom. To model a finite detection efficiency, suppose the radiated field is split by a beam splitter into two components weighted by $\sqrt{\eta_1}$ and $\sqrt{\eta_2}$, respectively (i.e., the *intensities* are weighted by η_1 and η_2), which are then monitored by homodyne detection on homodyne detectors 1 and 2, respectively.



We are assuming the amplitudes $\sqrt{\eta_1}$ and $\sqrt{\eta_2}$ to be real, with $\eta_1 + \eta_2 = 1$. Then the SME above is modified to include the *two* measurement processes, where $\sigma \longrightarrow \sqrt{\eta_1}\sigma$ for the process on detector 1, and $\sigma \longrightarrow \sqrt{\eta_2}\sigma$ for the process on detector 2:

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sqrt{\eta_1}\sigma]\rho\, dt + \sqrt{\Gamma}\mathcal{H}[\sqrt{\eta_1}\sigma]\rho\, dW_1 + \Gamma\mathcal{D}[\sqrt{\eta_2}\sigma]\rho\, dt + \sqrt{\Gamma}\mathcal{H}[\sqrt{\eta_2}\sigma]\rho\, dW_2. \tag{18.115}$$

We can combine the two dissipation terms, since they are linear in $\eta_{1,2}$, and factor the $\eta_{1,2}$ out of the measurement terms to obtain

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\, dt + \sqrt{\eta_1\Gamma}\mathcal{H}[\sigma]\rho\, dW_1 + \sqrt{\eta_2\Gamma}\mathcal{H}[\sigma]\rho\, dW_2. \tag{18.116}$$

(SME for two observers)

The dissipation term is precisely the same as if the light were not split: the disturbance does not depend on th details of the measurement. Of course, if we define

$$dW := \sqrt{\eta_1}dW_1 + \sqrt{\eta_2}dW_2, \tag{18.117}$$

we recover the original master equation where the light was not split. The measurement record (18.66) is correspondingly modified into two measurement records for the two detectors:

$$\begin{aligned} dr_1(t) &= \Gamma\langle\sqrt{\eta_1}\sigma + \sqrt{\eta_1}\sigma^\dagger\rangle\, dt + \sqrt{\Gamma}\, dW_1 \\ dr_2(t) &= \Gamma\langle\sqrt{\eta_2}\sigma + \sqrt{\eta_2}\sigma^\dagger\rangle\, dt + \sqrt{\Gamma}\, dW_2. \end{aligned}$$

(measurement records for two observers) (18.118)

The original measurement record is recovered by taking the combination

$$\sqrt{\eta_1}dr_1(t) + \sqrt{\eta_2}dr_2(t) = \Gamma\langle\sigma + \sigma^\dagger\rangle\, dt + \sqrt{\Gamma}\, dW = dr(t). \tag{18.119}$$

We can rescale these records so that the expectation values have the same amplitudes as the original measurement record:

$$\begin{aligned} d\tilde{r}_1(t) &= \frac{dr_1}{\sqrt{\eta_1}} = \Gamma\langle\sigma + \sigma^\dagger\rangle\, dt + \sqrt{\frac{\Gamma}{\eta_1}}\, dW_1 \\ d\tilde{r}_2(t) &= \frac{dr_2}{\sqrt{\eta_2}} = \Gamma\langle\sigma + \sigma^\dagger\rangle\, dt + \sqrt{\frac{\Gamma}{\eta_2}}\, dW_2. \end{aligned}$$

(measurement records for two observers) (18.120)

In this case, the quantum noise is effectively amplified for each of the two detectors compared to the case of a single detector. As we will discuss below, the increased quantum noise is due to the presence of another information channel, which necessarily disturbs, or back-acts, on the quantum system.

Now what we essentially have is the theory for two observers monitoring the same atomic fluorescence. Suppose that there are two observers 1 and 2, each of which has access to *only* their respective detector 1 or 2. Observer 1 does not have access to detector 2, and thus must trace over all possible results on detector 2. We do this by taking an ensemble average over all possible realizations of dW_2, which we again do by effectively setting $dW_2 = 0$, and thus obtain the SME for the state of knowledge ρ_1 of observer 1:

$$d\rho_1 = -\frac{i}{\hbar}[H, \rho_1]dt + \Gamma\mathcal{D}[\sigma]\rho_1\, dt + \sqrt{\eta_1\Gamma}\mathcal{H}_1[\sigma]\rho_1\, dW_1'.$$

(SME for observer 1) (18.121)

Here, we have used the notation

$$\begin{aligned} \mathcal{H}_1[c]\rho &:= c\rho + \rho c^\dagger - \mathrm{Tr}[c\rho_1 + \rho_1 c^\dagger]\rho \\ &= c\rho + \rho c^\dagger - \langle c + c^\dagger\rangle_1 \rho, \end{aligned} \tag{18.122}$$

so that the superscript on the $\mathcal{H}$ superoperator denotes that the expectation value is taken with respect to ρ_1. The measurement record for observer 1 has the same form as in Eq. (18.120),

$$d\tilde{r}_1(t) = \Gamma\langle\sigma + \sigma^\dagger\rangle_1\, dt + \sqrt{\frac{\Gamma}{\eta_1}}\, dW_1',$$

(measurement record for observer 1) (18.123)

but now the expectation value is taken with respect to ρ_1. Correspondingly, we cannot assume that the noise process dW_1' according to observer 1 is the same as the original dW_1. Similarly, the SME for observer 2 is

$$d\rho_2 = -\frac{i}{\hbar}[H, \rho_2]dt + \Gamma\mathcal{D}[\sigma]\rho_2\, dt + \sqrt{\eta_2\Gamma}\mathcal{H}_2[\sigma]\rho_2\, dW_2',$$

(SME for observer 2) (18.124)

and the measurement record according to observer 2 is

$$d\tilde{r}_2(t) = \Gamma\langle\sigma + \sigma^\dagger\rangle_2\, dt + \sqrt{\frac{\Gamma}{\eta_2}}\, dW_2'.$$

(measurement record for observer 2) (18.125)

We can interpret the first SME (18.116) and corresponding measurement records (18.120) that include *both* detectors as those of another *omniscient* observer that has access to both detectors. We can now equate the two expressions for each measurement record to obtain

$$\begin{aligned} dW_1' &= \sqrt{\eta_1\Gamma}\Big[\langle\sigma+\sigma^\dagger\rangle - \langle\sigma+\sigma^\dagger\rangle_1\Big] + dW_1 \\ dW_2' &= \sqrt{\eta_2\Gamma}\Big[\langle\sigma+\sigma^\dagger\rangle - \langle\sigma+\sigma^\dagger\rangle_2\Big] + dW_2, \end{aligned} \tag{18.126}$$

relating the noise processes dW_1' and dW_2' for the individual observers to the noise processes dW_1 and dW_2 for the omniscient observer. Note that we have derived the above equations assuming the two observers are making the *same* measurement on the light, but this is easily generalized to the case of two *different* measurements by the two observers.

The case of inefficient detection is exactly the case of a single observer in the presence of a second observer, where the observer does not have all the possible information and needs to trace over all the undetected information. Then the observer's SME becomes

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\mathcal{D}[\sigma]\rho\,dt + \sqrt{\eta\Gamma}\mathcal{H}[\sigma]\rho\,dW,$$

(inefficient homodyne detection SME) (18.127)

and the measurement record is

$$d\tilde{r}(t) = \Gamma\langle\sigma+\sigma^\dagger\rangle\,dt + \sqrt{\frac{\Gamma}{\eta}}\,dW,$$

(inefficient homodyne detection measurement record) (18.128)

where η is the efficiency of the detector (i.e., the fraction of total intensity that is actually registers on the detector).

18.3 Conditioned Dynamics and Squeezing

Now that we have fairly general forms of the master equation for homodyne and heterodyne measurement, we would like to interpret the measurement terms in the master equations to see their physical meaning. In particular, the $\mathcal{H}[c]\rho$ terms (i.e., the noise terms) represent the information gain due to the measurement process, while the $\mathcal{D}[c]\rho$ terms represent the disturbance to, or the *backaction* on, the state of the system due to the measurement. Of course, as we see from the dependence on the efficiency η, the backaction occurs independently of whether the observer uses or discards the measurement information (corresponding to $\eta = 1$ or 0, respectively).

Interpreting the master equation in this way is an important exercise because the measurement record dr tells us what the observer actually *measured*, but to find out what the observer actually *knows* about the system in light of the measurement, *we must actually solve the master equation.* Of course, this is quite difficult to do in general, but we can consider the evolution of the lowest *moments* (expectation values of powers of x and p) of the canonical variables. This will give us the observer's time-dependent estimates of position and momentum, as well as the associated uncertainties.

Let us consider the case of homodyne detection of a *cavity field*—which we recall from Chapter 12 is equivalent to homodyne detection of the *atomic* field under the replacements $\sigma \longrightarrow a$, $\Gamma \longrightarrow \kappa$—with the additional evolution under the action of a Hamiltonian H:

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + \kappa\mathcal{D}[a]\rho\,dt + \sqrt{\eta\kappa}\mathcal{H}[a]\rho\,dW.$$

(SME for cavity homodyne detection) (18.129)

Correspondingly, the measurement record for measurement efficiency η is

$$d\tilde{r}(t) = \kappa\langle a + a^\dagger\rangle\, dt + \sqrt{\frac{\kappa}{\eta}}\, dW.$$

(measurement record for cavity homodyne detection) (18.130)

Here, a is the cavity annihilation operator, and we are only considering a measurement of the X_1 quadrature (proportional to $a + a^\dagger$, as we will define below) to simplify things. For an arbitrary operator A, we can use the master equation and $d\langle A\rangle = \mathrm{Tr}[A\, d\rho]$ to obtain following equation of motion for the expectation value $\langle A\rangle$:

$$\begin{aligned} d\langle A\rangle = &-\frac{i}{\hbar}\langle [A, H]\rangle\, dt \\ &+\kappa\left\langle a^\dagger A a - \frac{1}{2}\left(a^\dagger a A + A a^\dagger a\right)\right\rangle dt \\ &+\sqrt{\eta\kappa}\langle a^\dagger A + A a - \langle A\rangle\langle a + a^\dagger\rangle\rangle\, dW. \end{aligned}$$

(expectation-value evolution under SME) (18.131)

The first line gives the Hamiltonian evolution, the second line the effect of the dissipation/disturbance $\mathcal{D}[a]\rho$, and the last line is the effect of the measurement information $\mathcal{H}[a]\rho$.

The evolution of the *isolated* cavity is then given by the harmonic oscillator Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2, \tag{18.132}$$

(cavity Hamiltonian)

which we write in canonical coordinates rather than the raising and lowering operators (m here is an "effective mass," which while quantizing the field we decided was the permittivity ϵ_0). We will thus derive the the lowest few moments of x and p using the above formula for $d\langle A\rangle$. We will also make the simplifying assumption that the initial state is *Gaussian*, so that we only need to consider the simplest five moments: the means $\langle x\rangle$ and $\langle p\rangle$, the variances V_x and V_p, where $V_\alpha := \langle \alpha^2\rangle - \langle\alpha\rangle^2$, and the symmetrized covariance $C_{xp} := (1/2)\langle [x, p]_+\rangle - \langle x\rangle\langle p\rangle$. These moments completely characterize arbitrary Gaussian states (including mixed states). Recall from Section 5.6.1 that we already decided that the Gaussian state was a "natural" state for the *damped* harmonic oscillator, where $\mathcal{D}[a]\rho$ is precisely the damping term that we used then.

18.3.1 Moment Equations

To set up the calculation, we recall for the harmonic oscillator that the annihilation operator is related to the canonical coordinates by

$$a = \frac{1}{\sqrt{2}x_0}x + i\frac{x_0}{\sqrt{2}\hbar}p, \tag{18.133}$$

where the length scale x_0 is defined by

$$x_0 := \sqrt{\frac{\hbar}{m\omega}}. \tag{18.134}$$

Evaluating the terms in the above evolution equation for the various moments gives the following moment equations for the conditioned evolution in this case:

$$\begin{aligned}
d\langle x\rangle &= \frac{1}{m}\langle p\rangle\, dt - \frac{\kappa}{2}\langle x\rangle\, dt + \sqrt{2\eta\kappa\frac{m\omega}{\hbar}}\left(V_x - \frac{\hbar}{2m\omega}\right) dW \\
d\langle p\rangle &= -m\omega^2\langle x\rangle\, dt - \frac{\kappa}{2}\langle p\rangle\, dt + \sqrt{2\eta\kappa\frac{m\omega}{\hbar}}C_{xp}\, dW \\
\partial_t V_x &= \frac{2}{m}C_{xp} - \kappa\left(V_x - \frac{\hbar}{2m\omega}\right) - 2\eta\kappa\frac{m\omega}{\hbar}\left(V_x - \frac{\hbar}{2m\omega}\right)^2 \\
\partial_t V_p &= -2m\omega^2 C_{xp} - \kappa\left(V_p - \frac{m\omega\hbar}{2}\right) - 2\eta\kappa\frac{m\omega}{\hbar}C_{xp}^2 \\
\partial_t C_{xp} &= \frac{1}{m}V_p - m\omega^2 V_x - \kappa C_{xp} - 2\eta\kappa\frac{m\omega}{\hbar}C_{xp}\left(V_x - \frac{\hbar}{2m\omega}\right).
\end{aligned}$$

(moment evolution under SME) (18.135)

Here, we have used the following moment relations, valid for a Gaussian state:[11]

$$\begin{aligned}
\left\langle x^3\right\rangle &= 3\langle x\rangle V_x + \langle x\rangle^3 \\
\tfrac{1}{2}\left\langle [x,p^2]_+\right\rangle &= 2\langle p\rangle C_{xp} + \langle x\rangle\left[V_p + \langle p\rangle^2\right] \\
\tfrac{1}{4}\left\langle [x,[x,p]_+]_+\right\rangle &= 2\langle x\rangle C_{xp} + \langle p\rangle\left[V_x + \langle x\rangle^2\right].
\end{aligned} \tag{18.136}$$

This approximation decouples the variances from any higher-order moments and removes any noise terms from the variance equations.

18.3.2 Quadrature Moments

We can make these equations look a bit more symmetric by defining the stationary quadrature operators

$$\begin{aligned}
X_1 &:= \frac{1}{2}\left(a + a^\dagger\right) = \sqrt{\frac{m\omega}{2\hbar}}\, x \\
X_2 &:= \frac{1}{2i}\left(a - a^\dagger\right) = \sqrt{\frac{1}{2m\omega\hbar}}\, p,
\end{aligned} \tag{18.137}$$

(field quadratures)

in which case the moment equations transform to

$$\begin{aligned}
d\langle X_1\rangle &= \omega\langle X_2\rangle\, dt - \frac{\kappa}{2}\langle X_1\rangle\, dt + \sqrt{4\eta\kappa}\left(V_{X_1} - \frac{1}{4}\right) dW \\
d\langle X_2\rangle &= -\omega\langle X_1\rangle\, dt - \frac{\kappa}{2}\langle X_2\rangle\, dt + \sqrt{4\eta\kappa}C_{X_1X_2}\, dW \\
\partial_t V_{X_1} &= 2\omega C_{X_1X_2} - \kappa\left(V_{X_1} - \frac{1}{4}\right) - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2 \\
\partial_t V_{X_2} &= -2\omega C_{X_1X_2} - \kappa\left(V_{X_2} - \frac{1}{4}\right) - 4\eta\kappa C_{X_1X_2}^2 \\
\partial_t C_{X_1X_2} &= \omega(V_{X_2} - V_{X_1}) - \kappa C_{X_1X_2} - 4\eta\kappa C_{X_1X_2}\left(V_{X_1} - \frac{1}{4}\right).
\end{aligned}$$

(quadrature-moment evolution) (18.138)

Here it is more obvious that X_1 and X_2 are treated symmetrically, and the Hamiltonian evolution simply involves a rotation in the X_1-X_2 plane at a frequency ω, corresponding to the free evolution of the cavity field.

[11] Salman Habib, "Gaussian Dynamics is Classical Dynamics," arXiv.org preprint (arXiv: quant-ph/0406011); though the relations here simply reflect the fact that for Gaussians, odd-order centered moments always vanish.

18.3.3 Interpretation

Now on to the interpretation of the moment equations (18.138). First, consider the unconditioned evolution of the means $\langle X_1\rangle$ and $\langle X_2\rangle$, where we average over all possible noise realizations. Again, since $\langle\!\langle \rho\, dW\rangle\!\rangle = 0$, we can simply set $dW = 0$ in the above equations, and we will drop the double angle brackets for brevity. The Hamiltonian evolution terms are of course the same, but now we see extra damping terms. Decoupling these two equations gives an equation of the usual form for the damped harmonic oscillator for the mean position:

$$\partial_t^2\langle X_1\rangle + \kappa\partial_t\langle X_1\rangle + \left(\omega^2 + \frac{\kappa^2}{4}\right)\langle X_1\rangle = 0. \tag{18.139}$$

The same equation of motion follows for the other quadrature X_2. Note that we identify the frequency ω here as the actual oscillation frequency ω_κ of the damped oscillator, given by $\omega_\kappa^2 = \omega^2 - \kappa^2/4$, and not the resonance frequency ω that appears the usual form of the classical formula. Then both quadratures undergo damped harmonic oscillation, so that the trajectory in the X_1-X_2 plane is a radially symmetric spiral towards the origin (or a circle in the limit $\kappa = 0$).

The noise terms in these equations correspond to *nonstationary diffusion*, or diffusion where the transport rate depends on the state of the system. Note that under such a diffusive process, the system will tend to come to rest in configurations where the diffusion coefficient vanishes, an effect closely related to the "blowtorch theorem."[12] Here, this corresponds to $V_{X_1} = 1/4$ and $C_{X_1X_2} = 0$, or $V_x = \hbar/2m\omega$ and $C_{xp} = 0$ in the original coordinates, which correspond to the values of the ground state (or any coherent state).

The variance equations also contain unconditioned damping terms (proportional to κ but not η). These damping terms cause the system to equilibrate with the same variance values as noted above; they also produce the extra equilibrium value $V_{X_2} = 1/4$ or $V_p = m\omega\hbar/2$. The conditioning term (proportional to η) in the equation for V_{X_1} merely accelerates the contraction of V_{X_1}, and thus represents information gain in the X_1 quadrature (i.e., the one we are measuring). It also accelerates the settling to the equilibrium value $V_{X_1} = 1/4$. The measurement term in the V_{X_2} equation involves only the covariance: this says that if $C_{X_1X_2} \neq 0$, then the two quadratures are correlated, and thus a measurement on X_1 also provides some information about X_2.

18.3.4 Squeezing (or Lack Thereof)

Thus, we see that the essential effect of the antihermitian measurement operator is to damp the energy from the system, whether it is stored in the centroids or in the variances. In fact, what we see is that this measurement process selects *coherent states*, states that have the same shape as the harmonic-oscillator ground state, but whose centroids oscillate along the classical harmonic-oscillator trajectories. One thing that we can immediately conclude from this analysis is that even though the homodyne measurement obtains information about the "x quadrature," since the measurement accelerates the decay of V_x, we can see that the measurement *does not squeeze* the quadradure—that is, the uncertanty does not become smaller than that of the coherent state (ground state) in steady state. Squeezing *can* be produced by a measurement, but it requires the measurement operator to be of the form $a + a^\dagger$ (i.e., we must realize a direct, Hermitian "position measurement") rather than simply a measurement via a. That is, the master equation should be of the form

$$d\rho = \kappa\mathcal{D}[X_1]\rho\, dt + \sqrt{\eta\kappa}\mathcal{H}[X_1]\rho\, dW \tag{18.140}$$

to produce squeezing in X_1, but this does not correspond to a photodetection measurement that we have considered thus far. Because the measurement operator is Hermitian in this case, the measurement would not cause damping, and would have to be realized by a *dispersive* (i.e., nonabsorbing) measurement interaction, say by firing a beam of atoms through a lossless cavity and measuring atomic phase shifts due to nonresonant interaction with the cavity field. This master equation also has the form of a position measurement, which we will consider in depth in the next chapter.

[12] Term coined by Rolf Landauer, "Statistical physics of machinery: forgotten middle-ground," *Physica A* **194**, 551 (1993) (doi: 10.1016/0378-4371(93)90385-H).

18.3.5 Homodyne Detection

Now recall that in homodyne detection, the frequency of the local oscillator matches that of the cavity, which has the same effect in the above equations of setting $\omega \longrightarrow 0$:

$$\begin{aligned}
d\langle X_1\rangle &= -\frac{\kappa}{2}\langle X_1\rangle\, dt + \sqrt{4\eta\kappa}\left(V_{X_1} - \frac{1}{4}\right) dW \\
d\langle X_2\rangle &= -\frac{\kappa}{2}\langle X_2\rangle\, dt + \sqrt{4\eta\kappa} C_{X_1X_2}\, dW \\
\partial_t V_{X_1} &= -\kappa\left(V_{X_1} - \frac{1}{4}\right) - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2 \\
\partial_t V_{X_2} &= -\kappa\left(V_{X_2} - \frac{1}{4}\right) - 4\eta\kappa C_{X_1X_2}^2 \\
\partial_t C_{X_1X_2} &= -\kappa C_{X_1X_2} - 4\eta\kappa C_{X_1X_2}\left(V_{X_1} - \frac{1}{4}\right).
\end{aligned}$$

(quadrature-moment evolution under homodyne detection) (18.141)

In this way, the measurement always gets information about a single quadrature (in this case, the measurement provides information about the X_1 quadrature).

18.3.6 Heterodyne Detection

To treat *heterodyne* detection, note that the local oscillator and cavity frequencies do *not* match, which amounts to letting $\omega \longrightarrow \Delta$ in Eqs. (18.138), where Δ is again the detuning between the local oscillator and the cavity. Then defining the corotating quadratures

$$\begin{aligned}
\tilde{X}_1 &:= X_1 \cos\Delta t - X_2 \sin\Delta t \\
\tilde{X}_2 &:= X_1 \sin\Delta t + X_2 \cos\Delta t,
\end{aligned}$$

(18.142)
(corotating quadratures)

we can transform the variances to the new variables using

$$\begin{aligned}
V_{\tilde{X}_1} &= V_{X_1}\cos^2\Delta t + V_{X_2}\sin^2\Delta t - 2C_{X_1X_2}\sin\Delta t\cos\Delta t \\
V_{\tilde{X}_2} &= V_{X_1}\sin^2\Delta t + V_{X_2}\cos^2\Delta t + 2C_{X_1X_2}\sin\Delta t\cos\Delta t \\
C_{\tilde{X}_1\tilde{X}_2} &= V_{X_1}\sin\Delta t\cos\Delta t - V_{X_2}\sin\Delta t\cos\Delta t + C_{X_1X_2}(\cos^2\Delta t - \sin^2\Delta t)
\end{aligned} \tag{18.143}$$

to rewrite the moment equations as

$$\begin{aligned}
d\left\langle \tilde{X}_1 \right\rangle &= -\frac{\kappa}{2}\left\langle \tilde{X}_1 \right\rangle dt + \sqrt{4\eta\kappa}\left(V_{X_1} - \frac{1}{4}\right)\cos\Delta t\, dW - \sqrt{4\eta\kappa}C_{X_1X_2}\sin\Delta t\, dW \\
d\left\langle \tilde{X}_2 \right\rangle &= -\frac{\kappa}{2}\left\langle \tilde{X}_2 \right\rangle dt + \sqrt{4\eta\kappa}\left(V_{X_1} - \frac{1}{4}\right)\sin\Delta t\, dW + \sqrt{4\eta\kappa}C_{X_1X_2}\cos\Delta t\, dW \\
\partial_t V_{\tilde{X}_1} &= -\kappa\left(V_{\tilde{X}_1} - \frac{1}{4}\right) - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2\cos^2\Delta t - 4\eta\kappa C_{X_1X_2}^2\sin^2\Delta t \\
&\quad +8\eta\kappa C_{X_1X_2}\left(V_{X_1} - \frac{1}{4}\right)\sin\Delta t\cos\Delta t \\
\partial_t V_{\tilde{X}_2} &= -\kappa\left(V_{X_2} - \frac{1}{4}\right) - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2\sin^2\Delta t - 4\eta\kappa C_{X_1X_2}^2\cos^2\Delta t \\
&\quad -8\eta\kappa C_{X_1X_2}\left(V_{X_1} - \frac{1}{4}\right)\sin\Delta t\cos\Delta t \\
\partial_t C_{\tilde{X}_1\tilde{X}_2} &= -\kappa C_{\tilde{X}_1\tilde{X}_2} - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2\sin\Delta t\cos\Delta t + 4\eta\kappa C_{X_1X_2}^2\sin\Delta t\cos\Delta t \\
&\quad +2\eta\kappa C_{X_1X_2}\left(V_{X_1} - \frac{1}{4}\right)(\cos^2\Delta t - \sin^2\Delta t).
\end{aligned} \tag{18.144}$$

In doing so, we have eliminated the Hamiltonian free-evolution terms, but we have introduced some explicit time dependence in the equations. We will now consider the variance equations in the limit of large Δ, and we will replace terms oscillating at frequencies of order Δ by their time-averaged values, being careful to implement the inverse relations for the original variances,

$$\begin{aligned}
V_{X_1} &= V_{\tilde{X}_1}\cos^2\Delta t + V_{\tilde{X}_2}\sin^2\Delta t + 2C_{\tilde{X}_1\tilde{X}_2}\sin\Delta t\cos\Delta t \\
V_{X_2} &= V_{\tilde{X}_1}\sin^2\Delta t + V_{\tilde{X}_2}\cos^2\Delta t - 2C_{\tilde{X}_1\tilde{X}_2}\sin\Delta t\cos\Delta t \\
C_{X_1X_2} &= -V_{\tilde{X}_1}\sin\Delta t\cos\Delta t + V_{\tilde{X}_2}\sin\Delta t\cos\Delta t + C_{\tilde{X}_1\tilde{X}_2}(\cos^2\Delta t - \sin^2\Delta t),
\end{aligned} \tag{18.145}$$

so that we obtain

$$\begin{aligned}
\partial_t V_{\tilde{X}_1} &= -\kappa\left(V_{\tilde{X}_1} - \frac{1}{4}\right) - 2\eta\kappa C_{\tilde{X}_1\tilde{X}_2}^2 - 2\eta\kappa\left(V_{\tilde{X}_1} - \frac{1}{4}\right)^2 \\
\partial_t V_{\tilde{X}_2} &= -\kappa\left(V_{\tilde{X}_2} - \frac{1}{4}\right) - 2\eta\kappa C_{\tilde{X}_1\tilde{X}_2}^2 - 2\eta\kappa\left(V_{\tilde{X}_2} - \frac{1}{4}\right)^2 \\
\partial_t C_{\tilde{X}_1\tilde{X}_2} &= -\kappa C_{\tilde{X}_1\tilde{X}_2} - \frac{\eta\kappa}{2}C_{\tilde{X}_1\tilde{X}_2}\left(V_{\tilde{X}_1} + V_{\tilde{X}_2} - \frac{1}{2}\right).
\end{aligned}$$

(variance evolution under heterodyne detection) (18.146)

The heterodyne variance equations are now relatively simple. In particular, notice that in homodyne detection, the measurement was represented in the X_1 quadrature by the term

$$-4\eta\kappa\left(V_{\tilde{X}_1} - \frac{1}{4}\right)^2, \tag{18.147}$$

while in heterodyne detection, *both* quadratures have measurement terms of this form, but with an overall factor of $2\eta\kappa$ instead of $4\eta\kappa$. So again, while heterodyne detection provides measurment information about *both* quadratures, it does so at only *half* the rate at which homodyne detection provides information about a *single* quadrature. Note that you can get similar results by splitting the field on a 50/50 beam splitter, and use *two* homodyne detection setups, set to monitor complementary quadratures, to monitor each output field of the beam splitter. In this case, the factor of 1/2 in the information rate is more obvious.

18.3.7 Explicit Solutions for the Uncertainty Dynamics

Incidentally, for homodyne detection, the X_1 variance evolves from Eqs. (18.141) as

$$\partial_t V_{X_1} = -\kappa \left(V_{X_1} - \frac{1}{4} \right) - 4\eta\kappa \left(V_{X_1} - \frac{1}{4} \right)^2 . \tag{18.148}$$

If we assume the variance is very broad, the measurement term dominates the dissipation term, so that

$$\partial_t V_{X_1} = -4\eta\kappa \left(V_{X_1} - \frac{1}{4} \right)^2 , \tag{18.149}$$

and this equation has the solution

$$V_{X_1}(t) = \frac{V_{X_1}(0) - 1/4}{1 + 4\eta\kappa[V_{X_1}(0) - 1/4]t} + \frac{1}{4}. \tag{18.150}$$

That is, the variance decreases like $1/t$, which makes sense since the uncertainty should decrease as $1/\sqrt{t}$ for averaging a noisy process. Of course, as the variance approaches the steady-state value of $1/4$, the damping term becomes dominant, and the decay becomes exponential. This is represented by the more general solution of (18.148),

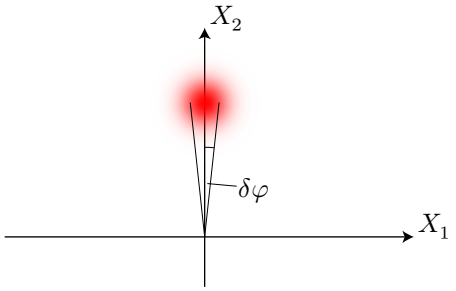
$$V_{X_1}(t) = \frac{(V_{X_1}(0) - 1/4)}{e^{\kappa t} + 4\eta[V_{X_1}(0) - 1/4](e^{\kappa t} - 1)} + \frac{1}{4}, \tag{18.151}$$

which contains the $1/t$ behavior at short times as well as the exponential decay at long times.

18.3.8 Phase Estimation

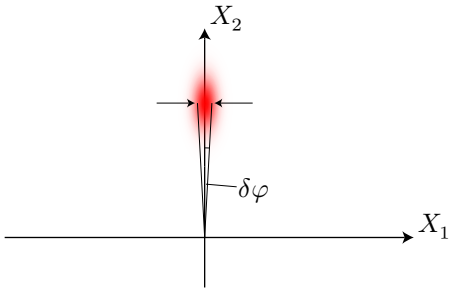
As an example of an application that illustrates the difference between homodyne and heterodyne detection, we will consider the problem of estimating the phase of a pulse of light. The phase uncertainty and quantum noise here, for example, put limits on how much information can be encoded in the phase of an optical pulse. To keep this treatment simple, we will consider as as simple model the field pulse to be modeled in the same way as the single-mode cavity field above.

First, consider the homodyne detection of the cavity phase in terms of the two complementary quadrature variables X_1 and X_2. We will suppose the phase of the field to be reasonably well defined and the field amplitude to be larger than its uncertainty. We will assume that $\langle X_1 \rangle = 0$ initially; we can treat the general case by simply applying a rotation to this basic configuration. We will further assume a Gaussian state for simplicity, and to capture the essence of the problem (recall that the state will be asymptotically Gaussian under the measurement anyway). We can then represent the quantum state by a distribution in phase space (e.g., a Wigner distribution, as in Section 4.3).



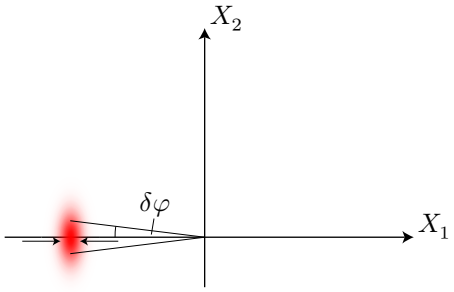
The question of the phase of the quantum state is then essentially recasting the same problem of measuring the field quadratures into polar coordinates. In the particular case shown here, the uncertainty in the phase φ of the quantum state is clearly related to the uncertainty in X_1. But it should also be clear that the phase uncertainty $\Delta\varphi$ is related to the *amplitude* of the field (i.e., the distance of the wave-packet centroid from the origin): a larger amplitude implies a smaller phase uncertainty, given that the uncertainty ΔX_1 is fixed.

The measurement, as we discovered above, causes the X_1 uncertainty to contract, assuming the local-oscillator phase $\phi = 0$. For the particular case we are considering now, this means that the phase uncertainty decreases—we gain knowledge about the phase.

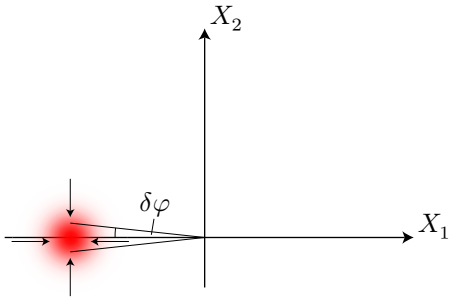


The other effect of the measurement is that the amplitude of the field decreases, so that the centroid of the wave packet moves towards the origin. After all, photodetection proceeds as the photodetector absorbs the field. (This is assuming that the cavity is not driven to counteract the cavity damping.) The dissipation has the opposite effect on the phase uncertainty; as the wave packet moves towards the origin with ΔX_1 fixed, the angle subtended by the wave packet increases, and thus $\delta\varphi$ increases. Of course, we have seen that the dissipation also reduces ΔX_1. However, once the variance V_{X_1} reaches its steady-state value 1/4, any decrease in amplitude increases the phase uncertainty. Of course, the point of all this is that when the phase uncertainty is much larger than the minimum quantum limit and the amplitude of the field is large, the dominant effect of the measurement is the rapid decrease of ΔX_1, which reduces the phase uncertainty.

Now, however, consider the case where the phase φ of the field is near π rather than near $\pi/2$. In this case, a reduction in the uncertainty ΔX_1 reduces the uncertainty of the field *amplitude*, but *not* its phase.



Thus, gaining information on the phase in homodyne detection depends on having a particular phase to begin with. That is, the phase sensitivity of homodyne detection has a phase-dependent sensitivity. Heterodyne detection has the advantage that the measurement reduces the phase uncertainty for any phase, since as we saw above, the uncertainties of *both* X_1 and X_2 decrease in response to the meaasurement.



The price of this ''omnidirectional'' phase sensitivity is that the phase uncertainty decreases (assuming the decrease to be dominated by the measurement information) at half the rate for the best case of homodyne detection. Of course, it greatly outperforms the homodyne measurement in its worst case, and the heterodyne measurement simultaneously provides information about the field amplitude.

To see this mathematically, note that for the case of homodyne detection in the best case of $\langle\varphi\rangle = \pi/2$, to see the reduction in phase uncertainty, we need the equation for V_{X_1} in Eqs. (18.141):

$$\partial_t V_{X_1} = -\kappa\left(V_{X_1} - \frac{1}{4}\right) - 4\eta\kappa\left(V_{X_1} - \frac{1}{4}\right)^2. \tag{18.152}$$

If the wave packet is localized and the amplitude of the field is large, then to lowest order in the phase uncertainty we may convert between the Cartesian and polar variances according to

$$\langle R\rangle^2 V_\varphi \approx V_{X_1}, \quad \text{where} \quad \langle R\rangle = \langle X_2\rangle. \tag{18.153}$$

Then assuming the measurement-induced reduction of V_φ is much faster than the damping of $\langle R\rangle$, we may write

$$\partial_t V_\varphi \approx -\kappa\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right) - 4\eta\kappa\langle R\rangle^2\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right)^2. \tag{18.154}$$

We can see here again the first, damping term (which should also be negligible in this limit), and the second, information term. We can explicitly see how the quantum-limited phase uncertainty is related to the field amplitude: the best phase uncertainty is $\delta\phi \approx 1/2\langle R\rangle$, at least in the absence of squeezing. Furthermore, the rate of information collapse increases with the field amplitude, which again reflects the fact that the phase uncertainty $\delta\varphi$ is related both to ΔX_1 *and* the field amplitude.

Then to treat the general case, we must rotate the phase space while maintaining the measurement of X_1. We have already done this in Eq. (18.144), and adapting the variance equation for $V_{\tilde{X}_1}$ in the same way, we replace Δt by ϕ and for simplicity ignore any covariance between the amplitude and phase to find

$$\partial_t V_\varphi \approx -\kappa\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right) - 4\eta\kappa\langle R\rangle^2\cos^2\phi\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right)^2. \tag{18.155}$$

Thus the measurement-induced collapse rate is now modulated by $\cos^2\phi$, where ϕ now represents the phase change of the local oscillator from our base case above. That is, this is the sensitivity of the homodyne measurement when the expected phase is $\langle\varphi\rangle = \phi + \pi/2$ and the measurement is of the X_1 quadrature. The heterodyne case follows from adapting the same variance equation in Eq. (18.146):

$$\partial_t V_\varphi \approx -\kappa\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right) - 2\eta\kappa\langle R\rangle^2\left(V_\varphi - \frac{1}{4\langle R\rangle^2}\right)^2. \tag{18.156}$$

This is, of course, the same as the homodyne expression averaged over the local-oscillator phase ϕ. Again, we have lost the dependence on the local-oscillator phase, but at the cost of a factor of 2 in the information-collapse rate.

18.3.8.1 Adaptive Measurements

Then in making a phase measurement, how is it possible to take advantage of the extra phase sensitivity in homodyne detection, when it seems to require already knowing the phase to begin with? One strategy is to use *both* heterodyne and homodyne detection in an **adaptive phase measurement**.[13] The idea is to start out a measurement of a light pulse without any knowledge of the phase using heterodyne detection. As the observer begins to get an idea of the phase, the observer switches the local oscillator to homodyne detection, with a local-oscillator phase set to maximize the sensitivity based on the heterodyne estimate. As the homodyne measurement continues, the observer feeds back to the local oscillator phase to track the estimated phase (which is diffusing stochasically due to the measurement) to ensure that sensitivity is always maximized.

[13] H. M. Wiseman, "Adaptive Phase Measurements of Optical Modes: Going Beyond the Marginal Q Distribution," *Physical Review Letters* **75**, 4587 (1995) (doi: 10.1103/PhysRevLett.75.4587); this adaptive scheme was implemented experimentally by Michael A. Armen, John K. Au, John K. Stockton, Andrew C. Doherty, and Hideo Mabuchi, "Adaptive Homodyne Measurement of Optical Phase," *Physical Review Letters* **89**, 133602 (2002) (doi: 10.1103/PhysRevLett.89.133602).

18.4 Exercises

Problem 18.1

Verify that the stochastic Schrödinger equation (SSE) [Eq. (18.18)] for quantum jumps,

$$d|\psi\rangle = -\frac{i}{\hbar}H|\psi\rangle dt + \frac{\Gamma}{2}\left(\langle\sigma^\dagger\sigma\rangle - \sigma^\dagger\sigma\right)|\psi\rangle dt + \left(\frac{\sigma}{\sqrt{\langle\sigma^\dagger\sigma\rangle}} - 1\right)|\psi\rangle\, dN, \tag{18.157}$$

is equivalent to the stochastic master equation (SME) [Eq. (18.158)]

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+ dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN. \tag{18.158}$$

Chapter 19

Position Measurement

Here we will study the continuous observation of a Hermitian observable, namely the position of a quantum particle. We will do so fairly abstractly, but then give a physical example of how a position measurement can arise in atomic resonance fluorescence.

19.1 Prelude: General Form for the Master Equation

Before working out another continuous measurement process, we can ask the question, what is the *most general* form of the measurement master equation when the measurements involve Gaussian noise? A simple but nonrigorous argument[1] that establishes the general form for the *unconditioned* master equation, and then extend it to examine stochastic master equations (SMEs). Thus, we will see that the form of the (Markovian) SME involving Wiener noise is quite constrained, and it is intuitively easy to adapt the SME to many different measurement processes simply by choosing the correct measurement operator (which we denote below by c).

19.1.1 Positive Maps

Under unitary (unconditioned) evolution, the Schrödinger equation tells us that in a short time interval dt, the state vector undergoes the transformation

$$|\psi\rangle \longrightarrow |\psi\rangle + d|\psi\rangle = \left(1 - i\frac{H}{\hbar}\,dt\right)|\psi\rangle, \tag{19.1}$$

where H is the Hamiltonian. The same transformation applied to the density operator gives the Schrödinger–von Neumann equation (from Section 4.1):

$$\rho + d\rho = \left(1 - i\frac{H}{\hbar}\,dt\right)\rho\left(1 + i\frac{H}{\hbar}\,dt\right) = \rho - \frac{i}{\hbar}[H,\rho]\,dt. \tag{19.2}$$

To be physical, any transformation of the density operator must be *completely positive*. That is, the transformation must preserve the fact that the density operator has only nonnegative eigenvalues. This property guarantees that the density operator can generate only sensible (nonnegative) probabilities. (To be more precise, *complete* positivity means that the transformation for a system's density operator must preserve the positivity of the density operator—the fact that the density operator has no negative eigenvalues—of any larger system containing the system.) It turns out that the most general form of a linear, completely positive

[1] S. L. Adler, "Derivation of the Lindblad generator structure by use of the Itô stochastic calculus," *Physics Letters A* **265**, 58 (2000) (doi: 10.1016/S0375-9601(99)00847-6).

transformation is[2]

$$\rho \longrightarrow \sum_n A_n \rho A_n^\dagger, \tag{19.3}$$

where the A_n are arbitrary operators. The Hamiltonian evolution above corresponds to a single infinitesimal transformation operator $A = 1 - iH\,dt/\hbar$.

19.1.2 Lindblad Form

Now let's examine the transformation for a more general, *stochastic* operator of the form

$$A = 1 - i\frac{H}{\hbar}\,dt + b\,dt + c\,dW, \tag{19.4}$$

where b and c are operators. We will use this operator to ''derive'' a Markovian master equation, then indicate how it can be made more general. We may assume here that b is Hermitian, since we can absorb any antihermitian part into the Hamiltonian. Putting this into the transformation (19.3), we find

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + [b,\rho]_+ dt + c\rho c^\dagger\,dt + \left(c\rho + \rho c^\dagger\right)\,dW, \tag{19.5}$$

recalling that $[A,B]_+ := AB + BA$ is the anticommutator. We can then take an average over all possible Wiener processes, which again we denote by the double angle brackets $\langle\!\langle\ \rangle\!\rangle$. To compute the ensemble average, we again use the property $\langle\!\langle \rho\,dW \rangle\!\rangle = 0$ of Itō calculus, so that

$$d\langle\!\langle\rho\rangle\!\rangle = -\frac{i}{\hbar}\left[H, \langle\!\langle\rho\rangle\!\rangle\right]\,dt + \left[b, \langle\!\langle\rho\rangle\!\rangle\right]_+ dt + c\langle\!\langle\rho\rangle\!\rangle c^\dagger\,dt. \tag{19.6}$$

Since the operator $\langle\!\langle\rho\rangle\!\rangle$ is an average over valid density operators, it is also a valid density operator and must therefore satisfy $\mathrm{Tr}[\langle\!\langle\rho\rangle\!\rangle] = 1$. Hence we must have $d\mathrm{Tr}[\langle\!\langle\rho\rangle\!\rangle] = \mathrm{Tr}[d\langle\!\langle\rho\rangle\!\rangle] = 0$. Using the cyclic property of the trace, this gives the constraint

$$\mathrm{Tr}\left[\langle\!\langle\rho\rangle\!\rangle\left(2b + c^\dagger c\right)\right] = 0. \tag{19.7}$$

This holds for an arbitrary density operator only if

$$b = -\frac{c^\dagger c}{2}. \tag{19.8}$$

Thus we obtain the *Lindblad form*[3] of the unconditioned master equation (averaged over all possible noise realizations):

$$d\langle\!\langle\rho\rangle\!\rangle = -\frac{i}{\hbar}\left[H, \langle\!\langle\rho\rangle\!\rangle\right]\,dt + \mathcal{D}[c]\langle\!\langle\rho\rangle\!\rangle\,dt. \tag{19.9}$$

As before, we have defined the Lindblad superoperator

$$\mathcal{D}[c]\rho := c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right), \tag{19.10}$$

where ''superoperator'' refers to the fact that $\mathcal{D}[c]$ operates on ρ from both sides. It is worth reiterating here that the $c\rho c^\dagger$ term results from the dW part of the transformation, and thus this term cannot be represented by a Hamiltonian transformation, even if the Hamiltonian is non-Hermitian, as we noted in Section 5.5.3. This is the most general (Markovian) form of the unconditioned master equation for a single dissipation process. Different choices for the operator c thus give the quantum backaction (disturbance) for different measurement processes.

[2]K.-E. Hellwig and K. Kraus, ''Operations and Measurements. II,'' *Communications in Mathematical Physics* **16**, 142 (1970); Benjamin Schumacher, ''Sending entanglement through noisy quantum channels,'' *Physical Review A* **54**, 2614 (1996) (doi: 10.1103/PhysRevA.54.2614).

[3]G. Lindblad, ''On the generators of quantum dynamical semigroups,'' *Communications in Mathematical Physics* **48**, 199 (1976) (doi: 10.1007/BF01608499).

19.1.3 Stochastic Terms

The full transformation from Eq. (19.5) then becomes

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + \mathcal{D}[c]\rho\,dt + \left(c\rho + \rho c^\dagger\right)\,dW. \tag{19.11}$$

This is the *linear* SME, which we will discuss again elsewhere. We know from our treatment in Section 18.2 of homodyne detection (where $c \longrightarrow \sigma$) that this equation is not "complete," since it is missing the nonlinear term. The problem is that this form of the master equation does not in general preserve the trace of the density operator, since the condition $\mathrm{Tr}[d\rho] = 0$ implies

$$\mathrm{Tr}\left[\rho\left(c + c^\dagger\right)\,dW\right] = 0. \tag{19.12}$$

We could interpret this relation as a constraint[4], on c, but we will instead keep c an arbitrary operator and explicitly renormalize $\rho + d\rho$ by adding a term proportional to the left-hand side of (19.12). The result is the nonlinear form

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + \mathcal{D}[c]\rho\,dt + \mathcal{H}[c]\rho\,dW, \tag{19.13}$$

where again the measurement superoperator is

$$\mathcal{H}[c]\rho := c\rho + \rho c^\dagger - \left\langle c + c^\dagger\right\rangle \rho. \tag{19.14}$$

This corresponds to using the normalizing transformation

$$\rho \longrightarrow \frac{A\rho A^\dagger}{\mathrm{Tr}[A\rho A^\dagger]}, \tag{19.15}$$

instead of the unnormalized transformation that we considered,

$$\rho \longrightarrow A\rho A^\dagger, \tag{19.16}$$

to first order in dt. The normalized transformation has exactly the form of a POVM-type reduction as we consider below.

When $c = \sqrt{\Gamma}\sigma$, we recover precisely the master equation for homodyne detection of spontaneous emission. In general, c can be chosen differently to model different continuous measurement processes.

19.1.4 Generalization

More generally, we may have any number of measurements, or *output channels*, happening simultaneously. The result is

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + \sum_n \left(\mathcal{D}[c_n]\rho\,dt + \mathcal{H}[c_n]\rho\,dW_n\right). \tag{19.17}$$

This is the same as Eq. (19.13), but this time summed (integrated) over multiple possible measurement operators c_n, each with a separate Wiener noise process independent of all the others. This simply corresponds to having multiple terms in the general positive map (19.3).

In view of the arguments from our treatment of detector efficiency in homodyne detection in Section (18.2.8), when the measurements are inefficient, we have[5]

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt + \sum_n \left(\mathcal{D}[c_n]\rho\,dt + \sqrt{\eta_n}\mathcal{H}[c_n]\rho\,dW_n\right), \tag{19.18}$$

[4] S. L. Adler, *op. cit.*

[5] This form is *close* to the most general form of the master equation, but can still be generalized further. See H. M. Wiseman and L. Diosi, "Complete parameterization, and invariance, of diffusive quantum trajectories for Markovian open systems," *Chemical Physics* **268**, 91 (2001) (doi: 10.1016/S0301-0104(01)00296-8).

where η_n is the efficiency of the nth detection channel. The corresponding measurement record for the nth process can be written (with an arbitrary normalization) as

$$dr_n(t) = \frac{\langle c_n + c_n^\dagger \rangle}{2} dt + \frac{dW_n}{\sqrt{4\eta_n}}. \tag{19.19}$$

Again, for homodyne detection, we recover the right results if we let $c = \sigma$ and interpret $dr_n/2\Gamma$ as a rescaled measurement record.

19.2 A Second Prelude: Positive-Operator-Valued Measures

To help handle generalized measurements, we will now introduce the somewhat mathematical concept of a positive-operator-valued measure (POVM). By referring to *generalized* measurements, we mean to differentiate these measurements from the usual projective, or *von Neumann*, measurements, which is what you normally find in introductory quantum-mechanics texts. The usual description goes like this: for a quantum system in state $|\psi\rangle$, a measurement of the observable Q leaves the system in an eigenstate $|q\rangle$ of Q with probability $\langle\psi|Q|\psi\rangle$, in which case the "result" of the measurement is the eigenvalue q. We can see that this notion of a measurement is lacking in two situations. First, it does not properly describe the situation in photodetection of atomic radiation, where each detection event results in the loss of energy (i.e., the atom is always found to be in the ground state), and we gain information even during instants when a photon is not detected. Thus, POVMs are crucial to the formal definition of a continuous measurement. The second situation is when the observable is the position operator, where eigenstate collapse is unphysical: a position eigenstate is a state of infinite energy. POVMs allow us to define an *imprecise* or *partial* measurement of an observable, which will be a stepping stone on the way to defining a *continuous* measurement of position.

19.2.1 Discrete, Finite Spaces

Consider a discrete, finite Hilbert space of dimension N. That is, the Hilbert space is spanned by the set of eigenstates

$$\{|q\rangle : q = 1, \ldots, N\} \tag{19.20}$$

of the observable Q. Then we can define a **positive-operator-valued measure** (POVM) as a set of positive-semidefinite operators $\Omega_q^\dagger \Omega_q$ that sum to the identity operator:

$$\sum_{q=1}^{N_q} \Omega_q^\dagger \Omega_q = 1. \tag{19.21}$$

Note that we are writing the qth positive operator $\Omega_q^\dagger \Omega_q$ as a factorization in terms of the **Kraus operator** Ω_q, since any positive operator always has such a factorization. We also note that the number N_q of positive operators is not necessarily the same as the dimension N of the Hilbert space.

Now the important *physical* point here is that a POVM defines a **quantum measurement** on the Hilbert space. The qth possible outcome of the measurement is that the state vector changes according to the replacement

$$|\psi\rangle \longrightarrow \frac{\Omega_q|\psi\rangle}{\sqrt{\left\langle \Omega_q^\dagger \Omega_q \right\rangle}}, \tag{19.22}$$

or in terms of the density operator,

$$\rho \longrightarrow \frac{\Omega_q \rho \Omega_q^\dagger}{\mathrm{Tr}[\Omega_q \rho \Omega_q^\dagger]} = \frac{\Omega_q \rho \Omega_q^\dagger}{\left\langle \Omega_q^\dagger \Omega_q \right\rangle}. \tag{19.23}$$

That is, in the qth outcome, the state is "hit" by the operator Ω_q and then renormalized if necessary. The probability that the qth outcome occurs is

$$P(q) = \mathrm{Tr}[\Omega_q \rho \Omega_q^\dagger] = \left\langle \Omega_q^\dagger \Omega_q \right\rangle . \tag{19.24}$$

The (classical) "result" of the quantum measurement in this case is simply q (or some physically meaningful function of q). This notion may seem rather abstract, but we can note that the usual projective measurement comes out as a special case of the POVM-based measurement. In particular, the usual measurement arises from a **projection-valued measure**, where we partition the Hilbert space according to a set of (Hermitian) projection operators

$$P_q := |q\rangle\langle q| \tag{19.25}$$

that also sum to the identity:

$$\sum_{q=1}^{N} P_q^2 = 1. \tag{19.26}$$

Of course, $P_q^2 = P_q$, but we have written the sum in this form to emphasize the similarity with Eq. (19.21) by taking $\Omega_q = P_q$ and $N_q = N$. Then the standard projective measurement of the observable Q results in the qth outcome of a reduction to the qth eigenstate $|q\rangle$,

$$|\psi\rangle \longrightarrow \frac{P_q|\psi\rangle}{\sqrt{\langle P_q^2 \rangle}} = |q\rangle, \tag{19.27}$$

or in terms of the density operator,

$$\rho \longrightarrow \frac{P_q \rho P_q^\dagger}{\mathrm{Tr}[P_q \rho P_q^\dagger]} = \frac{P_q \rho P_q^\dagger}{\langle P_q^2 \rangle} = |q\rangle\langle q|. \tag{19.28}$$

This outcome happens with probability

$$P(q) = \mathrm{Tr}[P_q \rho P_q^\dagger] = \left\langle P_q^2 \right\rangle = \langle P_q \rangle , \tag{19.29}$$

which for a pure state $|\psi\rangle$ becomes the familiar Born rule

$$P(q) = |\langle q|\psi\rangle|^2. \tag{19.30}$$

Thus, the POVM-based measurement above is a reasonably straightforward generalization of the usual projective measurements, at least when the standard measurements are cast in the proper way.

19.2.2 Measure

Why is a POVM called a "POVM"? The answer requires an excursion into mathematics, and so the short answer, if you feel the need to skip forward, is that a measure is usually something that assigns numbers to sets, and so a *positive-operator-valued* measure is a measure that instead associates *positive operators* with sets, and thence probabilities to the same sets via the expectation value as above. To *really* answer this question, we need to define what we usually mean by a *measure*, and then adapt it to the operator case. Informally, a measure is a rule for assigning numbers to subsets of some set, or **space**. This is a very useful notion in probability theory, where you would consider the set of all possible outcomes or events, and the measure would assign probabilities to each outcome or collection of outcomes. Alternately, a measure is an abstraction of the notion of volume, where the measure represents the "volume" of subsets of the main set.

Before formally defining a measure, though, we should first note that for a given space, it is problematic to try to define a measure on *every* subset. Instead, we will define the measure on only a limited collection of subsets, chosen to make the definition of the measure consistent. Formally, this collection is a **σ-algebra**, which we define as a collection $\mathscr{S}$ of subsets of the space X such that:

1. The empty set is included: $\emptyset \in \mathscr{S}$.
2. Countable, disjoint unions are included (with countable here meaning finite or countably infinite): if $\mathscr{U} \subset \mathscr{S}$ with $A \cap B = \emptyset$ for any $A, B \in \mathscr{U}$, and $\mathscr{U}$ is countable, then
$$\bigcup_{A \in \mathscr{U}} A \in \mathscr{S}. \tag{19.31}$$
3. Complements are included: if $A \in \mathscr{S}$, then $X - A \in \mathscr{S}$.

Any element of a σ-algebra is said to be a **measurable set**. This definition can be contrasted with the possibly familiar definition for a **topology** on a space X, which is a collection $\mathscr{T}$ of subsets of X such that:[6]

1. The empty set and the whole space are included: $\emptyset \in \mathscr{T}$, $X \in \mathscr{T}$.
2. Arbitrary unions are included: if $\mathscr{U} \subset \mathscr{S}$, then
$$\bigcup_{A \in \mathscr{U}} A \in \mathscr{S}. \tag{19.32}$$
3. Finite intersections are included: if $\mathscr{U} \subset \mathscr{S}$ with $\mathscr{U}$ finite, then
$$\bigcap_{A \in \mathscr{U}} A \in \mathscr{S}. \tag{19.33}$$

Any element of a topology is said to be an **open set**, while the complement of an open set is said to be a **closed set**. Thus, while topologies contain in general only open sets, σ-algebras contain both open and closed sets. For example, on the real line $\mathbb{R}$, the **standard topology** is the topology consisting of all open intervals of the form (a, b) and all possible unions of such intervals (and the empty set). It turns out there is a unique σ-algebra associated with the standard topology, which is the smallest σ-algebra containing it. This is called the **Borel σ-algebra** on $\mathbb{R}$, which would contain all open intervals as well as all closed intervals of the form $[a, b]$ (and many other sets). The notion of a σ-algebra may not be intuitively clear at this stage, but the definition is basically concocted to make the definition of measure work out, as we will now see.

A **measure** is a function $\mu : \mathscr{S} \longrightarrow [0, \infty]$ defined on a σ-algebra $\mathscr{S}$ on a space X, which satisfies

1. The empty set has zero measure: $\mu(\emptyset) = 0$.
2. The measure for countable, disjoint unions adds: if $\mathscr{U} \subset \mathscr{S}$ with $A \cap B = \emptyset$ for any $A, B \in \mathscr{U}$, and $\mathscr{U}$ is countable, then
$$\mu\left(\bigcup_{A \in \mathscr{U}} A\right) = \sum_{A \in \mathscr{U}} \mu(A). \tag{19.34}$$

These two requirements are sensible considering our analogies to probabilities and volumes, and we can also see how the requirements for a σ-algebra guarantee that we don't have any problems in defining a measure (the last axiom for a σ-algebra imposes the sensible constraint that if A is a measureable subset, then so is $X - A$). Note that the point ∞ is explicitly included in the range of a measure, which is intuitively a "good" measure for something like the entire real line. Also, strictly speaking, we have defined a **positive measure**, since we have only allowed nonnegative values in the range of μ. As an example of measure, the **Lebesgue measure** on the real line is defined on the Borel σ-algebra. We can define it in several cases as follows:

1. It turns out that any *open* set A can be written as the union of a countable set of open intervals (a_j, b_j), in which case the Lebesgue measure of A is the sum of the interval lengths:
$$\mu(A) := \sum_j (b_j - a_j). \tag{19.35}$$

[6] For further reading, see, e.g., James R. Munkres, *Topology: a First Course* (Prentice-Hall, 1975).

2. It turns out that any *closed* set B can be written as a closed interval $[a, b]$ with the union of a countable set of open intervals (a_j, b_j) removed from it,

$$B = [a, b] - \bigcup_j (a_j, b_j), \tag{19.36}$$

where every $a_j > a$ and every $b_j < b$, in which case the Lebesgue measure of B is the length of the closed interval minus the Lebesgue measure of the removed component:

$$\mu(A) := (b - a) - \sum_j (b_j - a_j). \tag{19.37}$$

3. For any other set C in the Borel σ-algebra, the Lebesgue measure is the infimum (greatest lower bound) of the set of Lebesgue measures of all open sets containing C:

$$\mu(C) := \inf\{\mu(A) : A \text{ is open and } C \subset A\} \tag{19.38}$$

Note that there exist sets that do not have Lebesgue measures according to the above definitions, and thus they are excluded by considering only the σ-algebra. The Lebesgue measure is useful in that it extends the notion of length to more complicated and subtle sets: the set of rational numbers, being countable, is a set of Lebesgue measure zero on the real line; and the Cantor middle-thirds set, a fractal set constructed by starting with the interval $[0, 1]$, removing the open ''middle third'' interval $(1/3, 2/3)$, removing the middle-thirds of the two remaining closed intervals, and so on *ad infinitum*, is an uncountable set but of zero Lebesgue measure.

For measurements, the concept of a **probability measure** is more useful, and it is simply that of a measure, but where the range of the measure is $[0, 1]$ rather than $[0, \infty]$, with a measure of the whole space being unity. For example, the Lebesgue measure on the space $[0, 1]$ is a probability measure, and corresponds to a uniform probability density on the same interval.

19.2.3 General Definition

Now with the above mathematical concepts, we can now give a more general definition of a POVM than in the finite case above. In more general terms, a **positive-operator-valued measure** (POVM) defined on a σ-algebra $\mathscr{S}$ on a space X is a function Π that takes as values positive semidefinite, Hermitian operators on a Hilbert space $\mathscr{H}$ such that for any $|\psi\rangle \in \mathscr{H}$, the function $\mu : X \longrightarrow [0, 1]$, defined by

$$\mu(A) := \langle\psi|\Pi(A)|\psi\rangle \tag{19.39}$$

for any measurable subset A of X, defines a *probability measure* on $\mathscr{S}$. In particular, this implies that $\Pi(X)$ is the identity operator, which is the generalization of the sum rule (19.21). Thus, the POVM associates positive operators with measurable subsets of the space of outcomes, which are then associated with probabilities by appropriate expectation values. In this way, we can define a family of probability measures, ''parameterized'' by the quantum state.

We could, of course, write the probability measure more generally in terms of the density operator as

$$\mu(A) = \mathrm{Tr}[\Pi(A)\rho]. \tag{19.40}$$

Incidentally, a trace of this form is (for a Hilbert space of dimension larger than two) the *only* way to construct a quantum probability measure; this is essentially the content of **Gleason's theorem**.[7]

[7] Andrew M. Gleason, ''Measures on the Closed Subspaces of a Hilbert Space,'' *Journal of Mathematics and Mechanics* **6**, 885 (1957) (doi: 10.1512/iumj.1957.6.56060).

19.2.4 Realization

It is important to note that measurements induced by POVMs, while generalizing projective measurements, don't introduce anything *fundamentally* new to quantum mechanics: any of these more general measurements can be realized by introducing an auxiliary system (ancilla), performing a unitary transformation on the combined system, and then perform a *projective* measurement on the ancilla. Thus, generalized measurements correspond to *indirect* measurements, where information about a system comes from projective measurements on the ''environment'' with which the system has interacted (and thus become entangled with).

This result is known as **Naimark's theorem** (or **Neumark's theorem**),[8] and we will only sketch the argument for the finite case here. Starting with the system in the state $|\psi\rangle$, we will extend the Hilbert space to contain the environment, whose dimension is equal to the number of Kraus operators defining the POVM, $|\psi\rangle \longrightarrow |\psi\rangle|0_{\rm E}\rangle \equiv |\psi\ 0_{\rm E}\rangle$. We will assume the environment to always start in a particular state that we label $|0_{\rm E}\rangle$. Note that we are assuming a pure state for the system, which we may as well do as long as we are extending the Hilbert space by invoking purification (Section 4.4.5). We can thus define an operator U that acts on the composite state as

$$U|\psi\ 0_{\rm E}\rangle = \sum_q \left(\Omega_q|\psi\rangle\right)|q_{\rm E}\rangle = \sum_q \sqrt{\langle\psi|\Omega_q^\dagger\Omega_q|\psi\rangle}\left(\frac{\Omega_q|\psi\rangle}{\sqrt{\langle\psi|\Omega_q^\dagger\Omega_q|\psi\rangle}}\right)|q_{\rm E}\rangle, \tag{19.41}$$

where the Kraus operators Ω_q only operate on the original system, and the $|q_{\rm E}\rangle$ environment states are orthogonal. In the last step we have written the part of the state of the original system as a normalized state, leading to explicit coefficients of the superposition. We can also see explicitly how the system and environment are entangled after the operation U. Now computing the norm of the transformed composite state,

$$\begin{aligned}\langle\psi\ 0_{\rm E}|U^\dagger U|\psi\ 0_{\rm E}\rangle &= \sum_{qq'}\langle q_{\rm E}|\langle\psi|\Omega_q^\dagger\ \Omega_{q'}|\psi\rangle|q'_{\rm E}\rangle \\ &= \sum_q \langle\psi|\Omega_q^\dagger\ \Omega_q|\psi\rangle \\ &= \langle\psi|\psi\rangle,\end{aligned} \tag{19.42}$$

so that U preserves the norm of states in the subspace of the original system. The operator U is thus unitary on this subspace, but is not fixed uniquely by the above argument. In principle, the action of U on the environment can be chosen to make U unitary on the composite Hilbert space. Basically, this is because taken as a matrix, the columns of U span the subspace of the original system (i.e., a subset of them form an orthonormal basis), and the extra degrees of freedom (elements of the extra rows) in expanding U to the composite Hilbert space may then be chosen to make the columns of U form an orthonormal basis on the *entire* composite space. Now after the transformation, a projective measurement of the state of the environment leads to the result $|q_{\rm E}\rangle$ with probability

$$\mathrm{Tr}\left[|q_{\rm E}\rangle\langle q_{\rm E}|\ U|\psi\ 0_{\rm E}\rangle\langle\psi\ 0_{\rm E}|U^\dagger\right] = \langle\psi|\Omega_q^\dagger\Omega_q|\psi\rangle. \tag{19.43}$$

Furthermore, the projection of the environment into state $|q_{\rm E}\rangle$ induces the transformation

$$|\psi\rangle \longrightarrow \frac{\Omega_q|\psi\rangle}{\sqrt{\langle\psi|\Omega_q^\dagger\Omega_q|\psi\rangle}} \tag{19.44}$$

on the original system. Thus we have constructed the POVM-based measurement based on the larger projective measurement.

[8] Asher Peres, *Quantum Theory: Concepts and Methods* (Springer, 1995), Section 9-6, p. 285. For a similar argument to what we present here for the unitary representation of linear positive maps, see Benjamin Schumacher, *op. cit.*

19.2.5 Example: Spontaneous Emission

As an example of a POVM, we return to the stochastic master equation for photodetection of atomic resonance fluorescence with quantum jumps from Section 18.1:

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+ dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt + \left(\frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle} - \rho\right)dN. \tag{19.45}$$

In any given time interval of duration dt, there are only two possible outcomes: no photon is detected, or one photon is detected. We can define this evolution in terms of a POVM as follows. Let $U(dt)$ denote the evolution operator for the combined atom–field system. Before each infinitesimal time interval, the field starts in the vacuum state $|0\rangle$, and after each infinitesimal time interval, the detector projectively measures the field and registers a detection event if a photon is emitted into *any* mode. Since the detector does not distinguish modes, we will simply denote the field state as $|1\rangle$ in the case of an emitted photon. Then the two "jump operators" for the two measurement outcomes are[9]

$$\begin{aligned}\Omega_0(dt) &= \langle 0|U(dt)|0\rangle = 1 - i\frac{H}{\hbar}\,dt - \frac{\Gamma}{2}\sigma^\dagger\sigma\,dt \\ \Omega_1(dt) &= \langle 1|U(dt)|0\rangle = \sqrt{\Gamma\,dt}\,\sigma.\end{aligned} \tag{19.46}$$

In the case of no photon detected, the state is transformed according to

$$\rho \longrightarrow \frac{\Omega_0(dt)\rho\Omega_0^\dagger(dt)}{\mathrm{Tr}\left[\Omega_0(dt)\rho\Omega_0^\dagger(dt)\right]} = \rho - \frac{i}{\hbar}[H,\rho]\,dt - \frac{\Gamma}{2}[\sigma^\dagger\sigma,\rho]_+\,dt + \Gamma\langle\sigma^\dagger\sigma\rangle\,\rho\,dt, \tag{19.47}$$

keeping terms to first order in dt, and in the case of a detector click the state is transformed according to

$$\rho \longrightarrow \frac{\Omega_1(dt)\rho\Omega_1^\dagger(dt)}{\mathrm{Tr}\left[\Omega_1(dt)\rho\Omega_1^\dagger(dt)\right]} = \frac{\sigma\rho\sigma^\dagger}{\langle\sigma^\dagger\sigma\rangle}. \tag{19.48}$$

These two transformations correspond exactly to the transformations induced by the SME (19.45) in the cases $dN = 0$ and $dN = 1$, respectively. The probabilities also work out as expected. For example, a photon is detected with probability

$$P(1) = \mathrm{Tr}\left[\Omega_1(dt)\rho\Omega_1^\dagger(dt)\right] = \Gamma\langle\sigma^\dagger\sigma\rangle\,dt, \tag{19.49}$$

and the probability for not detecting a photon is the complement of this, as seen by taking the appropriate trace using $\Omega_0(dt)$.

Notice that this POVM tends to drive the atom towards the ground state, as compared to the unconditioned Hamiltonian evolution (and for either possible outcome $\Omega_{0,1}$). By involving the atomic annihilation operator, we see in this case that the POVM generalizes projective measurements by modeling *dissipation* due to the measurement process. In the case at hand, the physical origin of the dissipation in the case at hand is absorption of radiated photons by the photodetector.

19.2.6 Example: Gaussian Projectors

POVMs can also generalize projective measurements to model *partial* or *imprecise* measurements. Partial measurements leave some uncertainty in the measured observable, whereas projective measurements leave the system in a state where the observable is perfectly defined—that is, an eigenstate of the observable. As a simple example, we can model partial measurements by defining the measurement operators Ω_q to be Gaussian-weighted sums over projection operators for the discrete set of eigenstates $|q\rangle$ ($q \in \mathbb{Z}$) of the observable Q:

$$\Omega_q = \frac{1}{\mathscr{N}}\sum_{j=-\infty}^{\infty} e^{-\kappa(j-q)^2/4}|j\rangle\langle j|. \tag{19.50}$$

[9] H. M. Wiseman, "Quantum Trajectories and Quantum Measurement theory," *Quantum and Semiclassical Optics* **8**, 205 (1996) (doi: 10.1088/1355-5111/8/1/015).

Here,

$$\mathscr{N}^2 := \sum_{j=-\infty}^{\infty} e^{-\kappa(j-q)^2/2}, \tag{19.51}$$

so that

$$\sum_{q=-\infty}^{\infty} \Omega_q^\dagger \Omega_q = 1, \tag{19.52}$$

as required for the operators to form a POVM. The Gaussian weights lead to having only partial information about Q after the measurement. For example, in a highly uncertain mixed state, where $\langle q|\rho|q\rangle$ is approximately the same for any q and $\langle q|\rho|q'\rangle = 0$ for any $q \neq q'$, the measurement leads to the collapse

$$\rho \longrightarrow \frac{\Omega_q \rho \Omega_q^\dagger}{\mathrm{Tr}[\Omega_q \rho \Omega_q^\dagger]} \approx \frac{1}{\mathscr{N}} \sum_j e^{-\kappa(j-q)^2/2} |q\rangle\langle q|. \tag{19.53}$$

The qth possible final state is thus peaked about the eigenvalue q, and additionally has an uncertainty $\Delta Q = 1/\sqrt{\kappa}$.

In the limit $\kappa \longrightarrow \infty$, the measurements here reduce to the usual projective measurements. Thus, for large κ, the variance in the measurement results (taken over an ensemble of measurements on identically prepared systems) is dominated by the uncertainty in the quantum state, while for small κ, the measurement variance is dominated by the uncertainty introduced by the measurement operators Ω_q. This distinction divides two categories of measurments, **strong measurements** where κ is large, and **weak measurements**, where κ is small.[10]

We can also generalize these Gaussian projectors to the continuous-variable case. For example, for a position measurement, the properly normalized measurement operators have the form

$$\Omega(\alpha) = \left(\frac{\kappa}{2\pi}\right)^{1/4} \int_{-\infty}^{\infty} dx\, e^{-\kappa(x-\alpha)^2/4} |x\rangle\langle x|. \tag{19.54}$$

Again, if this operator is applied to a an initially uncertain state (such as a momentum eigenstate), the resulting position variance in the collapsed state is $1/\kappa$ (i.e., the uncertainty is $1/\sqrt{\kappa}$). In what follows, we will consider sequences of weak position measurements of this form, and thus construct *continuous* quantum measurements of position. For this it is useful to consider the product of two operators,

$$\begin{aligned}
\Omega(\alpha';\kappa')\Omega(\alpha;\kappa) &= \left(\frac{\kappa\kappa'}{(2\pi)^2}\right)^{1/4} \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dx\, e^{-\kappa'(x'-\alpha')^2/4} |x'\rangle\langle x'| e^{-\kappa(x-\alpha)^2/4} |x\rangle\langle x| \\
&= \left(\frac{\kappa\kappa'}{(2\pi)^2}\right)^{1/4} \int_{-\infty}^{\infty} dx\, e^{-\kappa'(x-\alpha')^2/4} e^{-\kappa(x-\alpha)^2/4} |x\rangle\langle x| \\
&= \left(\frac{\kappa\kappa'}{(2\pi)^2}\right)^{1/4} \exp\left[-\frac{\kappa\kappa'}{\kappa+\kappa'}(\alpha-\alpha')^2\right] \int_{-\infty}^{\infty} dx\, \exp\left[-\frac{(\kappa+\kappa')}{4}\left(x - \frac{\alpha\kappa+\alpha'\kappa'}{\kappa+\kappa'}\right)^2\right] |x\rangle\langle x|,
\end{aligned} \tag{19.55}$$

which corresponds to a sequence of two Gaussian position measurements, the first of strength κ and the second of strength κ', with measurement outcomes α and then α', respectively. This operator product is still Gaussian, but it is not normalized properly in the sense that $\Omega(\alpha)$ is normalized (note that the norm vanishes if $\alpha - \alpha'$ becomes large), but we can see from its form that applying this operator to an initially uncertain state gives $1/(\kappa+\kappa')$ for the resulting position variance of the state. Hence, a sequence of two Gaussian measurements is effectively equivalent to a *single* Gaussian measurement, where the strength is

[10] Yakir Aharonov, David Z. Albert, and Lev Vaidman, "How the result of a measurement of a component of the spin of a spin-$\frac{1}{2}$ particle can turn out to be 100," *Physical Review Letters* **60**, 1351 (1988) (doi: 10.1103/PhysRevLett.60.1351). See also comments by A. J. Leggett, "Comment on 'How the result of a measurement of a component of the spin of a spin-$\frac{1}{2}$ particle can turn out to be 100,' " *Physical Review Letters* **62**, 2325 (1988) (doi: 10.1103/PhysRevLett.62.2325), and Asher Peres, "Quantum Measurements with Postselection," *Physical Review Letters* **62**, 2326 (1988) (doi: 10.1103/PhysRevLett.62.2326), as well as the reply by Y. Aharonov and L. Vaidman, *Physical Review Letters* **62**, 2327 (1988) (doi: 10.1103/PhysRevLett.62.2327).

the sum of the individual measurement strengths, as long as no other transformation or evolution occurs between the two measurements.

Notice how the information from the second measurement is incorporated with that of the first. After the first measurement, the best estimate for the position of the quantum system is α, with uncertainty $1/\sqrt{\kappa}$. After the second measurement (where the result is α'), the new best position estimate is an average of the old estimate and the new measurement result,

$$\langle x\rangle = \frac{\alpha\kappa + \alpha'\kappa'}{\kappa + \kappa'}, \tag{19.56}$$

weighted by the respective uncertainties. The new uncertainty of the estimate is reduced to $1/\sqrt{\kappa+\kappa'}$.

19.3 Continuous Position Measurement

Now, to construct a *continuous* measurement of position, we will arrange to have a sequence of weak position measurements, separated in time by Δt. We will also let the measurement strength depend on time by making the rescaling $\kappa \longrightarrow 8\kappa\Delta t$ in the measurement operator (19.54), so that

$$\Omega(\alpha;\Delta t) = \left(\frac{4\kappa\Delta t}{\pi}\right)^{1/4} \int_{-\infty}^{\infty} dx\, e^{-2\kappa\Delta t(x-\alpha)^2} |x\rangle\langle x|. \tag{19.57}$$

The factor of 8 here simply gives a convenient normalization for the measurement strength. We will return to the dependence on the time interval Δt below, but this particular scaling is necessary to obtain a sensible limit as $\Delta t \longrightarrow 0$.

Now with this set of measurement operators, the probability of obtaining a particular measurement result α is

$$\begin{aligned} P(\alpha) &= \mathrm{Tr}[\Omega(\alpha)\rho\Omega^\dagger(\alpha)] \\ &= \mathrm{Tr}\left[\left(\frac{4\kappa\Delta t}{\pi}\right)^{1/4} \int_{-\infty}^{\infty} dx\, e^{-2\kappa\Delta t(x-\alpha)^2} |x\rangle\langle x|\rho \left(\frac{4\kappa\Delta t}{\pi}\right)^{1/4} \int_{-\infty}^{\infty} dx'\, e^{-2\kappa\Delta t(x'-\alpha)^2} |x'\rangle\langle x'|\right] \\ &= \sqrt{\frac{4\kappa\Delta t}{\pi}} \int_{-\infty}^{\infty} dx\, e^{-4\kappa\Delta t(x-\alpha)^2} \langle x|\rho|x\rangle. \end{aligned} \tag{19.58}$$

In the limit of small Δt, the Gaussian factor in the integrand is much broader than the position probability density $\langle x|\rho|x\rangle$ for the quantum state. Since it varies slowly over the scale of $\langle x|\rho|x\rangle$, the Gaussian factor can be pulled out of the integral, with x replaced by $\langle x\rangle$, near which $\langle x|\rho|x\rangle$ is peaked. The integral then becomes trivial, and we obtain

$$P(\alpha) = \sqrt{\frac{4\kappa\Delta t}{\pi}} e^{-4\kappa\Delta t[\langle x\rangle - \alpha]^2}, \tag{19.59}$$

so that the measurement result α is a Gaussian random variable with variance $1/8\kappa\Delta t$. Noting also that $\langle\alpha\rangle = \langle x\rangle$, we can write α as an explicit Gaussian random variable in terms of a Wiener increment ΔW (also a Gaussian random variable) as

$$\alpha = \langle x\rangle + \frac{\Delta W}{\sqrt{8\kappa}\Delta t} \tag{19.60}$$

since the mean and variance agree with those from the probability density (19.59). Recall that α is an index for the measurement operator—equivalently, the measurement result for a particular time interval of duration Δt—but we may regard it in a sense as a stochastic, dynamical variable, since the measurement is repeated in time.

19.3.1 State Collapse and the Stochastic Schrödinger Equation

As the stream of measurement results $\alpha(t)$ comes in, the quantum state must be correspondingly modified in light of the new measurement information. Recall that a measurement result of α in a particular time interval of duration Δt causes the state to transform according to

$$|\psi(t+\Delta t)\rangle = \frac{\Omega(\alpha)|\psi(t)\rangle}{\sqrt{\langle\psi(t)|\Omega^\dagger(\alpha)\Omega(\alpha)|\psi(t)\rangle}}. \tag{19.61}$$

To simplify the evaluation, we can ignore the renormalization factor, so that we can use Eq. (19.57) for the measurement operator to see that the state change is given by

$$\begin{aligned}|\tilde{\psi}(t+\Delta t)\rangle &= \Omega(\alpha)|\psi(t)\rangle \\ &= \left(\frac{4\kappa\Delta t}{\pi}\right)^{1/4}\int_{-\infty}^{\infty} dx\, e^{-2\kappa\Delta t(x-\alpha)^2}|x\rangle\langle x|\psi(t)\rangle \\ &= \left(\frac{4\kappa\Delta t}{\pi}\right)^{1/4} e^{-2\kappa\Delta t(x-\alpha)^2}|\psi(t)\rangle,\end{aligned} \tag{19.62}$$

since the x in the exponential is the position operator. Here, the twiddle indicates an unnormalized state vector. Dropping the normalization factor and inserting expression (19.60) for α, we find

$$\begin{aligned}|\tilde{\psi}(t+\Delta t)\rangle &= \exp\left[-2\kappa\,\Delta t\left(x-\langle x\rangle-\frac{\Delta W}{\sqrt{8\kappa}\Delta t}\right)^2\right]|\psi(t)\rangle \\ &\propto \exp\left[-2\kappa\,\Delta t\,x^2 + x\left(4\kappa\langle x\rangle\,\Delta t+\sqrt{2\kappa}\,\Delta W\right)\right]|\psi(t)\rangle,\end{aligned} \tag{19.63}$$

where we have dropped the terms in the exponential that do not involve the position operator. In the infinitesimal limit, we can thus write

$$\begin{aligned}|\tilde{\psi}(t+dt)\rangle &= \exp\left[-2\kappa x^2\,dt + x\left(4\kappa\langle x\rangle\,dt+\sqrt{2\kappa}\,dW\right)\right]|\psi(t)\rangle \\ &= \left[1-2\kappa x^2\,dt + x\left(4\kappa\langle x\rangle\,dt+\sqrt{2\kappa}\,dW\right)+\kappa x^2\,dW^2\right]|\psi(t)\rangle \\ &= \left[1-\left(\kappa x^2-4\kappa x\langle x\rangle\right)dt+\sqrt{2\kappa}\,x\,dW\right]|\psi(t)\rangle,\end{aligned} \tag{19.64}$$

where we have (without approximation) expanded the exponential to first order in dt and second order in dW, setting $dW^2 = dt$ according to the rules of Itō calculus. Normalizing the new state vector and expanding to first order in dt (and second in dW),

$$\begin{aligned}|\psi(t+dt)\rangle &= \frac{|\tilde{\psi}(t+dt)\rangle}{\sqrt{\langle\tilde{\psi}(t+dt)|\tilde{\psi}(t+dt)\rangle}} \\ &= \frac{\left[1-\left(\kappa x^2-4\kappa x\langle x\rangle\right)dt+\sqrt{2\kappa}\,x\,dW\right]|\psi(t)\rangle}{\sqrt{\left\langle 1+8\kappa x\langle x\rangle\,dt+2\sqrt{2\kappa}\,x\,dW\right\rangle}} \\ &= \left[1-\left(\kappa x^2-4\kappa x\langle x\rangle+4\kappa\langle x\rangle^2\right)dt-\left(3\kappa\langle x\rangle^2+2\kappa x\langle x\rangle\right)dW^2+\sqrt{2\kappa}(x-\langle x\rangle)\,dW\right]|\psi(t)\rangle \\ &= \left[1-\kappa\,(x-\langle x\rangle)^2\,dt+\sqrt{2\kappa}(x-\langle x\rangle)\,dW\right]|\psi(t)\rangle,\end{aligned} \tag{19.65}$$

so that we arrive at the stochastic Schrödinger equation for the continuous position measurement:

$$d|\psi\rangle = -\kappa\,(x-\langle x\rangle)^2\,|\psi\rangle\,dt+\sqrt{2\kappa}(x-\langle x\rangle)|\psi\rangle\,dW. \tag{19.66}$$

Again, accompanying the SSE is the measurement record (19.60), which in the infinitesimal limit becomes

$$dy = \langle x \rangle \, dt + \frac{dW}{\sqrt{8\kappa}}, \tag{19.67}$$

where $dy := \alpha \, dt$, or directly in terms of α,

$$\alpha(t) = \langle x \rangle + \frac{\xi(t)}{\sqrt{8\kappa}}, \tag{19.68}$$

where as before $\xi(t) \equiv dW(t)/dt$. In terms of α, the variance sensibly diverges in the infinitesimal limit, because the information gained in time dt is zero. To obtain position information, the observer must average $dy(t)$ over some finite time interval:

$$\begin{aligned} \frac{y(t)}{t} &= \frac{1}{t} \int_0^t dt' \, \langle x \rangle + \frac{1}{\sqrt{8\kappa}\, t} \int_0^t dW(t') \\ &= \frac{1}{t} \int_0^t dt' \, \langle x \rangle + \frac{W(t)}{\sqrt{8\kappa}\, t} \end{aligned} \tag{19.69}$$

The second term represents uncertainty in the measurement, and generically converges to zero as $1/\sqrt{t}$. The first term, which represents the position information, would be simply $\langle x \rangle$ if this expectation value were time-independent, but the integral requires knowledge of the time evolution of the state, and thus its calculation requires the solution to the SSE.

19.3.1.1 Gaussian Noise

In constructing the above SSE, we explicitly assumed *Gaussian* collapse operators $\Omega(\alpha)$. This resulted in the Gaussian noise process dW appearing in the SSE, because in the infinitesimal limit (weak-measurement limit), the width and thus also the shape of the collapse operator determined the noise statistics of measurement results $\alpha(t)$. The question is, how general is this? If we had assumed a different form for $\Omega(\alpha)$, would we have obtained a different noise process? The answer is that the infinitesimal limit is an idealization, and really we should only consider increments

$$\Delta y(t) := \int_t^{t+\Delta t} dy(t) \tag{19.70}$$

for the measurement record (and corresponding finite increments for the quantum-state evolution). Such an increment is a sum over arbitrarily many infinitesimal noise increments, and thus under the continuous idealization, *any* ''reasonable'' form for the probability distribution of the noise increments will give results equivalent to the choice of Gaussian noise and Gaussian collapse operators, according to the central-limit theorem. By ''reasonable,'' we first mean that in the above finite construction, before taking the limit $\Delta t \longrightarrow 0$, the variance of the collapse operator (i.e., the variance of the state $\Omega(\alpha)|\psi\rangle$ for an initially very uncertain state $|\psi\rangle$) should exist and be finite, so that the statistics of many combined collapses are Gaussian. Further, in the limit $\Delta t \longrightarrow 0$, the variance of the infinitesimal increments should be proportional to dt: otherwise, for a variance scaling of the form dt^β, the above integral Δy will be either vanish ($\beta > 1$) or diverge ($\beta < 1$). Thus, Gaussian noise is general in the sense that any (appropriately normalized) continuous noise process dV representing a continuous measurement may be regarded as the Wiener process dW, so long as the integrated increments ΔV have finite variance.

19.3.2 Stochastic Master Equation

As in the case of photodetection, we can generalize the SSE by using it to derive a stochastic master equation, expanding to second order:[11]

$$d\rho = \Big(d|\psi\rangle\Big)\langle\psi| + d|\psi\rangle\Big(d|\psi\rangle\Big) + \Big(d|\psi\rangle\Big)\Big(d|\psi\rangle\Big). \tag{19.71}$$

[11] The Gaussian-projector method in the continuous limit was introduced to derive the *unconditioned* form for this master equation first by Carlton M. Caves and G. J. Milburn, ''Quantum-mechanical model for continuous position measurements,'' *Physical Review A* **36**, 5543 (1987) (doi: 10.1103/PhysRevA.36.5543).

The resulting SME is

$$\begin{aligned} d\rho &= -\kappa[x,[x,\rho]]\,dt + \sqrt{2\kappa}\,(x\rho + \rho x - 2\langle x\rangle\,\rho)\,dW \\ &= 2\kappa\mathcal{D}[x]\rho\,dt + \sqrt{2\kappa}\mathcal{H}[x]\rho\,dW, \end{aligned} \tag{19.72}$$

where recall the superoperators we defined previously in Chapter 18:

$$\begin{aligned} \mathcal{D}[c]\rho &:= c\rho c^\dagger - \frac{1}{2}\left(c^\dagger c\rho + \rho c^\dagger c\right) \\ \mathcal{H}[c]\rho &:= c\rho + \rho c^\dagger - \mathrm{Tr}[c\rho + \rho c^\dagger]\rho. \end{aligned} \tag{19.73}$$

Of course, the dW here, while effectively a stochastic variable, is defined in terms of the measurement record (19.67), so that we may eliminate it and write the SME in terms of the measurement results dy:

$$\begin{aligned} d\rho &= -\kappa[x,[x,\rho]]\,dt + 4\kappa\,(x\rho + \rho x - 2\langle x\rangle\,\rho)\,(dy - \langle x\rangle\,dt) \\ &= 2\kappa\mathcal{D}[x]\rho\,dt + 4\kappa\mathcal{H}[x]\rho\,(dy - \langle x\rangle\,dt). \end{aligned} \tag{19.74}$$

Also, in writing down this SME (as well as the corresponding SSE above), we have ignored any Hamiltonian evolution that proceeds in parallel with the measurement process. Thus, Hamiltonian terms should be added as necessary, so that

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt - \kappa[x,[x,\rho]]\,dt + \sqrt{2\kappa}\,(x\rho + \rho x - 2\langle x\rangle\,\rho)\,dW \tag{19.75}$$

in the case of the SME with system Hamiltonian H. Of course, the corresponding term may be added to the SSE.

19.3.3 Inefficient Detection and Multiple Observers

Notice that we may write the SME (19.72) as

$$d\rho = -\kappa[x,[x,\rho]]\,dt + \sqrt{2\kappa_1}\mathcal{H}[x]\rho\,dW_1 + \sqrt{2\kappa_2}\mathcal{H}[x]\rho\,dW_2, \tag{19.76}$$

and the measurement record as

$$dy = \langle x\rangle\,dt + \frac{dW_1}{\sqrt{8\kappa_1}} + \frac{dW_2}{\sqrt{8\kappa_2}}, \tag{19.77}$$

where dW_1 and dW_2 are Wiener processes, so long as $\kappa_1 + \kappa_2 = \kappa$ and we thus identify $dW = \sqrt{\kappa_1}\,dW_1 + \sqrt{\kappa_2}\,dW_2$ as the Wiener process from before. We can then associate dW_1 and dW_2 with different observers, or with the information detected and not detected by a single observer.[12] This is precisely the same construction as for photodetection (Section 18.2.8), except now the detector for position information is more abstract. However, the same ideas apply. Taking $\kappa_1 = \eta\kappa$ (with $\eta \in [0,1]$), we can take an ensemble average over all realizations of dW_2 (and then relabel $dW_1 \longrightarrow dW$) to obtain the SME for inefficient detection, where η is the fraction of information actually received by the observer:

$$d\rho = -\kappa[x,[x,\rho]]\,dt + \sqrt{2\eta\kappa}\mathcal{H}[x]\rho\,dW. \tag{19.78}$$

Correspondingly, the measurement record becomes

$$dy = \langle x\rangle\,dt + \frac{dW}{\sqrt{8\eta\kappa}} \tag{19.79}$$

[12] see also A. Barchielli, "Stochastic differential equations and *a posteriori* states in quantum mechanics," *Int. J. Theor. Phys.* **32**, 2221 (1993) (doi: 10.1007/BF00672994); and Jacek Dziarmaga, Diego A. R. Dalvit, Wojciech H. Zurek, "Conditional quantum dynamics with several observers," *Phys. Rev. A* **69**, 022109 (2004) (doi: 10.1103/PhysRevA.69.022109).

for the case of inefficient detection. The bad-detection limit $\eta = 0$ leads to

$$d\rho = -\kappa[x,[x,\rho]]\,dt, \tag{19.80}$$

which is simply the unconditioned master equation for a position measurement. In this case, only the disturbance is left, and we will see shortly that the disturbance term here corresponds to momentum diffusion.

Similarly, the SME (19.76) represents the evolution for position measurement by two observers that do not share information, but from the point of view of an omniscient observer. Observers 1 and 2 thus have their own SMEs, given by tracing out the other noise process,

$$\begin{aligned} d\rho_1 &= -\kappa_1[x,[x,\rho_1]]\,dt - \kappa_2[x,[x,\rho_1]]\,dt + \sqrt{2\kappa_1}\mathcal{H}[x]\rho_1\,dW_1' \\ d\rho_2 &= -\kappa_1[x,[x,\rho_2]]\,dt - \kappa_2[x,[x,\rho_2]]\,dt + \sqrt{2\kappa_1}\mathcal{H}[x]\rho_2\,dW_2'. \end{aligned} \tag{19.81}$$

Of course, the *disturbance* for both measurement processes are present, independent of the ensemble averages. The corresponding measurement records for each observer are

$$\begin{aligned} dy_1 &= \langle x\rangle_1\,dt + \frac{dW_1'}{\sqrt{8\kappa_1}} \\ dy_2 &= \langle x\rangle_2\,dt + \frac{dW_2'}{\sqrt{8\kappa_2}}. \end{aligned} \tag{19.82}$$

The Wiener processes $dW_{1,2}'$ for the individual observers are in general different from the corresponding processes $dW_{1,2}$ from the omniscient observer, because in the individual master equations, the expectation values are taken with respect to $\rho_{1,2}$ rather than ρ. Equating $(\kappa_1\,dy_1 + \kappa_2\,dy_2)/\kappa$ with dy,

$$\frac{\kappa_1}{\kappa}\langle x\rangle_1\,dt + \frac{\kappa_2}{\kappa}\langle x\rangle_2\,dt + \frac{\sqrt{\kappa_1}\,dW_1'}{\sqrt{8}\,\kappa} + \frac{\sqrt{\kappa_2}\,dW_2'}{\sqrt{8}\,\kappa} = \langle x\rangle + \frac{dW}{\sqrt{8\kappa}}, \tag{19.83}$$

we can split up the two terms on the right-hand side consistently and separate the parts depending on κ_1 and κ_2, so that for example

$$\frac{\kappa_1}{\kappa}\langle x\rangle_1\,dt + \frac{\sqrt{\kappa_1}\,dW_1'}{\sqrt{8}\,\kappa} = \frac{\kappa_1}{\kappa}\langle x\rangle\,dt + \sqrt{\frac{\kappa_1}{\kappa}}\frac{dW}{\sqrt{8\kappa}}, \tag{19.84}$$

and find

$$\begin{aligned} dW_1' &= \sqrt{8\kappa_1}\,[\langle x\rangle - \langle x\rangle_1]\,dt + dW_1 \\ dW_2' &= \sqrt{8\kappa_2}\,[\langle x\rangle - \langle x\rangle_2]\,dt + dW_2 \end{aligned} \tag{19.85}$$

for the relation between the noise sources of the individual observers to the noise sources of the omniscient observer.

19.3.4 Interpretation

To better see the effects of the measurement terms in the SME (19.72), we will look at the evolution of the first- and second-order moments, as we did in the case of photodetection. Again deriving the equation of motion for the expectation value of an arbitrary operator A, essentially by taking Eq. (18.131) and setting $a \longrightarrow x$, we obtain (now including Hamiltonian evolution)

$$d\langle A\rangle = -\frac{i}{\hbar}\langle[A,H]\rangle\,dt - \kappa\langle[x,[x,A]]\rangle\,dt + \sqrt{2\eta\kappa}\Big[\,\langle[x,A]_+\rangle - 2\langle A\rangle\langle x\rangle\,\Big]\,dW. \tag{19.86}$$

Assuming for simplicity a Hamiltonian evolution according to the harmonic-oscillator Hamiltonian

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2, \tag{19.87}$$

it follows that the means and variances obey the evolution equations[13]

$$\begin{aligned}
d\langle x\rangle &= \frac{1}{m}\langle p\rangle\, dt + \sqrt{8\eta\kappa}V_x\, dW \\
d\langle p\rangle &= -m\omega_0^2\langle x\rangle\, dt + \sqrt{8\eta\kappa}C_{xp}\, dW \\
\partial_t V_x &= \frac{2}{m}C_{xp} - 8\eta\kappa {V_x}^2 \\
\partial_t V_p &= -2m\omega_0^2 C_{xp} + 2\hbar^2\kappa - 8\eta\kappa C_{xp}^2 \\
\partial_t C_{xp} &= \frac{1}{m}V_p - m\omega_0^2 V_x - 8\eta\kappa V_x C_{xp},
\end{aligned} \tag{19.88}$$

where again the variances V_x and V_p are defined by $V_\alpha := \langle \alpha^2\rangle - \langle\alpha\rangle^2$, and the symmetrized covariance is $C_{xp} := (1/2)\langle [x,p]_+\rangle - \langle x\rangle\langle p\rangle$. In deriving these evolution equations, we have again explicity assumed a Gaussian state, for which the moments obey [see Eqs. (18.136)]

$$\begin{aligned}
\langle x^3\rangle &= 3\langle x\rangle\, V_x + \langle x\rangle^3 \\
\tfrac{1}{2}\langle [x,p^2]_+\rangle &= 2\langle p\rangle\, C_{xp} + \langle x\rangle\left[V_p + \langle p\rangle^2\right] \\
\tfrac{1}{4}\langle [x,[x,p]_+]_+\rangle &= 2\langle x\rangle\, C_{xp} + \langle p\rangle\left[V_x + \langle x\rangle^2\right].
\end{aligned} \tag{19.89}$$

These relations explicitly decouple the means and variances from the higher-order moments. Also, since the Gaussian state is preserved both by the harmonic Hamiltonian evolution and by the position measurement (which amounts to a Gaussian collapse at each instant in time), there is no loss of generality involved in assuming a Gaussian state provided that the system *starts* in a Gaussian state, and even if the system starts in some *other* state, the position measurement will eventually force the system into a Gaussian state.

In examining Eqs. (19.88), we can simply use the coefficients to identify the source and thus the interpretation of each term. The first term in each equation is due to the natural Hamiltonian evolution of the harmonic oscillator. Terms originating from the $\mathcal{D}[c]\rho$ component are proportional to $\kappa\, dt$ but not η; in fact, the only manifestation of this term is the $2\hbar^2\kappa$ term in the equation of motion for V_p. Thus, a position measurement with rate constant k produces momentum diffusion (heating) at a rate $2\hbar^2\kappa$, as is required to maintain the uncertainty principle as the position uncertainty contracts due to the measurement. (This can also be seen by deriving the Fokker–Planck equation for the Wigner function; see Problem problem:fokker-planck-x-measurement.)

There are more terms here originating from the $\mathcal{H}[c]\rho$ component of the master equation, and they are identifiable since they are proportional to either $\sqrt{\eta\kappa}$ or $\eta\kappa$. The dW terms in the equations for $\langle x\rangle$ and $\langle p\rangle$ represent the stochastic nature of the position measurement. That is, during each small time interval, the wave function collapses slightly, but we don't know exactly where it collapses to. The stochastic term in the $\langle x\rangle$ equation is proportional to V_x, since the larger the variance, the wider the range of potential collapses. The stochastic term in the $\langle p\rangle$ equation is proportional to C_{xp}, since a position measurement only induces momentum collapses if x and p are correlated ($C_{xp} \neq 0$). This stochastic behavior is precisely the same behavior that we saw in Eq. (19.60).

The more subtle point here lies with the nonstochastic terms proportional to $\eta\kappa$, which came from the second-order term, as in the last term of

$$dV_x = d\langle X^2\rangle - 2\langle X\rangle\, d\langle X\rangle - (d\langle X\rangle)^2, \tag{19.90}$$

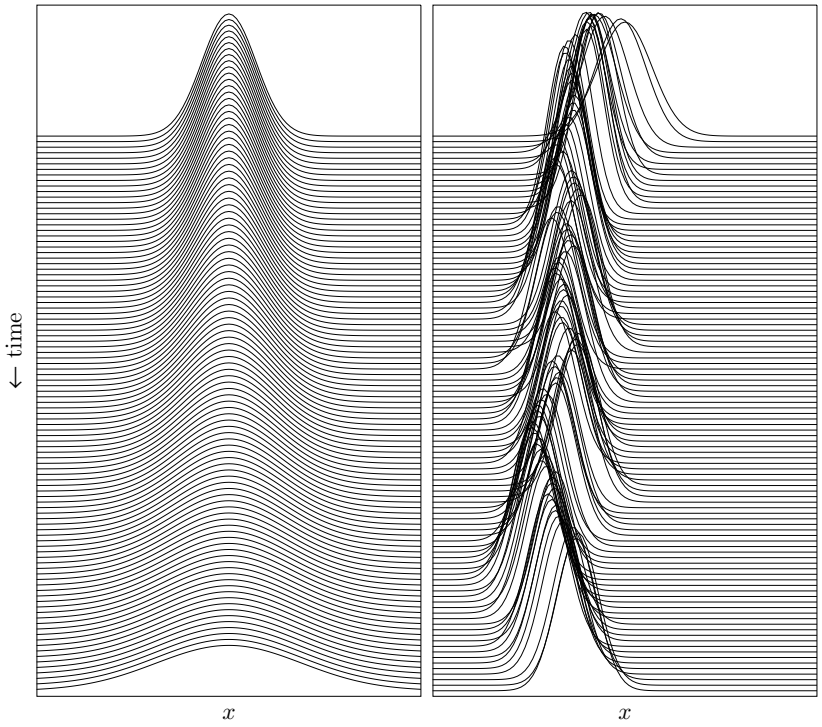
where Itō calculus generates a nonstochastic term from $dW^2 = dt$. Notice in particular the term of this form in the V_x equation, which acts as a *damping* term for V_x. This term represents the certainty gained via the measurement process. The other similar terms are less clear in their interpretation, but they are

[13] A. C. Doherty and K. Jacobs, "Feedback control of quantum systems using continuous state estimation," *Physical Review A* **60**, 2700 (1999) (doi: 10.1103/PhysRevA.60.2700).

necessary to maintain consistency of the evolution. Essentially, again V_p and C_{xp} are only modified by a position measurement, which you might expect would only influence V_x, if x and p are correlated. Hence the presence of C_{xp} in these terms.

Note that we have made the assumption of a Gaussian initial state in deriving these equations, but this assumption is not very restrictive. Due to the linear potential and the Gaussian collapse operators, these equations of motion preserve the Gaussian form of the initial state. The Gaussian collapses additionally converts arbitrary initial states into Gaussian states at long times. Furthermore, as we have mentioned, the assumption of a Gaussian measurement is not restrictive—under the assumption of sufficiently high noise bandwidth, the central-limit theorem guarantees that temporal coarse-graining yields Gaussian noise for *any* measurement process giving random deviates with bounded variance.

As a simple example to illustrate the conditioned dynamics, below is plotted the evolution of a free particle (with Hamiltonian $H = p^2/2m$, in units where $m = \hbar = 1$), with and without a continuous position measurement. The time runs from $t = 0$ to $t = 10$, and the measurement strengths are $k = 0$ (left) and $k = 0.1$ (right). In the unconditioned case, the initial variance $V_x = 2$ spreads ballistically, characteristic of a free particle. In the measured case, the initial variance $V_x = 2$ *contracts* due to the measurement, because position becomes more certain under the measurement process, until dispersion and the measurement balance. The centroid also moves stochastically due to the random nature of the measurement process. Of course, the stochastic motion is different for each possible realization of the measurement process.



19.3.5 Linear Stochastic Evolution Equations

Recall from Section 19.1.2 that the infinitesimal transformation $\rho \longrightarrow \Omega(dW)\rho\Omega^\dagger(dW)$, with collapse operator

$$\Omega(dW) = 1 - i\frac{H}{\hbar}\,dt - \frac{1}{2}c^\dagger c\,dt + c\,dW, \tag{19.91}$$

leads to the *unnormalized* evolution equation

$$d\tilde{\rho} = -\frac{i}{\hbar}[H,\tilde{\rho}]\,dt + \mathcal{D}[c]\tilde{\rho}\,dt + \left(c\tilde{\rho} + \tilde{\rho}c^\dagger\right)\,dW, \tag{19.92}$$

where $\tilde{\rho}$ is the unnormalized density operator. Adding the nonlinear term $-\langle c + c^\dagger \rangle \, \rho \, dW$ restores the proper trace of ρ, and thus leads to the familiar normalized SME

$$\begin{aligned} d\rho &= -\frac{i}{\hbar}[H,\rho]\,dt + \mathcal{D}[c]\rho\,dt + \left(c\rho + \rho c^\dagger\right)\,dW - \langle c + c^\dagger \rangle \, \rho \, dW \\ &= -\frac{i}{\hbar}[H,\rho]\,dt + \mathcal{D}[c]\rho\,dt + \mathcal{H}[c]\rho\,dW. \end{aligned} \tag{19.93}$$

This linear SME is clearly equivalent to the linear SSE

$$d|\tilde{\psi}\rangle = -i\frac{H}{\hbar}|\tilde{\psi}\rangle\,dt - \frac{1}{2}c^\dagger c|\tilde{\psi}\rangle\,dt + c|\tilde{\psi}\rangle\,dW, \tag{19.94}$$

where $|\tilde{\psi}\rangle$ is the unnormalized state vector, since this corresponds to the infinitesimal transformation $|\tilde{\psi}\rangle \longrightarrow \Omega(dW)|\tilde{\psi}\rangle$. The corresponding normalized SSE reads

$$d|\psi\rangle = -\frac{i}{\hbar}H|\psi\rangle\,dt - \frac{1}{2}\left[c^\dagger c - \langle c + c^\dagger \rangle \, c + \frac{1}{4}\langle c + c^\dagger \rangle^2\right]|\psi\rangle\,dt + \left[c - \frac{1}{2}\langle c + c^\dagger \rangle\right]|\psi\rangle\,dW, \tag{19.95}$$

as we saw before for homodyne detection, Eq. (18.74). We may regard the corresponding measurement record (up to an arbitrary factor) in either case to be

$$dr(t) = \langle c + c^\dagger \rangle \, dt + dW. \tag{19.96}$$

Despite the fact that it does not preserve the norm of the state, the linear SME (19.92) and linear SSE (19.94) are still useful—provided that they are interpreted properly—precisely because they are linear and thus facilitate analytic solutions.[14]

19.3.5.1 Norm of the Linear Solution

To see the effect of using the unnormalized equation, consider the evolution of the norm of the state in an infinitesimal time interval, assuming the state is initially normalized:

$$\mathrm{Tr}[\tilde{\rho}(t+dt)] = \mathrm{Tr}[\Omega(dW)\rho(t)\Omega^\dagger(dW)]. \tag{19.97}$$

That is, the norm is just the *probability that the outcome labeled by the value of dW occurred*. The normalization factors for evolution in subsequent time intervals simply multiply, and so for an initially normalized state at $t = 0$, the normalization (trace) of the final state $\tilde{\rho}(t)$ *gives the probability that the particular realization $dW(t)$ of the measurement actually occurred*.

19.3.5.2 Interpretation of the Solution

However, for the same realization $dW(t)$, the evolution according to the linear SSE (SME) is *not* equivalent to evolution according to the normalized SSE (SME). We can see this because the linear evolution is given in terms of the transformation operator $\Omega(dW)$ (19.91), which is labeled solely by the stochastic variable dW, which is *independent of the state*. By contrast, the *normalized* evolution of the state corresponds to collapse operators $\Omega(dr)$, which are labeled by the measurement result $dr(t)$ (19.96), which depends on the state via the expectation value. The linear evolution thus corresponds to taking $dr(t) = dW$, that is, *shifting* the measurement results by the amount $\langle c + c^\dagger \rangle$. Thus, the unnormalized evolution according to the linear SSE (SME), after renormalization at the end, corresponds to the correct *normalized* evolution for *some* possible measurement realization $dr(t)$, just not the one given by Eq. (19.96). In fact, the measurement record corresponding to the linear evolution is generically much less likely than, say, the realization according to Eq. (19.96), which occurs with the same probability as that for choosing the realization of $dW(t)$. However,

[14] Peter Goetsch and Robert Graham, "Linear stochastic wave equations for continuously measured quantum systems," *Physical Review A* **50**, 5242 (1994) (doi: 10.1103/PhysRevA.50.5242); H. M. Wiseman, "Quantum trajectories and quantum measurement theory," *Quantum and Semiclassical Optics* **8**, 205 (1996) (doi: 10.1088/1355-5111/8/1/015).

as we argued above, the *norm* of the unnormalized state gives the probability that this realization $dr(t) = dW$ actually occurred, and so as long as we weight these trajectories appropriately in any ensemble average, we have no problem.

Thus, to summarize: linear SSEs and SMEs give different conditioned evolution than their normalized counterparts, and in fact they realize particular evolutions with the wrong probability (assuming $dW(t)$ is chosen with the proper probability), but the probability for realizing any particular simulated trajectory is given by the final norm of the state. Simulation according to the normalized equations always gives solutions that are equally likely if $dW(t)$ is chosen with the proper probability. To see another example of this in terms of the position measurement, recall from Eq. (19.64) that the infinitesimal evolution under a continuous position measurement is given by

$$|\tilde{\psi}(t+dt)\rangle = \left[1 - \left(\kappa x^2 - 4\kappa x\langle x\rangle\right) dt + \sqrt{2\kappa}\, x\, dW\right] |\psi(t)\rangle, \tag{19.98}$$

where, although the result is not normalized, the result is chosen with the correct probability by our construction. The corresponding position measurement record was

$$dy = \langle x\rangle\, dt + \frac{dW}{\sqrt{8\kappa}}, \tag{19.99}$$

in terms of which we can write the above evolution as

$$|\tilde{\psi}(t+dt)\rangle = \left[1 - \kappa x^2\, dt + 4\kappa\, x\, dy\right] |\psi(t)\rangle. \tag{19.100}$$

Now if we make the replacement $dy \longrightarrow dW/\sqrt{8\kappa}$, that is we shift the mean of the distribution of position-measurement results, we obtain the infinitesimal evolution

$$|\tilde{\psi}(t+dt)\rangle = \left[1 - \kappa x^2\, dt + \sqrt{2\kappa}\, x\, dW\right] |\psi(t)\rangle. \tag{19.101}$$

corresponding to the linear SSE for position measurement,

$$d|\tilde{\psi}\rangle = -\kappa x^2 |\tilde{\psi}\rangle\, dt + \sqrt{2\kappa}\, x|\tilde{\psi}\rangle\, dW, \tag{19.102}$$

or equivalently the SME

$$d\tilde{\rho} = -\kappa[x,[x,\rho]]\, dt + \sqrt{2\kappa}\left(x\tilde{\rho} + \tilde{\rho}x\right)\, dW \tag{19.103}$$

for position measurement, equivalent to the general forms (19.102) and (19.103) with $c = \sqrt{2\kappa}\, x$.

19.3.5.3 Explicit Solutions of Measurement Dynamics

In the linear form, then, analytic solutions become more tractable. The solution to the linear SSE (19.94) is thus given by noting that

$$e^{a\, dt + b\, dW} = 1 + a\, dt + b\, dW + \frac{1}{2} b^2\, dW^2 = 1 + \left(a + \frac{b^2}{2}\right)\, dt + b\, dW, \tag{19.104}$$

so that composing infinitesimal exponential evolution operations gives

$$|\tilde{\psi}(t)\rangle = \exp\left[-\frac{i}{\hbar}\int_0^t H(t')\, dt' - \frac{1}{2}\int_0^t \left(c^\dagger c + c^2\right)\, dt' + \int_0^t c\, dW(t')\right] |\psi(0)\rangle \tag{19.105}$$

if c and H commute, which in the case of having no explicit time dependence of c (i.e., the measurement is not changed with time) and a time-independent Hamiltonian becomes

$$|\tilde{\psi}(t)\rangle = \exp\left[-\frac{iHt}{\hbar} - \frac{1}{2}\left(c^\dagger c + c^2\right)\, t + c\, W(t)\right] |\psi(0)\rangle. \tag{19.106}$$

Recall that $W(t)$ is a Gaussian random variable with zero mean and variance t, so that we do not need the entire history $dW(t)$ to compute the final solution.

Now consider a **nondemolition measurement** of a Hermitian observable operator Q (i.e., the measurement leaves eigenstates of Q in eigenstates of Q), so that we let $c = \sqrt{2\kappa}Q$ in the above equations, and have a corresponding measurement record

$$dy = \langle Q \rangle + \frac{dW}{\sqrt{8\kappa}} \tag{19.107}$$

for the conditioned evolution. Then the final state is

$$|\tilde{\psi}(t)\rangle = \exp\left[-\frac{iHt}{\hbar} - 2\kappa Q^2\, t + \sqrt{2\kappa}\, Q\, W(t)\right] |\psi(0)\rangle. \tag{19.108}$$

Decomposing the initial state into eigenstates $|q\rangle$ of Q,

$$|\psi(0)\rangle = \sum_q c_q |q\rangle, \tag{19.109}$$

we may write the final state as

$$|\tilde{\psi}(t)\rangle = \sum_q c_q \exp\left[-\frac{iE_q t}{\hbar} - 2\kappa q^2\, t + \sqrt{2\kappa}\, q\, W(t)\right] |q\rangle, \tag{19.110}$$

where $H|q\rangle = E_q|q\rangle$. Now recall that the probability of this particular outcome is given by the norm of the final state

$$\langle\tilde{\psi}(t)|\tilde{\psi}(t)\rangle = \sum_q |c_q|^2 \exp\left[-4\kappa q^2\, t + \sqrt{8\kappa}\, q\, W(t)\right]. \tag{19.111}$$

or in terms of the rescaled Wiener process

$$Y(t) := \frac{W(t)}{\sqrt{8\kappa}\, t}, \tag{19.112}$$

the norm becomes

$$\langle\tilde{\psi}(t)|\tilde{\psi}(t)\rangle = \sum_q |c_q|^2 \exp\left[-4\kappa q^2\, t + 8\kappa t q\, Y(t)\right]. \tag{19.113}$$

But the actual probability for realizing $Y(t)$ is the probability for picking $Y(t)$ in a simulation (i.e., the corresponding Gaussian probability for picking $W(t)$), multiplied by the norm of the final state:

$$\begin{aligned} P(Y,t) &= \frac{1}{\sqrt{2\pi t}} e^{-W^2/2t} \frac{dW}{dY} \langle\tilde{\psi}(t)|\tilde{\psi}(t)\rangle \\ &= \sqrt{\frac{4\kappa t}{\pi}} e^{-4\kappa t Y^2} \sum_q |c_q|^2 \exp\left[-4\kappa q^2\, t + 8\kappa t q\, Y(t)\right] \\ &= \sqrt{\frac{4\kappa t}{\pi}} \sum_q |c_q|^2 e^{-4\kappa t (Y-q)^2}. \end{aligned} \tag{19.114}$$

That is, the probability distribution for $Y(t)$ is a sum of Gaussians of width $1/\sqrt{8\kappa t}$, centered about each eigenvalue q and weighted by the usual Born probability c_q^2. But recalling that using the *linear* SSE amounts to choosing the measurement record with the expectation value of the observable removed,

$$dy = \frac{dW}{\sqrt{8\kappa}}, \tag{19.115}$$

and thus we interpret

$$Y(t) = \frac{1}{t}\int_0^t dy(t') \tag{19.116}$$

as simply the time-averaged measurement result. Thus, we have shown in the case where there is no real interplay between the measurement and the Hamiltonian evolution, we can solve the measurement evolution explicitly and see that the time average of the measurement record gives the observed value of Q for a particular trajectory (particular experiment). This observed value converges (almost always) to an eigenvalue of the discrete observable as $1/\sqrt{t}$.

19.4 Imaged Resonance Fluorescence as a Position Measurement

Now we will consider several physical examples of position measurements, taking a single atom as a concrete example of a quantum-mechanical particle. The first one we will consider is the case of photodetection of ordinary resonance fluorescence, where the atom is excited by a plane wave, as we have already treated in Chapter 18. However, to gain position information, it is not sufficient to simply *detect* the photons, you have to *image* the scattered light, just as you would use a camera or a microscope to locate a small object.

19.4.1 Center-of-Mass Dynamics

We can write the SME for spontaneous emission as usual as

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\sigma]\rho\,dN, \tag{19.117}$$

with superoperators

$$\begin{aligned}\mathcal{J}[c]\rho &:= \left(\frac{c\rho c^\dagger}{\langle c^\dagger c\rangle} - \rho\right)\\ \mathcal{H}[c]\rho &:= c\rho + \rho c^\dagger - \left\langle c + c^\dagger\right\rangle \rho,\end{aligned} \tag{19.118}$$

and a Poisson process characterized by

$$\langle\!\langle dN\rangle\!\rangle = \Gamma\left\langle\sigma^\dagger\sigma\right\rangle dt. \tag{19.119}$$

Now we want to consider how the evolution of the atomic internal state influences the atomic center-of-mass motion. We need to explicitly include the mechanical effects of the resonance fluorescence. First, we will model the situation of *angle-resolved* photodetection, where we break up the Poisson process dN into many infinitesimal Poisson processes $dN(\theta,\phi)/d\Omega$, corresponding to emission in any possible direction (θ,ϕ). Additionally, if the photon is detected in the direction $\mathbf{k}$, then the atom must recoil with momentum $-\hbar\mathbf{k}$, which is equivalent to applying the operator

$$\sigma e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{19.120}$$

to the atomic state at each detection event, rather than just the lowering operator σ. Thus, for the term describing each subprocess $dN(\theta,\phi)/d\Omega$, we make the replacement

$$\sigma \longrightarrow \sigma e^{-i\mathbf{k}\cdot\mathbf{r}}, \tag{19.121}$$

where $\mathbf{k}$ points along the direction (θ,ϕ), and then sum over all angles to obtain

$$d\rho = -\frac{i}{\hbar}[H,\rho]\,dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \int d\Omega\,\mathcal{J}\left[\sigma e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\rho\;\frac{dN(\theta,\phi)}{d\Omega}. \tag{19.122}$$

(Compare to the form (5.404) of the atom–field interaction Hamiltonian, where the spatial dependence of the field enters via the Rabi frequency.) The Poisson processes are characterized by the means

$$\langle\!\langle\frac{dN(\theta,\phi)}{d\Omega}\rangle\!\rangle = \Gamma\left\langle\sigma^\dagger\sigma\right\rangle f_{\hat{\varepsilon}}(\theta,\phi)\,dt, \tag{19.123}$$

where $f_{\hat{\varepsilon}}(\theta,\phi)$ is the dipole radiation pattern for a dipole unit vector of $\hat{\varepsilon}$. In the Weisskopf–Wigner treatment of spontaneous emission in Chapter 11, we obtained this master equation by simply accounting for the spatial dependence of the vacuum field modes by letting

$$g_{\mathbf{k}} \longrightarrow g_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{19.124}$$

in the atom–field interaction Hamiltonian to obtain

$$H_{\text{AF}} = \sum_{\mathbf{k},\zeta} \hbar \left(g_{\mathbf{k},\zeta} a_{\mathbf{k},\zeta} \sigma^\dagger e^{i\mathbf{k}\cdot\mathbf{r}} + g^*_{\mathbf{k},\zeta} a^\dagger_{\mathbf{k},\zeta} \sigma e^{-i\mathbf{k}\cdot\mathbf{r}} \right), \tag{19.125}$$

and then expanding the state vector also in the momentum basis

$$|\psi\rangle = \int d^3p\, \psi_{\text{e}}(\mathbf{p}) |\mathbf{p}, \text{e}\rangle + \int d^3p \sum_{\mathbf{k},\zeta} \psi_{\mathbf{k},\zeta}(\mathbf{p}) |\mathbf{p}, \text{g}, 1_{\mathbf{k},\zeta}\rangle. \tag{19.126}$$

In the weak-excitation limit, we can take the magnitude of $\mathbf{k}$ to have the value of an externally applied probe field, which will generally be near enough to resonance that it will match the resonant wave number (if the detuning is very large, the Rayleigh-scattered photons are elastically scattered from the incident field and thus have the same wave number as the driving field). Recall that the measurement terms only account for the momentum recoil on *emission*; any additional recoil due to photon absorption is already accounted for by the Hamiltonian evolution (see, e.g., Section 5.8.6.6).

We can simplify the angle-resolved SME by carrying out the angular integral, defining dN to be one whenever $\max[dN(\theta,\phi)] = 1$. The result is

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\, dt + \mathcal{J}\left[\sigma e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\rho\, dN, \tag{19.127}$$

with

$$\langle\langle dN \rangle\rangle = \Gamma \left\langle \sigma^\dagger\sigma \right\rangle dt \tag{19.128}$$

as before. The angles θ and ϕ are then stochastic variables with probability density $f(\theta,\phi)\sin\theta$.

19.4.2 Imaging

The above master equation (19.122) is for an *angle-resolving* detector. What we see is that angle-resolved detection keeps explicit track of the atomic momentum kicks due to spontaneous emission. An *imaging* detector, on the other hand, gives up resolution of the direction of the emitted photon wave vector $\mathbf{k}$, thus obtaining instead some position information about the atom. An imaging system operates by summing fields from many directions together and then detecting the resulting interference pattern. The procedure for obtaining the measurement operators for the imaging system is as follows.[15] Notice that we can regard the master equation (19.122) as a normal jump process of the form (19.117), with measurement operators

$$\sigma(\theta,\phi) = \sqrt{f(\theta,\phi)}\, \sigma e^{-ikz\cos\theta}, \tag{19.129}$$

where we sum over all possible emission angles. In writing down this family of operators, we are specializing to one-dimensional motion along the z-axis ($x = y = 0$), so we only require the z-component $k\cos\theta$ of $\mathbf{k}$. This operator ranges from -1 to 1 in $\cos\theta$ and from 0 to 2π in ϕ. Thus, we can write down Fourier coefficients (operators) for $\sigma(\theta,\phi)$, since these functions are defined on a bounded domain, with two indices α and β:

$$\tilde{\sigma}_{\alpha\beta} = \frac{\sigma}{\sqrt{4\pi}} \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, \sqrt{f(\theta,\phi)}\, e^{-ikz\cos\theta} e^{i\alpha\pi\cos\theta} e^{i\beta\phi}. \tag{19.130}$$

[15] M. Holland, S. Marksteiner, P. Marte, and P. Zoller, "Measurement Induced Localization from Spontaneous Decay," *Physical Review Letters* **76**, 3683 (1996) (doi: 10.1103/PhysRevLett.76.3683); W. Greenwood, P. Pax, and P. Meystre, "Atomic transport on one-dimensional optical lattices," *Physical Review A* **56**, 2109 (1997) (doi: 10.1103/PhysRevA.56.2109). For one experimental implementation, see Hidetoshi Katori, Stefan Schlipf, and Herbert Walther, "Anomalous Dynamics of a Single Ion in an Optical Lattice," *Physical Review Letters* **79**, 2221 (1997) (doi: 10.1103/PhysRevLett.79.2221).

However, this expression corresponds to imaging via an ideal imaging system, where the aperture extends over the full 4π solid angle (requiring, for example, arbitrarily large lenses on either side of the atom). In practice it is rare to come anywhere close to this extreme. Thus, we include the effects of an aperture that only allows the imaging system to detect radiated light within a limited solid angle. We thus take the *intensity* transmission of the aperture to be represented by the function $T(\theta, \phi)$, so that we explicitly ignore any phase-shift effects of the aperture. The collapse operator for imaged detection is thus modified to be the Fourier transform of the angular distribution, which now includes the aperture function:

$$\tilde{\sigma}_{\alpha\beta} = \frac{\sigma}{\sqrt{4\pi}} \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, \sqrt{T(\theta,\phi) f(\theta,\phi)}\, e^{-ikz\cos\theta} e^{i\alpha\pi\cos\theta} e^{i\beta\phi}. \tag{19.131}$$

Of course, any phase mask could be modeled by introducing a factor $\exp[i\varphi(\theta, \phi)]$ in the above integrand. Now consider the ϕ part of the above integral, which we may write as

$$a_\beta := \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} d\phi\, \sqrt{T(\theta,\phi) f(\theta,\phi)}\, e^{i\beta\phi}. \tag{19.132}$$

Note that in this part of the measurement operator, there is no position dependence, and thus we will be able to eliminate it from the dynamics. With our normalization convention, we have chosen our normalized basis functions as $e^{i\beta\phi}/\sqrt{2\pi}$, and thus

$$\sum_{\beta=-\infty}^{\infty} e^{i\beta(\phi-\phi')} = 2\pi\delta(\phi - \phi'), \tag{19.133}$$

where the argument of the delta function is taken modulo 2π, so that we have the overall normalization

$$\sum_{\beta=-\infty}^{\infty} a_\beta^\dagger a_\beta = \int_0^{2\pi} d\phi\, T(\theta,\phi) f(\theta,\phi). \tag{19.134}$$

If we trace over the measurement result β, this amounts to using the reduced density operator

$$\sum_{\beta=-\infty}^{\infty} a_\beta \rho a_\beta^\dagger = \rho \int_0^{2\pi} d\phi\, T(\theta,\phi) f(\theta,\phi), \tag{19.135}$$

which is equivalent to making the replacement

$$a_\beta \longrightarrow \sqrt{\tilde{T}(\theta)} \tag{19.136}$$

in the operator $\tilde{\sigma}_{\alpha\beta}$, where we now have the *effective* aperture

$$\tilde{T}(\theta) := \int_0^{2\pi} d\phi\, T(\theta,\phi) f(\theta,\phi). \tag{19.137}$$

Thus, the measurement operator (19.131) loses the irrelevant index β, and reduces to

$$\tilde{\sigma}_\alpha = \mathscr{N} A(z - \alpha\lambda/2)\sigma, \tag{19.138}$$

where the effect on the position degree of freedom due to the photodetection is given by the operator

$$A(z) := \frac{1}{\mathscr{N}\sqrt{2}} \int_{-1}^{1} d(\cos\theta)\, \sqrt{\tilde{T}(\theta)}\, e^{-ikz\cos\theta}, \tag{19.139}$$

and $\mathscr{N}$ is a normalization constant that does not influence the effect of the operator. The $A(z)$ operators, for ''reasonable'' apertures, contain localized functions of the position z, and thus correspond to position

measurements. Again, the effect of the operator $\tilde{\sigma}_\alpha$ is equivalent to that of the original form $\tilde{\sigma}_{\alpha\beta}$, but with a trace over β:

$$\tilde{\sigma}_\alpha \rho \tilde{\sigma}_\alpha^\dagger = \sum_{\beta=-\infty}^{\infty} \tilde{\sigma}_{\alpha\beta} \rho \tilde{\sigma}_{\alpha\beta}^\dagger. \tag{19.140}$$

The idea here is that for motion along the z-axis, photons going into *any* azimuthal angle ϕ are equivalent as far as providing position information about the atom. Thus, the θ dependence of the aperture will be most important, but the ϕ dependence gives some *effective* θ dependence if the aperture is not separable in (θ, ϕ).

Notice that with the normalization convention for the Fourier coefficients here, if we remove the aperture by setting $T(\theta, \phi) = 1$, we have

$$\int d\Omega\, \sigma^\dagger(\theta,\phi)\sigma(\theta,\phi) = \sum_\alpha \tilde{\sigma}_\alpha^\dagger \tilde{\sigma}_\alpha = \sigma^\dagger \sigma, \tag{19.141}$$

so that the set of measurement operators is complete and properly normalized in either basis. An arbitrary aperture mask will then reduce the efficiency of the measurement, since not all of the photons will be detected. In this case,

$$\sum_\alpha \tilde{\sigma}_\alpha^\dagger \tilde{\sigma}_\alpha = \eta_\Omega \sigma^\dagger \sigma, \tag{19.142}$$

where we have defined the detection efficiency of the angular aperture by

$$\eta_\Omega := \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta)\, T(\theta,\phi) f(\theta,\phi) = \int_{-1}^{1} d(\cos\theta)\, \tilde{T}(\theta). \tag{19.143}$$

While this is the efficiency for photon detection, we will see that in general this is *not* the same as the efficiency for *information gain*.

If we then choose the normalization $\mathscr{N} = \sqrt{\eta_\Omega}$, we will thus have the detection operator

$$\tilde{\sigma}_\alpha = \sqrt{\eta_\Omega} A(z - \alpha\lambda/2)\sigma, \tag{19.144}$$

where the effect on the position degree of freedom due to the photodetection is given by the operator

$$A(z) = \frac{1}{\sqrt{2\eta_\Omega}} \int_{-1}^{1} d(\cos\theta)\, \sqrt{\tilde{T}(\theta)}\, e^{-ikz\cos\theta}, \tag{19.145}$$

so that we associate the efficiency of the detection explicitly with the collapse operator, and now in view of the normalization (19.142), the operators $A(z)$ form a POVM.

We can thus get the imaged-detection SME from the angle-resolved form (19.122) by first separating the angular part of the measurement term according to what fraction of light at a given angle makes it through the aperture:

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\, dt + \int d\Omega\, \mathcal{J}\left[\sigma e^{-ikz\cos\theta}\right]\rho\, \frac{dN_1(\theta,\phi)}{d\Omega} + \int d\Omega\, \mathcal{J}\left[\sigma e^{-ikz\cos\theta}\right]\rho\, \frac{dN_2(\theta,\phi)}{d\Omega}. \tag{19.146}$$

Here, $dN_1(\theta,\phi)/d\Omega$ enumerates the processes by which a photon is detected through the aperture, satisfying

$$\langle\!\langle \frac{dN_1(\theta,\phi)}{d\Omega} \rangle\!\rangle = \Gamma \langle \sigma^\dagger \sigma \rangle\, T(\theta,\phi) f(\theta,\phi)\, dt, \tag{19.147}$$

and $dN_2(\theta,\phi)/d\Omega$ represents the fictitious processes by which the photons that are blocked by the aperture are detected, and thus satisfies

$$\langle\!\langle \frac{dN_2(\theta,\phi)}{d\Omega} \rangle\!\rangle = \Gamma \langle \sigma^\dagger \sigma \rangle\, [1 - T(\theta,\phi)] f(\theta,\phi)\, dt, \tag{19.148}$$

so that taking the two processes together is equivalent to the original model. Eliminating the undetected photons, we take an ensemble average over $dN_2/d\Omega$ to obtain

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\int_{-1}^{1} d(\cos\theta)\,\tilde{R}(\theta)\mathcal{D}\left[\sigma e^{-ikz\cos\theta}\right]\rho\,dt - \eta_\Omega\frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \int d\Omega\,\mathcal{J}\left[\sigma e^{-ikz\cos\theta}\right]\rho\,\frac{dN_1(\theta,\phi)}{d\Omega}, \tag{19.149}$$

where the second term is the quantum backaction due to the *undetected* photons, the third term is the backaction due to the *detected* photons, and the last term represents the gain of measurement information. Here we have defined the angular distribution of *blocked* photons

$$\tilde{R}(\theta) := \int_0^{2\pi} d\phi\,[1 - T(\theta,\phi)]f(\theta,\phi), \tag{19.150}$$

so that the transmitted and blocked distributions add to the ''natural'' radiation pattern when there is no aperture:

$$\tilde{T}(\theta) + \tilde{R}(\theta) = \int_0^{2\pi} d\phi\, f(\theta,\phi). \tag{19.151}$$

Now from our argument relating the position-sensitive operators $\tilde{\sigma}_\alpha$ to the momentum-kicking operators $\sigma e^{-ikz\cos\theta}$, we may rewrite the last term as a sum over σ_α:

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\int_{-1}^{1} d(\cos\theta)\,\tilde{R}(\theta)\mathcal{D}\left[\sigma e^{-ikz\cos\theta}\right]\rho\,dt - \eta_\Omega\frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \sum_{\alpha=-\infty}^{\infty}\mathcal{J}[\tilde{\sigma}_\alpha]\rho\,dN_\alpha. \tag{19.152}$$

This amounts to a unitary transformation on the measurement operators, under which the rest of the master equation is invariant, as implied by the completeness in Eq. (19.140). Here, the Poisson processes are characterized by

$$\langle\langle dN_\alpha\rangle\rangle = \eta_\Omega\Gamma\mathrm{Tr}\left[\sigma^\dagger\sigma|A(z-\alpha\lambda/2)|^2\rho\right]\,dt, \tag{19.153}$$

so that the probability of the measurement outcome α goes as the squared modulus of the overlap of $A(z-\alpha\lambda/2)$ with the center-of-mass part of the atomic wave function $\psi(z)$. Again, we may combine the Poisson processes dN_α into a single process dN_Ω,

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + \Gamma\int_{-1}^{1} d(\cos\theta)\,\tilde{R}(\theta)\mathcal{D}\left[\sigma e^{-ikz\cos\theta}\right]\rho\,dt - \eta_\Omega\frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \mathcal{J}[\tilde{\sigma}_\alpha]\rho\,dN_\Omega, \tag{19.154}$$

where the combined Poisson process is characterized by

$$\langle\langle dN_\Omega\rangle\rangle = \eta_\Omega\Gamma\langle\sigma^\dagger\sigma\rangle\,dt, \tag{19.155}$$

and the probability distribution for the integer outcome α is the expectation value

$$P(\alpha) = \int dz\,\langle \mathrm{e}, z|\rho|\mathrm{e}, z\rangle|A(z-\alpha\lambda/2)|^2. \tag{19.156}$$

The expectation value is taken with respect to the excited-state part $|\mathrm{e}\rangle$ of the atomic state, since this is what survives after the reduction $\rho \longrightarrow \sigma\rho\sigma^\dagger/\langle\sigma^\dagger\sigma\rangle$ representing the photon detection. Notice also that the set of possible measurement values is not continuous, but rather is *discretely* spaced by $\lambda/2$, which is rather odd for the result of a position measurement. However, this is the case because the Fourier transform was taken with respect to a bounded domain, and is thus a consequence of the sampling theorem (Chapter 25): the information contained in a function on a bounded, continuous domain is equivalent in the Fourier domain to the information in a function on a discrete (but infinite) domain.

19.4.2.1 Example: 4π Detection

As an example of a particular form for the position-reduction operator $A(z)$, a radiating atomic dipole oriented along the z-axis has

$$f_{\hat{z}}(\theta, \phi) = \frac{3}{8\pi} \sin^2 \theta, \tag{19.157}$$

in which case the effective aperture becomes

$$\tilde{T}(\theta) = \frac{3}{4} \sin^2 \theta. \tag{19.158}$$

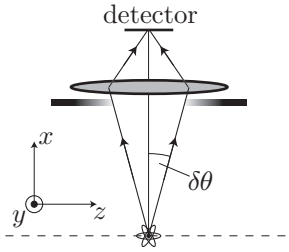
This gives the measurement operator

$$A(z) = \sqrt{\frac{3\pi^2}{8}} \frac{J_1(kz)}{kz}, \tag{19.159}$$

where $J_1(x)$ is an ordinary Bessel function. Since $J_1(x)$ decays as $x^{-1/2}$ for large x, the tails of the collapse operator here decay as $z^{-3/2}$.

19.4.2.2 Example: Small Gaussian Aperture

Often in real situations, the aperture subtends only a small solid angle. Intuitively, one expects a camera imaging system to be most effective when oriented normal to the z-axis, so we choose the aperture to be centered about $\theta = \pi/2$.



We can also arbitrarily take the aperture to be centered about $\phi = 0$. It is mathematically convenient to assume a Gaussian transmission function for the aperture, and we thus take the intensity transmission function of the aperture to be

$$T(\theta, \phi) = \exp\left[-\frac{2(\theta - \pi/2)^2}{(\delta\theta)^2}\right] \exp\left[-\frac{2\phi^2}{(\delta\phi)^2}\right] \tag{19.160}$$

where $\delta\theta$ and $\delta\phi$ are the same for a circular aperture. Now the effective aperture becomes

$$\tilde{T}(\theta) \approx \eta_\phi \exp\left[-\frac{2(\theta - \pi/2)^2}{(\delta\theta)^2}\right], \tag{19.161}$$

where the azimuthal angular integral is

$$\eta_\phi(\theta) := \int_{-\pi}^{\pi} d\phi \exp\left[-\frac{2\phi^2}{(\delta\phi)^2}\right] f(\theta, \phi). \tag{19.162}$$

We can suppress the dependence on θ by assuming that the distribution function $f(\theta, \phi)$ varies slowly over the width $\delta\theta$ of the aperture (in particular, we should not choose the aperture to be near a direction in which the radiation pattern *vanishes*). This happens when the aperture is narrow, and the result is the constant efficiency

$$\eta_\phi \approx \int_{-\pi}^{\pi} d\phi \exp\left[-\frac{2\phi^2}{(\delta\phi)^2}\right] f(\theta = \pi/2, \phi). \tag{19.163}$$

If $f(\theta,\phi)$ is the usual dipole radiation pattern (19.157), then we may take $f(\theta=\pi/2,\phi)=3/8\pi$, and thus

$$\eta_\phi \approx \frac{3}{8\pi}\int_{-\infty}^{\infty} d\phi \exp\left[-\frac{2\phi^2}{(\delta\phi)^2}\right] = \frac{3\delta\phi}{8\sqrt{2\pi}} \tag{19.164}$$

in the limit where $\delta\phi$ is small.

If $\delta\theta$ is small, then the integrand is only appreciable for θ near $\pi/2$ due to the Gaussian factor. Recentering the integrand in Eq. (19.145), making the small-angle approximation in the rest of the integrand, and extending the limits of integration, we find

$$\begin{aligned} A(z) &= \sqrt{\frac{\eta_\phi}{2\eta_\Omega}}\int_{-\pi/2}^{\pi/2} d\theta\, \cos^2\theta\, e^{-ikz\sin\theta}\exp\left[-\frac{\theta^2}{(\delta\theta)^2}\right] \\ &\approx \sqrt{\frac{\eta_\phi}{2\eta_\Omega}}\int_{-\infty}^{\infty} d\theta\, e^{-ikz\theta}\exp\left[-\frac{\theta^2}{(\delta\theta)^2}\right] \\ &= \sqrt{\frac{\pi\eta_\phi}{2\eta_\Omega}}\delta\theta\, \exp\left[-\left(\frac{k\,\delta\theta}{2}\right)^2 z^2\right]. \end{aligned} \tag{19.165}$$

Thus, the measurement operator in this case is also Gaussian. We can write the fraction of photons transmitted by the aperture as

$$\eta_\Omega = \int_{-1}^{1} d(\cos\theta)\,\tilde{T}(\theta) \approx \eta_\phi\sqrt{\frac{\pi}{2}}\,\delta\theta \tag{19.166}$$

in the same regime of small $\delta\theta$, and thus the Gaussian operator becomes

$$A(z) = \sqrt{\sqrt{\frac{\pi}{2}}\delta\theta}\exp\left[-\left(\frac{k\,\delta\theta}{2}\right)^2 z^2\right] \tag{19.167}$$

upon eliminating η_Ω.

19.4.3 Adiabatic Elimination of the Internal Atomic State

So far, we have seen how the internal and external dynamics of the atom are intrinsically linked. Now we want to focus on the external atomic dynamics. To do so, we will take advantage of the natural separation of the time scales of the dynamics. The internal dynamics are damped at the decay rate Γ, which is typically on the order of $\sim 10^7\ \mathrm{s}^{-1}$. The external dynamics are typically much slower, corresponding to kHz or smaller oscillation frequencies for typical laser dipole traps. The adiabatic approximation assumes that the internal dynamics equilibrate rapidly compared to the external dynamics, and are thus always in a quasi-equilibrium state with respect to the external state.

19.4.3.1 Internal Quasi-Equilibrium

At this point, we must consider the internal atomic dynamics more precisely, and as a review we will compactly rederive the necessary steady-state results from Section 5.5.1. A resonant, driving (classical) laser field enters in the usual form

$$\tilde{H}_{\mathrm{AF}} = \frac{\hbar\Omega}{2}\left(\sigma e^{-i\mathbf{k}_\mathrm{D}\cdot\mathbf{r}} + \sigma^\dagger e^{i\mathbf{k}_\mathrm{D}\cdot\mathbf{r}}\right), \tag{19.168}$$

where the *Rabi frequency* Ω characterizes the strength of the laser–atom interaction, and we have included the spatial dependence of a driving plane-wave driving field $\Omega(\mathbf{r}) = \Omega e^{i\mathbf{k}_\mathrm{D}\cdot\mathbf{r}}$ propagating along $\mathbf{k}_\mathrm{D}$. Restricting our attention again to the z-axis, the Hamiltonian becomes

$$\tilde{H}_{\mathrm{AF}} = \frac{\hbar\Omega}{2}\left(\sigma e^{-ik_\mathrm{D} z\cos\theta_\mathrm{D}} + \sigma^\dagger e^{ik_\mathrm{D} z\cos\theta_\mathrm{D}}\right), \tag{19.169}$$

In writing down this interaction, we have made (as before) the standard unitary transformation to a rotating frame where the free atomic Hamiltonian $\tilde{H}_{\rm A} = 0$. Note that if the driving field propagates along a normal to the z-axis, the spatial dependence of the field vanishes in $\tilde{H}_{\rm AF}$.

The usual unconditioned master equation with this interaction, but neglecting the external motion (that is equivalent to the usual, on-resonance optical Bloch equations) is

$$\partial_t \rho = -\frac{i}{\hbar}[\tilde{H}_{\rm D}, \rho] + \Gamma \mathcal{D}[\sigma]\rho. \tag{19.170}$$

This equation implies that the expectation value of an operator A evolves as

$$\partial_t \langle A \rangle = -\frac{i}{\hbar}\Big\langle [A, \tilde{H}_{\rm D}] \Big\rangle + \Gamma \Big\langle \sigma^\dagger A \sigma - \frac{1}{2}[\sigma^\dagger \sigma, A]_+ \Big\rangle . \tag{19.171}$$

This gives the following equations of motion for the density-matrix elements $\rho_{\alpha\beta} := \langle \alpha|\rho|\beta\rangle$:

$$\begin{aligned} \partial_t \rho_{\rm ee} = \partial_t \langle \sigma^\dagger \sigma \rangle &= \frac{i\Omega}{2}\Big[\langle \sigma \rangle e^{-ik_{\rm D} z \cos\theta_{\rm D}} - \langle \sigma^\dagger \rangle e^{ik_{\rm D} z \cos\theta_{\rm D}} \Big] - \Gamma \langle \sigma^\dagger \sigma \rangle , \\ \partial_t \rho_{\rm eg} = \partial_t \langle \sigma \rangle &= \frac{i\Omega e^{ik_{\rm D} z \cos\theta_{\rm D}}}{2} \left(\langle \sigma^\dagger \sigma \rangle - \langle \sigma \sigma^\dagger \rangle \right) - \frac{\Gamma}{2}\langle \sigma \rangle . \end{aligned} \tag{19.172}$$

The remaining matrix elements are determined by $\rho_{\rm ge} = \rho_{\rm eg}^*$ and $\rho_{\rm gg} = \langle \sigma \sigma^\dagger \rangle = 1 - \langle \sigma^\dagger \sigma \rangle$. Setting the time derivatives to zero, we can solve these equations to obtain (as we already derived in Section 5.5.1)

$$\begin{aligned} \langle \sigma^\dagger \sigma \rangle_{t\to\infty} &= \frac{\Omega^2/\Gamma^2}{1 + 2\Omega^2/\Gamma^2}, \\ \langle \sigma \rangle_{t\to\infty} &= \frac{-i\Omega e^{ik_{\rm D} z \cos\theta_{\rm D}}/\Gamma}{1 + 2\Omega^2/\Gamma^2}, \end{aligned} \tag{19.173}$$

for the internal steady-state of the atom.

19.4.3.2 External Master Equation

To make the adiabatic approximation and eliminate the internal dynamics, we note that there is no effect on the external dynamics apart from the slow center-of-mass motion in the potential $V(x)$ and the collapses due to the detection events. When the internal timescales damp much more quickly than the external time scales, we can make the replacements

$$\begin{aligned} \langle \sigma^\dagger \sigma \rangle &\longrightarrow \langle \sigma^\dagger \sigma \rangle_{t\to\infty} \\ \langle \sigma \rangle &\longrightarrow \langle \sigma \rangle_{t\to\infty} \end{aligned} \tag{19.174}$$

in the master equation (19.154) and mean Poisson process (19.155). In this approximation, we will similarly ignore the fast *fluctuations* of the atomic operators, which do not substantially couple to the slow atomic dynamics, and thus also make the replacements

$$\begin{aligned} \sigma^\dagger \sigma &\longrightarrow \langle \sigma^\dagger \sigma \rangle_{t\to\infty} \\ \sigma &\longrightarrow \langle \sigma \rangle_{t\to\infty} . \end{aligned} \tag{19.175}$$

Furthermore, we will work in the small-excitation regime ($\Omega \ll \Gamma$), so that

$$\langle \sigma^\dagger \sigma \rangle \approx \langle \sigma^\dagger \rangle \langle \sigma \rangle \approx \frac{\Omega^2}{\Gamma^2}. \tag{19.176}$$

In this case, the master equation simplifies and becomes

$$d\rho = -\frac{i}{\hbar}[H_{\rm CM}, \rho]dt + \gamma \int_{-1}^{1} d(\cos\theta)\, \tilde{R}(\theta) \mathcal{D}\big[e^{ik_{\rm D} z \cos\theta_{\rm D}} e^{-ikz\cos\theta}\big]\rho\, dt + \mathcal{J}\big[A(z - \alpha\lambda/2) e^{ik_{\rm D} z \cos\theta_{\rm D}}\big]\rho\, dN_\Omega, \tag{19.177}$$

where the mean of the Poisson process becomes

$$\langle\langle dN_\Omega \rangle\rangle = \eta_\Omega \gamma \, dt, \tag{19.178}$$

and we have defined the mean spontaneous-scattering rate

$$\gamma := \Gamma \langle \sigma^\dagger \sigma \rangle . \tag{19.179}$$

The Hamiltonian $H_{\rm CM}$ refers to the external, center-of-mass Hamiltonian for the atom, since the evolutions according to the atomic Hamiltonian $\tilde{H}_{\rm A}$ and $\tilde{H}_{\rm AF}$ are trivial in this regime. Notice that the disturbance (backaction) term contains two momentum-shift operators, a deterministic one for absorption and a random one for spontaneous emission. The absorption disturbance can be eliminated in this one-dimensional analysis by taking the driving wave vector $\mathbf{k}_{\rm D}$ to be orthogonal to the z-axis ($\theta_{\rm D} = \pi/2$).

There is effectively now no dependence on the internal atomic degrees of freedom in the master equation, since all such dependence has been reduced to constant averages. We can thus take a partial trace over the internal degrees of freedom by defining the external density operator

$$\rho_{\rm ext} := \langle {\rm e}|\rho|{\rm e}\rangle + \langle {\rm g}|\rho|{\rm g}\rangle , \tag{19.180}$$

so that we obtain

$$\begin{aligned} d\rho_{\rm ext} = -\frac{i}{\hbar}[H_{\rm CM}, \rho_{\rm ext}]dt + \gamma \int_{-1}^{1} d(\cos\theta)\, \tilde{R}(\theta) \mathcal{D}\left[e^{ik_{\rm D} z \cos\theta_{\rm D}} e^{-ikz\cos\theta}\right]\rho_{\rm ext}\, dt \\ + \mathcal{J}\left[A(z - \alpha\lambda/2) e^{ik_{\rm D} z\cos\theta_{\rm D}}\right]\rho_{\rm ext}\, dN_\Omega . \end{aligned} \tag{19.181}$$

Now we have what we want: a master equation for the atomic center-of-mass state that exhibits localizing collapses due to a physical measurement process. What we essentially have is continuous evolution, with the end of each interval of mean length $(\eta_\Omega \gamma)^{-1}$ punctuated by a measurement reduction of the form

$$\rho_{\rm ext} \longrightarrow \frac{e^{ik_{\rm D} z\cos\theta_{\rm D}} A(z - \alpha\lambda/2)\rho_{\rm ext} A^\dagger(z - \alpha\lambda/2) e^{-ik_{\rm D} z\cos\theta_{\rm D}}}{\langle |A(z - \alpha\lambda/2)|^2\rangle}. \tag{19.182}$$

But note that here there is extra disturbance for the amount of information we gain, because the aperture only picks up a fraction of the available information. We will return to this point shortly.

19.4.4 White-Noise Limit: Gaussian Aperture

Now we will take the white-noise limit, and we will thus obtain a master equation in the standard form for an inefficient, continuous position measurement. To do this, we will consider the case of a small Gaussian aperture, for which we showed the collapse operator was Gaussian and given by Eq. (19.167). As in the finite-step construction of the continuous position measurement, the Gaussian collapse operator $A(\alpha)$ is applied to the state after every time interval of average length $\Delta t = (\eta_\Omega \gamma)^{-1}$. In the regime of slow atomic center-of-mass motion, the collapses come quickly compared to the motion. Then it is a good approximation (in the temporal coarse-graining sense) to take the formal limit $\Delta t \longrightarrow 0$, while keeping the rate of information gain constant.

19.4.4.1 Spatial Continuum Approximation

If an atom is initially completely delocalized, after one photon is detected and the collapse operator $A(z - \alpha')$ applies, where $\alpha' = \alpha\lambda/2$, the atom is reduced to a width of order

$$\delta\alpha = \frac{1}{k\,\delta\theta} = \frac{\lambda}{2\pi\,\delta\theta}. \tag{19.183}$$

Since this is much larger than the spacing

$$\Delta\alpha = \frac{\pi}{k} = \frac{\lambda}{2}, \tag{19.184}$$

it is effectively impossible to ''see'' the discreteness of the measurement record, and it is a good approximation to replace the set of measurement operators with a set corresponding to a continuous range of possible measurement outcomes. Since in the limit of small spacing Δx, it is a good approximation to write an integral as a sum

$$\sum_n f(n\Delta x)\,\Delta x = \int dx\, f(x) \tag{19.185}$$

for an arbitrary function $f(x)$, we can make the formal replacement

$$A(z-\alpha') \longrightarrow \frac{A(z-\alpha')}{\sqrt{\Delta\alpha}} \tag{19.186}$$

to obtain the continuum limit of the position collapse operators with proper normalization, now regarding α' as a *continuous* position index rather than a discrete real index of spacing $\lambda/2$. Dropping the prime from α', and using Eq. (19.183) in Eq. (19.167), we now have the collapse operator

$$A(z-\alpha) = \sqrt{\frac{1}{\sqrt{2\pi}\delta\alpha}}\exp\left[-\frac{(z-\alpha)^2}{4(\delta\alpha)^2}\right]. \tag{19.187}$$

We have implicitly written down this operator in the position basis, so technically we should write

$$A(z-\alpha) = \int dz|z\rangle\langle z|\sqrt{\frac{1}{\sqrt{2\pi}\delta\alpha}}\exp\left[-\frac{(z-\alpha)^2}{4(\delta\alpha)^2}\right] \tag{19.188}$$

to be general. Again, α is now a continuous index with dimensions of length, rather than an integer index.

Modifying the SME (19.177) appropriately, and taking $\cos\theta_{\rm D} = 0$ (i.e., taking the direction of the pump beam to be orthogonal to the z-axis), we now have

$$d\rho_{\rm ext} = -\frac{i}{\hbar}[H_{\rm CM},\rho_{\rm ext}]dt + \gamma\int_{-1}^{1} d(\cos\theta)\,\tilde{R}(\theta)\mathcal{D}\big[e^{-ikz\cos\theta}\big]\rho_{\rm ext}\,dt + \mathcal{J}[A(z-\alpha)]\rho_{\rm ext}\,dN_\Omega. \tag{19.189}$$

The probability density for α is then, with the appropriate modification of Eq. (19.156),

$$P(\alpha) = \int dz\,\langle z|\rho_{\rm ext}|z\rangle|A(z-\alpha)|^2 = \int dz\,\langle z|\rho_{\rm ext}|z\rangle\frac{1}{\sqrt{2\pi}\delta\alpha}\exp\left[-\frac{(z-\alpha)^2}{2(\delta\alpha)^2}\right]. \tag{19.190}$$

If the atomic wave packet is well localized beyond the scale $\delta\alpha$, the probability distribution is thus Gaussian with variance $(\delta\alpha)^2$.

19.4.4.2 Quantum-State Diffusion

Comparing the collapse operator $A(z)$ of Eq. (19.188) with the collapse operator (19.57) we see that they are the same if we identify

$$4\kappa\Delta t = \frac{1}{2(\delta\alpha)^2}. \tag{19.191}$$

Solving for the measurement strength κ, using $\Delta t = 1/\eta_\Omega\gamma$,

$$\kappa = \frac{\eta_\Omega\gamma}{8(\delta\alpha)^2} = \frac{\pi^2\eta_\Omega\gamma(\delta\theta)^2}{2\lambda^2}. \tag{19.192}$$

Repeating the procedure of Section 19.3.1, we can take the limit $\Delta t \longrightarrow 0$ with κ fixed. Here, however, this is a mathematical, coarse-graining approximation, as the measurements are *really* still occuring with a nonzero mean time between collapses. The resulting master equation, in ''quantum-state diffusion'' form, is

$$\begin{aligned} d\rho_{\rm ext} = &-\frac{i}{\hbar}[H_{\rm CM},\rho_{\rm ext}]dt + \gamma\int_{-1}^{1} d(\cos\theta)\,\tilde{R}(\theta)\,\mathcal{D}[e^{-ikz\cos\theta}]\rho_{\rm ext}\,dt \\ &+ 2\kappa\mathcal{D}[z]\rho_{\rm ext}\,dt + \sqrt{2\kappa}\mathcal{H}[z]\rho_{\rm ext}\,dW. \end{aligned} \tag{19.193}$$

The form here is the same as in Eq. (19.72), except for an extra "disturbance term" representing the undetected photons.

19.4.4.3 Diffusion Rates

To simplify the master equation (19.193), we will analyze the diffusion rates due to the second and third terms (proportional to γ and κ, respectively). From the analysis of Eqs. (19.88), recall that the term $2\kappa\mathcal{D}[z]\rho_{\text{ext}}\,dt$ causes diffusion in momentum at the rate

$$D_\kappa = 2\hbar^2\kappa = \frac{\eta_\Omega}{4}\gamma\hbar^2k^2(\delta\theta)^2. \tag{19.194}$$

This is the disturbance corresponding to the information gain. The relation $\kappa = D_\kappa/(2\hbar^2)$ will be useful below.

We can compute the *total* diffusion rate due to the spontaneously emitted photons as follows. Each photon emission causes a momentum kick of magnitude $\hbar k\cos\theta$, and the spontaneous emission rate is γ. Averaging over the angular photon distribution (19.157), the diffusion rate becomes

$$D_{\text{SE}} = \gamma\hbar^2k^2\int d\Omega\, f(\theta,\phi)\cos^2\theta = \frac{\gamma\hbar^2k^2}{5}. \tag{19.195}$$

On the other hand, we can compute the diffusion rate due only to the *detected* photons. Using the Gaussian aperture function (19.161) with Eq. (19.164) for the azimuthal part (assuming the usual dipole-radiation pattern) to obtain

$$\tilde{T}(\theta) \approx \frac{\eta_\Omega}{\delta\theta}\sqrt{\frac{2}{\pi}}\exp\left[-\frac{2(\theta-\pi/2)^2}{(\delta\theta)^2}\right], \tag{19.196}$$

the partial diffusion rate for detected photons is

$$\begin{aligned} D_\Omega &= \gamma\hbar^2k^2\int_{-1}^{1} d(\cos\theta)\,\tilde{T}(\theta)\,\cos^2\theta \\ &= \gamma\hbar^2k^2\frac{\eta_\Omega}{\delta\theta}\sqrt{\frac{2}{\pi}}\int_{-1}^{1} d(\cos\theta)\,\cos^2\theta\exp\left[-\frac{2(\theta-\pi/2)^2}{(\delta\theta)^2}\right] \\ &\approx \frac{\eta_\Omega}{4}\gamma\hbar^2k^2(\delta\theta)^2, \end{aligned} \tag{19.197}$$

where we again used the fact that $\delta\theta$ is small. This is precisely the same rate as D_κ, since they are two different representations of the same physical process.

We see now that the second and third terms of Eq. (19.193) have the same effect of momentum diffusion, both corresponding to heating from photon scattering, but at different rates, corresponding to the partition between detected and undetected photons. We can combine them to obtain

$$d\rho_{\text{ext}} = -\frac{i}{\hbar}[H_{\text{CM}},\rho_{\text{ext}}]dt + 2k_{\text{eff}}\mathcal{D}[z]\rho_{\text{ext}}\,dt + \sqrt{2\eta_{\text{eff}}k_{\text{eff}}}\mathcal{H}[z]\rho_{\text{ext}}\,dW, \tag{19.198}$$

where the effective measurement strength is

$$k_{\text{eff}} = \frac{D_{\text{SE}}}{2\hbar^2} = \frac{\gamma k^2}{10}, \tag{19.199}$$

and the effective measurement efficiency is

$$\eta_{\text{eff}} = \frac{\kappa}{k_{\text{eff}}} = \frac{5}{4}\eta_\Omega(\delta\theta)^2. \tag{19.200}$$

Notice that since $\delta\theta$ is assumed small, the *apparent* efficiency η_{eff} derived from comparing the information rate to the disturbance rate, is much smaller than the photon-detection efficiency of η_Ω. Evidently, the photons radiated near $\theta = \pi/2$ are much less effective compared to the photons radiated near $\theta = 0$ or π. This result is counterintuitive when considering typical imaging setups as we have considered here, but suggests that other ways of processing the radiated photons (e.g., measuring the phase of photons radiated closer to the z-axis) are more effective than camera-like imaging.

19.5 Position Measurement via Excitation by a Local Probe Field

Now we will examine a position-measurement method for an atom that uses resonance fluorescence but uses the resolution of a focused probe beam instead of the resolution of an imaging system to gain the position information. This setup is cleaner in the sense that the efficiency is only determined by the photon detection probability, and the form and width of the collapse operator can be chosen essentially independently of the measurement efficiency.

19.5.1 Localized Probe Field

Recall from Eq. (19.122) that the quantum-jump SME for resonance fluorescence from a two-level atom is given by

$$d\rho = -\frac{i}{\hbar}[\tilde{H},\rho]dt - \frac{\Gamma}{2}\mathcal{H}[\sigma^\dagger\sigma]\rho\,dt + \int d\Omega\,\mathcal{J}\left[\sigma e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\rho\;\frac{dN(\theta,\phi)}{d\Omega}. \tag{19.201}$$

in the case of angle-resolved detection of the photons, where again the Poisson processes corresponding to each angular element are

$$\langle\!\langle \frac{dN(\theta,\phi)}{d\Omega} \rangle\!\rangle = \Gamma\langle\sigma^\dagger\sigma\rangle\, f_{\hat{\varepsilon}}(\theta,\phi)\,dt. \tag{19.202}$$

The Hamiltonian describes both the atomic motion and the atom–field coupling, so we can decompose it into these parts as

$$\tilde{H} = H_{\text{CM}} + \tilde{H}_{\text{AF}}, \tag{19.203}$$

where the center-of-mass Hamiltonian describes one-dimensional motion along the z-axis,

$$H = \frac{p_z^2}{2m_{\text{A}}} + V(z), \tag{19.204}$$

where $V(z)$ is some external atomic potential, and the atom–field coupling Hamiltonian is given in the rotating frame of the laser field from Eq. (5.406) by

$$\tilde{H}_{\text{AF}} = \frac{\hbar}{2}\left[\Omega^*(z)\sigma + \Omega(z)\sigma^\dagger\right], \tag{19.205}$$

where $\Omega(z)$ is the space-dependent Rabi frequency representing the resonant probe field, defined such that $|\Omega(z)|^2$ is proportional to the local intensity. We assume the probe to propagate normal to the z-axis, so that we can assume a zero average momentum recoil on absorption.

In the weak-excitation regime, where the rate of spontaneous emission is much smaller than the excited-state decay rate Γ, we can adiabatically eliminate the atomic internal state, which as in Section (19.4.3) amounts to replacing the operator σ by its space-dependent, steady-state value $\langle\sigma\rangle_{t\to\infty} \approx -i\Omega(z)/\Gamma$ (to lowest order in Ω/Γ). With this replacement, the SME (19.201) becomes

$$d\rho = -\frac{i}{\hbar}[H_{\text{CM}},\rho]dt - \frac{1}{2\Gamma}\mathcal{H}\left[|\Omega(z)|^2\right]\rho\,dt + \int d\Omega\,\mathcal{J}\left[\Omega(z)e^{-i\mathbf{k}\cdot\mathbf{r}}\right]\rho\;\frac{dN(\theta,\phi)}{d\Omega}. \tag{19.206}$$

while the mean of the Poisson process reduces to

$$\langle\!\langle dN \rangle\!\rangle = \frac{1}{\Gamma}\langle|\Omega(z)|^2\rangle\,dt. \tag{19.207}$$

Clearly, now, the electric-field profile $\Omega(z)$ of the probe acts as a collapse operator for a position measurement, and the rate of spontaneous scattering gives information about the atomic position. To make these equations a bit cleaner, we can define a normalized collapse operator by

$$A(z) := \frac{\Omega(z)}{\Omega_\int}, \tag{19.208}$$

where we have defined the integrated Rabi frequency

$$\Omega_{\int} := \sqrt{\int_{-\infty}^{\infty} dz\, |\Omega(z)|^2}. \tag{19.209}$$

Then in terms of the normalized collapse operators, the SME becomes

$$d\rho = -\frac{i}{\hbar}[H_{\rm CM}, \rho]dt - \frac{\Omega_{\int}^2}{2\Gamma}\mathcal{H}\big[|A(z)|^2\big]\,\rho\,dt + \int d\Omega\,\mathcal{J}\big[A(z)e^{-i\mathbf{k}\cdot\mathbf{r}}\big]\rho\;\frac{dN(\theta,\phi)}{d\Omega}. \tag{19.210}$$

and the Poisson means become

$$\langle\!\langle \frac{dN(\theta,\phi)}{d\Omega} \rangle\!\rangle = \frac{\Omega_{\int}^2}{\Gamma}\big\langle |A(z)|^2 \big\rangle\, f_{\hat{\varepsilon}}(\theta,\phi)\,dt. \tag{19.211}$$

The overall rate of information gain is thus given by the ratio $\Omega_{\int}^2/\Gamma$. Note that the two information-related terms have *opposite effects*: for a localized probe profile $|A(z)|$, the last (stochastic) term collapses (localizes) the atomic wave function by multiplying by $|A(z)|$ and renormalizing. By contrast, in the *absence* of photodetection events, the second term moves the atom *away* from the probe by transforming the probability density according to $\rho_{zz} \longrightarrow \rho_{zz} - 2(\Omega_{\int}^2/2\Gamma)(|A(z)|^2\rho_{zz} + \rho_{zz}|A(z)|^2)\,dt$, thus tending to reduce the population where $|A(z)|^2$ is maximum.

Physically, where does the information come from? From the steady-state relations $\langle\sigma\rangle_{t\to\infty} \approx -i\Omega(z)/\Gamma$ and $\langle\sigma^\dagger\sigma\rangle_{t\to\infty} \approx |\Omega(z)|^2/\Gamma^2$ (again, to lowest order in the field strength), we can see that the excited-state amplitude in steady state is related to the ground-state amplitude by $\psi_{\rm e}(z) \propto \Omega(z)\psi_{\rm g}$. The spontaneous-emission event then "flushes away" $\psi_{\rm g}(z)$, making $\psi_{\rm e}(z)$ the new atomic wave function. While the information is given in principle by the emission event, the information is "set up" by the *absorption* of the photon from the probe field. Correspondingly, the back-action on the quantum state due to the localized probe (i.e., the increase of the momentum width of the atomic state due to the position-space collapse) is due to absorption of a photon with a superposition of wave-vector orientations, as is consistent with having the localized probe field.

Of course, as we saw from the quantum theory of imaged resonance fluorescence above, the emitted photon contains yet more center-of-mass information about the atom, beyond the fact that it has merely scattered an atom, as a consequence of the random direction of the photon recoil. In principle, you could extract the most information about the atom by also imaging the resonance fluorescence, but if the fluorescence is merely detected without imaging or angle resolution, then we should trace over all possible photodetection angles in (19.210),

$$d\rho = -\frac{i}{\hbar}[H_{\rm CM}, \rho]dt - \frac{\Omega_{\int}^2}{2\Gamma}\mathcal{H}\big[|A(z)|^2\big]\,\rho\,dt + \int_{-1}^{1} d(\cos\theta)\,\tilde{f}_{\hat{\varepsilon}}(\theta)\,\mathcal{J}\big[A(z)e^{-ikz\cos\theta}\big]\rho\,dN, \tag{19.212}$$

so that

$$\langle\!\langle dN \rangle\!\rangle = \frac{\Omega_{\int}^2}{\Gamma}\big\langle |A(z)|^2 \big\rangle\,dt, \tag{19.213}$$

and the last term in the SME puts the atom in an incoherent superposition of having recoiled in all possible directions, weighted by the correct probabilities. Note that we have also carried out the ϕ part of the angular integral in the last term of (19.212), where

$$\tilde{f}_{\hat{\varepsilon}}(\theta) := \int_0^{2\pi} d\phi\, \tilde{f}_{\hat{\varepsilon}}(\theta,\phi) \tag{19.214}$$

is the effective angular distribution for the atomic resonance fluorescence, since the ϕ angle is immaterial as far as the atomic dynamics are concerned.

19.5.2 Scanning Probe Field

The fluorescent probe, as outlined above, only gives information about whether or not the atom is in the vicinity of the probe. To obtain a more standard position measurement, we can now consider the case of a *moving* probe field, where the center of the probe moves according to the trajectory $z_{\rm probe}(t)$. We will assume $z_{\rm probe}(t)$ to be a sawtooth function of constant velocity $v_{\rm probe}$, but jumping discontinuously from $z_{\rm max}$ to $-z_{\rm max}$ at the end of each sweep. We will also assume the time for a single scan to be much slower than the time scale Γ^{-1} for the internal state to equilibrate, but we will assume it to be much faster than any motional time scale for the atom. We also assume that the atom will remain localized within the region $(-z_{\rm max}, z_{\rm max})$. The effect of the moving probe is to make the replacement $A(z) \longrightarrow A[z - z_{\rm probe}(t)]$ in the above SME. Performing a time average on the probe-raster time scale in the second term then amounts to replacing $A(z - z_{\rm probe}(t))$ by a function that is approximately uniform over $(-z_{\rm max}, z_{\rm max})$ and zero elsewhere. Because of our assumption that the atom stays within the range of the probe, the second term has no effect on the atomic state, and can be dropped. What we are essentially saying is that the probe should always excite the atom equally no matter where it is, and thus there is no information to be gained by *not* detecting a photon. Thus,

$$d\rho = -\frac{i}{\hbar}[H_{\rm CM}, \rho]dt + \int_{-1}^{1} d(\cos\theta)\, \tilde{f}_{\hat{\varepsilon}}(\theta)\, \mathcal{J}\left[A[z - z_{\rm probe}(t)]e^{-ikz\cos\theta}\right]\rho\, dN, \tag{19.215}$$

with

$$\langle\!\langle dN(t) \rangle\!\rangle = \frac{\Omega_f^2}{\Gamma}\left\langle |A[z - z_{\rm probe}(t)]|^2 \right\rangle dt. \tag{19.216}$$

Now we can see that $z_{\rm probe}(t)$ acts as an index for the displaced collapse operator $A(z - z_{\rm probe}(t))$. If the probe raster time $\Delta t_{\rm r}$ is much shorter than both the time scale for atomic motion and the mean time between spontaneous-scattering events, but we carefully time-resolve the detection events, then as far as the motional dynamics are concerned, we can time-average the dynamics on time scales of $\Delta t_{\rm r}$ to write

$$d\rho = -\frac{i}{\hbar}[H_{\rm CM}, \rho]dt + \int_{-1}^{1} d(\cos\theta)\, \tilde{f}_{\hat{\varepsilon}}(\theta)\, \mathcal{J}\left[A(z - z_{\rm d})e^{-ikz\cos\theta}\right]\rho\, dN, \tag{19.217}$$

where $z_{\rm d} \in (-z_{\rm max}, z_{\rm max})$, which is simply $z_{\rm probe}(t)$ evaluated at the time of the detection event, is a stochastic, random variable with probability density

$$P(z_{\rm d}) = \left\langle |A(z - z_{\rm d})|^2 \right\rangle, \tag{19.218}$$

and the Poisson process now responds only to the time-averaged probe intensity,

$$\langle\!\langle dN(t) \rangle\!\rangle = \frac{\Omega_f^2}{\Gamma}\left[\int_{-\infty}^{\infty} dz\, |A(z)|^2\, {\rm rect}(z/2z_{\rm max})\right] dt =: \gamma\, dt, \tag{19.219}$$

where the bracketed quantity represents the convolution of the probe-intensity profile $|A(z)|^2$ with the time-averaged trajectory probe, evaluated at $z = 0$ (thus assuming $z_{\rm max}$ is much larger than the width of $A(z)$ and that the atom stays away from the edges of the scan range), and we have used ${\rm rect}(z)$ as the rectangular-pulse function of unit height and width.

19.5.3 Example: Gaussian Probe

Again, to see how much information we are getting, we can compare the diffusion rate due to the measurement process to the rate of information gain. If we take the collapse operators $A(z)$ to have a Gaussian profile, that is, we take the probe intensity to have the Gaussian form

$$|A(z - z_{\rm d})|^2 \propto \exp\left(-\frac{2(z - z_{\rm d})^2}{w_0^2}\right), \tag{19.220}$$

where w_0 is the beam-waist parameter for the Gaussian beam, we can then compare to the Gaussian form of Eq. (19.57) to identify the measurement strength as $\kappa = 1/2w_0^2\,\Delta t$. Noting that the average time Δt between detection events is $1/\gamma$ (assuming unit detection efficiency of the radiated photons), the measurement strength in the formal white-noise limit is $\kappa = \gamma/2w_0^2$. From our analysis of Eqs. (19.88), we conclude that the momentum-diffusion rate due to the measurement (i.e., the application of the $A(z)$ operators) is $D_p = \hbar^2\kappa$. However, this is only the diffusion rate due to *absorption*, which is where the position-dependent nature of the probe enters; the emission events cause additional diffusion.

From Eq. (19.195), the diffusion rate due to spontaneous emission (i.e., the application of the $e^{-ikz\cos\theta}$ factors), assuming an atomic dipole oriented along the z-axis, is

$$D_{\text{SE}} = \frac{\gamma\hbar^2k^2}{5}. \tag{19.221}$$

Thus, the effective measurement-information efficiency for this probe-measurement scheme is that fraction of the total diffusion rate that corresponds to the measurement gain (i.e., the absorption):

$$\eta_{\text{eff}} = \frac{D_p}{D_p + D_{\text{SE}}} = \frac{1}{1 + 2w_0^2k^2/5}. \tag{19.222}$$

In other words, the effective measurement efficiency goes *down* as the beam waist w_0 becomes larger, because the information gained becomes smaller but the disturbance due to spontaneous emission is the same. Practically, w_0 is limited to something on the order of λ, and for a very tight focus of $w_0 = \lambda$, the effective measurement efficiency would be limited to a maximum of $\eta_{\text{eff}} = 6\%$, a rather low value. The efficiency is correspondingly further reduced by the detection efficiency.

19.6 Continuous Momentum Measurement by EIT

As an alternative to continuous measurements of *position*, we can also consider schemes to continuously measure the *momentum* of a single atom. In some sense, continuous measurements of position and momentum are equivalent, since to some extent one can infer a momentum trajectory from a continuous record of position, and vice versa. In the simple but important example of the harmonic oscillator, position and momentum represent different yet equivalent directions in phase space. Electromagnetically induced transparency (EIT) provides a momentum-sensitive probe for a single atom that works without a cavity.[16]

19.6.1 General Remarks

Recalling the phenomenon of EIT from Section 6.2.2, the susceptibility of an atomic gas of number density N for a weak probe field (field 2) due to the pump (field 1) is [from Eq. (6.70)]

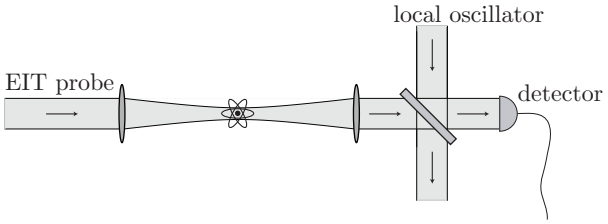
$$\chi = \frac{-4N|\langle \text{g}_2|\hat{\varepsilon}_2\cdot\mathbf{d}|\text{e}\rangle|^2}{\epsilon_0\hbar}\,\frac{[(\Delta_2-\Delta_1)+i\gamma_{\text{g}}]}{\Omega_1^2}, \tag{19.223}$$

where we have taken the limit of a strong pump field (large Ω_1). We assume now that for an atom at rest, the two fields are at Raman resonance ($\Delta_1 = \Delta_2 = \Delta$), and we also assume that the pump and probe counterpropagate, so that $\Delta_1 = \Delta + k_2v$ and $\Delta_2 = \Delta - k_2v$, where v is the atomic velocity in the direction of the EIT fields, and k_2 is the wave number of the probe field. Then

$$\chi = \frac{-4N|\langle \text{g}_2|\hat{\varepsilon}_2\cdot\mathbf{d}|\text{e}\rangle|^2}{\epsilon_0\hbar\Omega_1^2}(-2kv + i\gamma_{\text{g}}), \tag{19.224}$$

and assuming that the ground-state relaxation rate γ_{g} is negligible, we see that the susceptibility is real and proportional to the atomic velocity. The atoms thus present a refractive index $n = 1 + \text{Re}[\chi]/2$ whose deviation from the vacuum value is proportional to the velocity.

[16] P. Rabl, V. Steixner and P. Zoller, "Quantum limited velocity readout and quantum feedback cooling of a trapped ion via electromagnetically induced transparency," *Physical Review A* **72**, 043823 (2005) (doi: 10.1103/PhysRevA.72.043823).



By measuring the phase of the probe beam using homodyne detection, we thus continuously extract momentum (velocity) information about the atom. Of course, the "number density" N for a single atom is small, but EIT functions to amplify the effect of the atomic momentum on the phase of the field, and so we need to treat this system more carefully below.

19.6.2 Homodyne Detection of the EIT Probe

With direct detection of the EIT probe beam, the appropriate collapse operator for a detected photon in analogy with Section 18.2 is

$$C_{\text{direct}} = \sqrt{\Gamma_2}(\sigma_2 + \alpha_{\text{probe}}), \tag{19.225}$$

where Γ_2 is the partial decay rate on the EIT probe transition, $\sigma_2 := |\text{g}_2\rangle\langle\text{e}|$ is the atomic annihilation operator for the EIT probe transition, and α_{probe} is a complex number representing the coherent state of the probe field. Strictly speaking, this is the collapse operator for the mode of the probe field, and photons scattered into other modes must be treated by a separate detection process, as we will do later. This collapse operator already has the appropriate form for homodyne detection, as it represents the lack of knowledge about whether a detected photon came from the atom or from the probe field. We assume the pump field to be in a different mode, and we will otherwise ignore the pump except for its cooperative effects with the probe.

When the probe field is then monitored via simple homodyne detection, as in the above diagram, the collapse operator is then modified to include the local-oscillator field as

$$C_{\text{homodyne}} = \sqrt{\Gamma_2}(\sigma_2 + \alpha_{\text{probe}} + \beta), \tag{19.226}$$

where we have already taken the limit as the reflection coefficient of the beam splitter vanishes, taking the local-oscillator amplitude to be correspondingly large (and absorbing the reflection coefficient into the local-oscillator amplitude β). Combining the two classical fields, we can write

$$C_{\text{homodyne}} = \sqrt{\Gamma_2}(\sigma_2 + \beta'), \tag{19.227}$$

where $\beta' := \beta + \alpha_{\text{probe}}$. This collapse operator has precisely the form of the collapse operator for simple homodyne detection, and thus the analysis for homodyne detection from Section 18.2 carries through here. Thus, from our previous analysis of homodyne detection, the SME becomes

$$d\rho = -\frac{i}{\hbar}[H, \rho]dt + \Gamma_2\mathcal{D}\left[\sigma_2 e^{i\phi}\right]\rho\, dt + \sqrt{\eta\Gamma_2}\mathcal{H}\left[\sigma_2 e^{i\phi}\right]\rho\, dW, \tag{19.228}$$

where H is the atomic Hamiltonian, including the interaction with the pump and probe fields, ϕ is the phase of the combined field β',

$$\beta' =: |\beta'|e^{-i\phi}. \tag{19.229}$$

and the efficiency η represents the fraction of photons radiated by the atom on the $|\text{g}_2\rangle \longrightarrow |\text{e}\rangle$ transition into the mode of the probe beam—recall that the phase shift of the probe beam is due to the interference of the dipole radiation and the original probe field. The corresponding scaled photocurrent (measurement record) is

$$d\tilde{r}(t) = \Gamma_2\left\langle \sigma_2 e^{i\phi} + \sigma_2^\dagger e^{-i\phi}\right\rangle dt + \sqrt{\frac{\Gamma_2}{\eta}}\, dW, \tag{19.230}$$

so that we must still choose the local-oscillator phase ϕ to obtain the appropriate information.

19.6.3 Adiabatic Approximation

Since the atomic motion is much slower than the internal atomic dynamics, we can adiabatically eliminate the internal atomic state by replacing the internal atomic operators (namely σ_2) by their steady-state values. The steady-state coherence from Eq. (6.68) on the probe transition $|\mathrm{g}_2\rangle \longrightarrow |\mathrm{e}\rangle$ is

$$\langle\sigma_2(t\longrightarrow\infty)\rangle = \tilde{\rho}_{\mathrm{eg}_2}(t\longrightarrow\infty) = \frac{i(\Omega_2/2)[(\Delta_2-\Delta_1)+i\gamma_\mathrm{g}]}{(i\Delta_2-\Gamma_2/2)[(\Delta_2-\Delta_1)+i\gamma_\mathrm{g}]-i(\Omega_1/2)^2}. \tag{19.231}$$

Assuming Raman resonance ($\Delta_1=\Delta_2$) for atoms at rest, ignoring the ground-state relaxation ($\gamma_\mathrm{g}=0$), assuming a counterpropagating pump-probe pair, so that $\Delta_1=\Delta+k_2v$ and $\Delta_2=\Delta-k_2v$, and keeping only the first-order velocity term,

$$\langle\sigma_2(\longrightarrow\infty)\rangle = \frac{4k_2\Omega_2}{\Omega_1^2}v = \frac{4k_2\Omega_2}{m_\mathrm{a}\Omega_1^2}p. \tag{19.232}$$

Then choosing a measurement phase $\phi=0$, the measurement record (19.230) becomes

$$d\tilde{r}(t) = \sqrt{8\kappa\Gamma_2}\langle p\rangle\,dt + \sqrt{\frac{\Gamma_2}{\eta}}\,dW, \tag{19.233}$$

where the measurement strength is

$$\kappa := \frac{8k_2^2\Gamma_2\Omega_2^2}{m_\mathrm{a}^2\Omega_1^4}, \tag{19.234}$$

where m_a is the atomic mass. The SME (19.228) then becomes

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + 2\kappa\mathcal{D}[p]\,\rho\,dt + \sqrt{2\eta\kappa}\mathcal{H}[p]\,\rho\,dW, \tag{19.235}$$

and the measurement record (19.233) can be rescaled to appear in a more standard form for a position-type measurement:

$$dy := \frac{d\tilde{r}(t)}{\sqrt{8\kappa\Gamma_2}} = \langle p\rangle\,dt + \frac{dW}{\sqrt{8\eta\kappa}}. \tag{19.236}$$

This result is valid in the momentum range

$$|p| \ll \frac{m_\mathrm{a}\Omega_1^2}{8k_2\sqrt{\Delta_2^2+(\Gamma_2/2)^2}}, \tag{19.237}$$

which follows if the second term of the denominator of Eq. (19.231) is to dominate the first. If we assume a small probe detuning Δ_2 compared to Γ_2, a moderate pump of $\Omega_1=\Gamma_2\sim 20\times10^6\ \mathrm{s}^{-1}$, and we consider the case of ^{87}Rb ($\lambda=780$ nm), then this corresponds to the range $|p|\ll 200\hbar k_2$. Assuming further a weak probe of $\Omega_2=\Omega_1/10$, the measurement strength is $\kappa=9(\hbar k_2)^2\mathrm{s}^{-1}$.

19.6.4 Spontaneous Scattering

The spontaneous scattering rate should be small, at least when the motion of the atom stays near $p=0$, due to the nature of the dark state. To compute the rate of spontaneous scattering, we use the equation of motion

$$\partial_t\rho_{\mathrm{g_2g_2}} = \frac{i\Omega_2}{2}(\tilde{\rho}_{\mathrm{g_2e}}-\tilde{\rho}_{\mathrm{eg}_2}) + \Gamma_2\rho_{\mathrm{ee}}, \tag{19.238}$$

as follows from the unconditioned master equation for EIT, Eq. (6.63). In steady state, this gives the excited-state population in terms of the coherence on the probe-transition:

$$\rho_{\mathrm{ee}}(t\longrightarrow\infty) = \frac{i\Omega_2}{2\Gamma_2}(\tilde{\rho}_{\mathrm{eg}_2}-\tilde{\rho}_{\mathrm{g_2e}}). \tag{19.239}$$

We can expand the steady-state coherence (19.231) to second order in momentum to obtain

$$\tilde{\rho}_{\mathrm{eg}_2}(t \longrightarrow \infty) = \frac{4k_2\Omega_2}{m_{\mathrm{a}}\Omega_1^2}p - \frac{32k_2^2(i\Gamma_2/2+\Delta_2)\Omega_2}{m_{\mathrm{a}}^2\Omega_1^2}p^2 + O(p^3), \tag{19.240}$$

and then put this into Eq. (19.239) to find

$$\rho_{\mathrm{ee}}(t \longrightarrow \infty) = \frac{32k_2^2\Omega_2^2}{m_{\mathrm{a}}^2\Omega_1^4}p^2 + O(p^3). \tag{19.241}$$

The steady-state scattering rate is then simply

$$R_{\mathrm{sc}}(t \longrightarrow \infty) = (\Gamma_1+\Gamma_2)\rho_{\mathrm{ee}} = \frac{32k_2^2(\Gamma_1+\Gamma_2)\Omega_2^2}{m_{\mathrm{a}}^2\Omega_1^4}p^2 + O(p^3), \tag{19.242}$$

so we see that the scattering rate vanishes to first order in the atomic momentum. However, to account for this effect, we should include an extra diffusion term to the SME (19.235)

$$d\rho = -\frac{i}{\hbar}[H,\rho]dt + 2\kappa\mathcal{D}[p]\,\rho\,dt + \sqrt{2\eta\kappa}\mathcal{H}[p]\,\rho\,dW + \gamma\int_{-1}^{1} d(\cos\theta)\tilde{f}_{\hat{\varepsilon}}(\theta)\mathcal{D}[e^{-ikz\cos\theta}]\rho\,dt, \tag{19.243}$$

where $\gamma := \langle R_{\mathrm{sc}}\rangle$, the angular distribution $\tilde{f}_{\hat{\varepsilon}}(\theta)$ is defined by Eq. (19.214), and we are assuming that the extra spontaneous scattering is not detected. Note again that since the EIT measurement is *dispersive*, it scales as p, while the spontaneous emission scales as p^2. Again, this is because the dipole-radiated field scales as p, and the phase shift is due to an interference with this field and the forward EIT probe beam, so that the phase shift goes as the product of the two fields, while the spontaneous emission goes as the square of the dipole-radiated field.

19.6.5 Phase

It is illuminating to work out the relative phases of the atomic dipole, the EIT probe field, and the local-oscillator field. The phase of the atomic dipole, given by the phase of Eq. (19.232), is defined with respect to the phase of the EIT probe field. To lowest order in momentum, the atomic dipole (and thus the dipole field) is thus exactly in phase with the probe field. Further, we saw that the appropriate choice of local-oscillator phase was $\phi = 0$, so that the local oscillator is also exactly in phase with the atomic dipole and the driving field. This seems a bit strange: if the probe is, in fact, phase-shifted by the effective refractive index of the atom, then the atom should radiate *in quadrature* with the field, not in phase with it (radiation exactly in or exactly out of phase can affect only the amplitude, not the phase, of the probe field). Further, to detect the phase of the probe field, the local oscillator should again be in quadrature with the probe field, so that any phase shifts of the probe would act to modulate the detected intensity of the local oscillator.

An easy way to resolve this difficulty is to note that to have a strong coupling to the atom, the EIT probe field should be tightly focused onto the atom. If we take as a concrete example a Gaussian beam, we may write the probe field as

$$E^{(+)}(\mathbf{r}) = E_0^{(+)}\left(-\hat{x} + \frac{x}{z-iz_0}\hat{z}\right)\frac{w_0}{w(z)}\exp\left[-\frac{r^2}{w^2(z)}\right]\exp\left[ikz - i\tan^{-1}\left(\frac{z}{z_0}\right)\right]\exp\left[ik\frac{r^2}{2R(z)}\right], \tag{19.244}$$

where the beam propagates along the z-direction and is (mostly) polarized along the $-x$-direction, w_0 is the Gaussian beam-waist parameter that characterizes the beam waist at the focus, $z_0 = \pi w_0^2/\lambda$ is the Rayleigh length, $r^2 = x^2+y^2$ is the radial coordinate, $w(z) := w_0\sqrt{1+(z/z_0)^2}$ is the z-dependent spot size, and $R(z) := z[1+(z_0/z)^2]$ is the radius of curvature of the phase fronts. The Gaussian beam as written here solves the electromagnetic wave equation in the paraxial approximation (i.e., as long as the divergence angle of the beam is not too large). The important thing to notice is the second exponential factor, which gives the longitudinal phase. There is the usual plane-wave-type phase of ikz, but there is also the **Gouy phase** of $-i\tan^{-1}(z/z_0)$, which varies from $i\pi/2$ at $z \longrightarrow -\infty$ to 0 at the focus $z = 0$ to $-i\pi/2$ at $z \longrightarrow \infty$. The Gouy phase is generic for focused beams, and plays an important role here.

The dipole field lacks this Gouy phase, with a phase varying as ikz along the z-axis. Similarly, we take the local oscillator to be collimated (but still Gaussian), so its Gouy phase varies negligibly over the scale of the optical apparatus. Thus, even though the dipole, EIT-probe, and local-oscillator fields are all in phase at the atomic position (Gaussian-beam focus) $z = 0$, by the time the EIT probe is recollimated in the far field, it has accumulated an extra phase shift of $-i\pi/2$. Thus, as we expect, the EIT probe is in quadrature with both the atomic-dipole and local-oscillator fields.

From this, we can view the homodyne detection slightly differently. Rather than regarding the detection as a measurement of the phase of σ_2, we can regard it as a measurement of the phase of the phase-shifted EIT probe, $\sigma_2 + \alpha_{\text{probe}} = \sigma_2 - i|\alpha_{\text{probe}}|$, by adding the local-oscillator field β, as in Eq. (19.226). Our analysis above showed the phase of σ_2 to be near zero, and the same choice of phase was best for the local oscillator, while the phase of $-i$ in the EIT probe here is due to the Gouy phase accumulated as the probe beam travels from the atom to the detector. Viewed thusly, since the homodyne-detection signal measured $\left\langle \sigma_2 + \sigma_2^\dagger \right\rangle$ in the analysis above, here it measures $\left\langle (\sigma_2 - i|\alpha_{\text{probe}}|) + (\sigma_2 - i|\alpha_{\text{probe}}|)^\dagger \right\rangle = \left\langle \sigma_2 + \sigma_2^\dagger \right\rangle$, and thus the conclusions above still obtain: because the probe and local-oscillator fields are in quadrature, the homodyne measurement rejects any contribution to the measurement signal from the probe. Similarly, viewed this way, the analysis for *balanced* homodyne detection as in Section 18.2.6 carries through here, if we view the balanced homodyne detection as a measurement of the probe phase rather than the atomic phase. Then the same SME and measurement record result, except that there is no need to subtract a dc offset from the measurement signal.

19.6.6 Detection Efficiency

The detection efficiency η here is simply the probability that a photon radiated by the atom is scattered into the mode of the EIT probe. As an example case, we will take the EIT probe again to be a Gaussian beam, and compute explicitly the overlap between the probe and dipole waves. It is sufficient to consider the overlap in the far field, for which the Gaussian beam (19.244) becomes

$$E^{(+)}(\mathbf{r}) = E_0^{(+)} \frac{z_0}{z} \exp\left[-\frac{z_0^2 r^2}{w_0^2 z^2}\right], \tag{19.245}$$

noting that we have dropped the polarization and phase factors, as they will be irrelevant for the mode overlap (within the paraxial approximation, they will exactly match the same factors for the dipole wave). We can then write this field as a normalized field mode function

$$\sqrt{f_{\text{Gaussian}}(\mathbf{r})} = \frac{2z_0}{\sqrt{2\pi}\, w_0 z} \exp\left[-\frac{z_0^2 r^2}{w_0^2 z^2}\right], \tag{19.246}$$

normalized so that at any fixed z,

$$\int dx \int dy\; f_{\text{Gaussian}}(\mathbf{r}) = 1. \tag{19.247}$$

Assuming a linearly polarized atomic dipole, the usual *field* dipole pattern is

$$\sqrt{f_{\hat{x}}(\theta,\phi)} = \sqrt{\frac{3}{8\pi}} \sin\theta = \sqrt{\frac{3}{8\pi} \frac{y^2+z^2}{x^2+y^2+z^2}}, \tag{19.248}$$

but where θ and ϕ are defined with respect to the polarization (x) axis, not the z-axis. In the paraxial approximation, we may write this field in terms of a normalized spatial distribution (by dividing by z) as

$$\sqrt{f_{\hat{x}}(\mathbf{r})} \approx \frac{1}{z}\sqrt{\frac{3}{8\pi}} \left(1 - \frac{x^2}{2z^2}\right). \tag{19.249}$$

Then the efficiency is the overlap integral of the Gaussian field with the dipole field:

$$\eta = \int dx \int dy\; \sqrt{f_{\hat{x}}(\mathbf{r}) f_{\text{Gaussian}}(\mathbf{r})} = \frac{\sqrt{3} w_0 (4z_0^2 - w_0^2)}{8 z_0^3} \approx \frac{\sqrt{3}\,\lambda}{2\pi w_0}. \tag{19.250}$$

For a focus of $w_0 = 5\lambda$ (for a far-field divergence half-angle of 3.6°), the efficiency is 5.5%, while for a focus of $w_0 = \lambda$ (for a far-field divergence half-angle of 18°, where the paraxial approximation is no longer very good), the efficiency is 27%.

19.7 Exercises

Problem 19.1

Show that the Weyl correspondence (Section 4.3.5) for the stochastic master equation for position measurement,

$$\begin{aligned} d\rho &= 2\kappa\mathcal{D}[x]\rho\,dt + \sqrt{2\kappa}\mathcal{H}[x]\rho\,dW \\ &= 2\kappa x\rho x - \kappa(x^2\rho + \rho x^2)\,dt + \sqrt{2\kappa}\big[x\rho + \rho x - 2\langle x\rangle\,\rho\big]\,dW(t) \end{aligned} \tag{19.251}$$

(without Hamiltonian evolution) gives the Fokker-Planck equation for the Wigner function with no drift and a momentum diffusion coefficient of $D = 2\hbar^2\kappa$, plus a stochastic driving term:

$$dW(x,p) = \hbar^2\kappa\partial_p^2 W(x,p)\,dt + \sqrt{8\kappa}\big(x - \langle x\rangle\big)W(x,p)\,dW(t). \tag{19.252}$$

Problem 19.2

Show that the Weyl correspondence (Section 4.3.5) for the stochastic master equation for position-squared measurement,

$$\begin{aligned} d\rho &= 2\kappa\mathcal{D}[x^2]\rho\,dt + \sqrt{2\kappa}\mathcal{H}[x^2]\rho\,dW \\ &= 2\kappa x^2\rho x^2 - \kappa(x^4\rho + \rho x^4)\,dt + \sqrt{2\kappa}\big[x^2\rho + \rho x^2 - 2\big\langle x^2\big\rangle\,\rho\big]\,dW(t) \end{aligned} \tag{19.253}$$

(without Hamiltonian evolution) gives the Fokker-Planck equation for the Wigner function with no drift and a position-dependent momentum diffusion coefficient of $D = 8\hbar^2\kappa x^2$, plus a stochastic driving term:

$$dW(x,p) = 4\hbar^2\kappa x^2\partial_p^2 W(x,p)\,dt + \sqrt{8\kappa}\left(x^2 - \big\langle x^2\big\rangle - \frac{\hbar^2\partial_p^2}{4}\right)W(x,p)\,dW(t). \tag{19.254}$$

Problem 19.3

Consider a particle subjected to a noisy potential of the form

$$V(x,t) = \hbar\sqrt{\kappa}\,x \circ \frac{dW(t)}{dt}. \tag{19.255}$$

(a) Why is it appropriate to write the potential in Stratonovich form, rather than Itō form?

(b) Write down a stochastic Schrödinger equation for the particle, convert it to Itō form, and then use the result to derive a stochastic master equation.

(c) Show that in the ensemble average, this SME is equivalent to the unconditioned SME for position measuremeent. (In the *conditioned* case, the two master equations are *not* equivalent; for example, the noisy potential still generates a *linear* SME.)

Chapter 20

Path Integration

20.1 Feynman Path Integral

We will begin by developing the path-integral formulation of quantum mechanics,[1] at least for the evolution of a single particle in a potential. Recall that we introduced the **propagator** [Eq. (15.17)] as the matrix representation of the evolution operator,

$$K(\beta, t; \alpha, t_0) := \langle \beta | U(t, t_0) | \alpha \rangle, \tag{20.1}$$
(propagator)

where the evolution operator corresponds to a time-independent Hamiltonian $H(x, p)$:

$$U(t, t_0) = e^{-iH(x,p)(t-t_0)/\hbar}. \tag{20.2}$$

Here we will work more specifically with the position-representation propagator

$$K(x, t; x_0, t_0) := \langle x | U(t, t_0) | x_0 \rangle = \langle x, t | x_0, t_0 \rangle. \tag{20.3}$$
(propagator)

In the last expression here, the evolution operator has been ''hidden'' inside the states according to the notation

$$|x, t\rangle = e^{iHt/\hbar} |x\rangle \tag{20.4}$$

where the ''wrong'' time dependence comes not from $|x, t\rangle$ denoting time-evolving states in the sense of $|x(t)\rangle$, but as eigenstates of the Heisenberg-picture operator $x(t)$. Thus, the propagator here is roughly a ''transition amplitude'' from x_0 at time t_0 to x at time t. More precisely, if we recall that

$$|\psi(t)\rangle = U(t, t_0) |\psi(t_0)\rangle, \tag{20.5}$$

we can apply $\langle x|$ and inserting an identity to give

$$\langle x | \psi(t) \rangle = \langle x | U(t, t_0) | \psi(t_0) \rangle = \int dx_0 \, \langle x | U(t, t_0) | x_0 \rangle \langle x_0 | \psi(t_0) \rangle, \tag{20.6}$$

or

$$\psi(x, t) = \int dx_0 \, K(x, t; x_0, t_0) \, \psi(x_0, t_0), \tag{20.7}$$
(evolution via propagator)

so that $K(x, t; x_0, t_0)$ is a convolution kernel corresponding to the evolution (i.e., $K(x, t; x_0, t_0)$ is a representation of the Green function for the Schrödinger equation).

[1] R. P. Feynman, ''Space-Time Approach to Non-Relativistic Quantum Mechanics,'' *Reviews of Modern Physics* **20**, 367 (1948) (doi: 10.1103/RevModPhys.20.367); Richard P. Feynman and A. R. Hibbs, ''Quantum Mechanics and Path Integrals,'' (McGraw–Hill, 1965).

A key technique in the construction of the path integral is the splitting of the propagator at an intermediate time. This is easy to do, we simply insert the identity operator, expanded in the position basis at some intermediate time t', into the propagator inner product,

$$\begin{aligned} K(x,t;x_0,t_0) &= \langle x,t|x_0,t_0\rangle \\ &= \int dx' \langle x,t|x',t'\rangle\langle x',t'|x_0,t_0\rangle, \end{aligned} \tag{20.8}$$

with the result

$$K(x,t;x_0,t_0) = \int dx' K(x,t;x',t')K(x',t';x_0,t_0).$$

(composition property of propagator) (20.9)

This is the analogue of the **Chapman–Kolmogorov equation** (17.133) for conditional probability densities in diffusion problems. This relation basically says that we can regard the transition from x_0 to x at the corresponding times to be as if the particle ''passed through'' the intermediate x' point, so long as we sum the amplitude over all possible intermediate points x'.

20.1.1 Dividing the Evolution

Now the idea is to do the same thing, but split the time interval $(t-t_0)$ into N equal subintervals of length $\delta t = (t-t_0)/N$, and inserting an identity between each of the split factors of the evolution operator. We will assume N to be large (such that we will suppose δt to be infinitesimal). Sticking to the specific form (20.2) for the evolution operator, we find[2]

$$\begin{aligned} K(x,t;x_0,t_0) &= \langle x|e^{-iH(x,p)(t-t_0)/\hbar}|x_0\rangle \\ &= \langle x_N|e^{-iH(N\delta t)/\hbar}|x_0\rangle \\ &= \langle x_N|e^{-iH\delta t/\hbar}\left[\int dx_{N-1}|x_{N-1}\rangle\langle x_{N-1}|\right]e^{-iH\delta t/\hbar}\cdots e^{-iH\delta t/\hbar}\left[\int dx_1|x_1\rangle\langle x_1|\right]e^{-iH\delta t/\hbar}|x_0\rangle. \end{aligned} \tag{20.10}$$

Then we have an $(N-1)$-dimensional integral over a product of N matrix elements,

$$K(x,t;x_0,t_0) = \int\left(\prod_{j=1}^{N-1} dx_j\right)\langle x_N|e^{-iH\delta t/\hbar}|x_{N-1}\rangle\langle x_{N-1}|e^{-iH\delta t/\hbar}|x_{N-2}\rangle\cdots\langle x_1|e^{-iH\delta t/\hbar}|x_0\rangle. \tag{20.11}$$

This is now an infinite-dimensional integral, since we have in mind the limit $N\longrightarrow\infty$.

20.1.2 Splitting the Evolution Operator

To continue, we will assume a standard particle Hamiltonian of the form

$$H = \frac{p^2}{2m} + V(x) = T(p) + V(x), \tag{20.12}$$

where $p = (\hbar/i)\partial_x$ as usual, and while we are treating only one dimension, the generalization to multiple dimensions is straightforward. The Baker–Campbell–Hausdorff expansion[3] allows us to split the evolution operator with this Hamiltonian according to

$$e^{-iH\delta t/\hbar} = e^{-iT(p)\delta t/\hbar}e^{-iV(x)\delta t/\hbar} + O(\delta t^2). \tag{20.13}$$

[2]We are following here the phase-space construction of the path integral, as in Mark Srednicki, *Quantum Field Theory* (Cambridge, 2007), Chapter 6 (draft available at `http://web.physics.ucsb.edu/~mark/ms-qft-DRAFT.pdf`).

[3]R. M. Wilcox, ''Exponential Operators and Parameter Differentiation in Quantum Physics,'' *Journal of Mathematical Physics* **8**, 962 (1967) (doi: 10.1063/1.1705306).

(For more on this operator splitting, see Section 26.1.) The correction term here is negligible in the following sense: there are N such correction terms that add together, which gives an error $N\delta t^2 = \delta t$, which vanishes as $\delta t \longrightarrow 0$. Then considering any one of the matrix elements in the integrand of (20.11), we have

$$\begin{aligned}\langle x_2|e^{-iH\delta t/\hbar}|x_1\rangle &= \langle x_2|e^{-iT(p)\delta t/\hbar}e^{-iV(x)\delta t/\hbar}|x_1\rangle \\ &= \int dp_1\, \langle x_2|e^{-iT(p)\delta t/\hbar}|p_1\rangle\langle p_1|e^{-iV(x)\delta t/\hbar}|x_1\rangle,\end{aligned} \tag{20.14}$$

where we have inserted a momentum-space identity operator between the operators. We should be careful to emphasize that the position and momentum appearing in the exponentials are operators. Since we have inner products of eigenstates of these operators, we can simply replace them by the appropriate eigenvalues,

$$\begin{aligned}\langle x_2|e^{-iH\delta t/\hbar}|x_1\rangle &= \int dp_1\, \langle x_2|p_1\rangle\langle p_1|x_1\rangle e^{-iT(p_1)\delta t/\hbar}e^{-iV(x_1)\delta t/\hbar} \\ &= \frac{1}{2\pi\hbar}\int dp_1\, e^{ip_1(x_2-x_1)/\hbar}e^{-iT(p_1)\delta t/\hbar}e^{-iV(x_1)\delta t/\hbar},\end{aligned} \tag{20.15}$$

where we have used the inner product

$$\langle x|p\rangle = \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}}, \tag{20.16}$$

which is just the momentum eigenstate expressed in the position basis. Note that in introducing the momentum integral, we have also introduced a Fourier transformation. In fact, by tacking on an extra $|x_2\rangle$ on the front of Eq. (20.14) and then integrating over x_2, we have the recipe for propagating the initial state $|x_2\rangle$ over a small time δt: first propagate using the potential-energy exponential operator, which is diagonal in x, then Fourier transform to momentum space, where we may apply the kinetic operator, and finally transform back to position space to express the result in the x_2 basis. This is in fact the basis for split-operator methods for numerically solving the Schrödinger equation (Chapter 26).

Since $T(p)$ is quadratic, the momentum integral in the mini-propagator (20.15) is a Gaussian integral, so we will go ahead and carry it out by completing the square. Letting $\delta x_1 := x_2 - x_1$,

$$\begin{aligned}\langle x_2|e^{-iH\delta t/\hbar}|x_1\rangle &= \frac{1}{2\pi\hbar}\int dp_1\, e^{ip_1\delta x_1/\hbar}e^{-ip_1^2\delta t/2m\hbar}e^{-iV(x_1)\delta t/\hbar} \\ &= \frac{1}{2\pi\hbar}e^{-iV(x_1)\delta t/\hbar}\int dp_1\, e^{-i(p_1-m\delta x_1/\delta t)^2\delta t/2m\hbar}e^{im(\delta x_1)^2/2\hbar(\delta t)} \\ &= \frac{1}{2\pi\hbar}e^{im(\delta x_1)^2/2\hbar(\delta t)}e^{-iV(x_1)\delta t/\hbar}\int dp_1\, e^{-ip_1^2\delta t/2m\hbar},\end{aligned} \tag{20.17}$$

where we have shifted away the constant offset in the Gaussian integrand. The integral here is not strictly convergent, but we can help it by nudging it into the territory of convergent Gaussian integrals:

$$\int_{-\infty}^{\infty} dx\, e^{-i\alpha x^2} = \lim_{\beta\to 0}\int_{-\infty}^{\infty} dx\, e^{-(\beta+i\alpha)x^2} = \lim_{\beta\to 0}\sqrt{\frac{\pi}{\beta+i\alpha}} = \sqrt{\frac{\pi}{i\alpha}}. \tag{20.18}$$

Thus,

$$\begin{aligned}\langle x_2|e^{-iH\delta t/\hbar}|x_1\rangle &= \frac{1}{2\pi\hbar}e^{im(\delta x_1)^2/2\hbar(\delta t)}e^{-iV(x_1)\delta t/\hbar}\sqrt{\frac{2\pi m\hbar}{i\delta t}} \\ &= \sqrt{\frac{m}{i2\pi\hbar\delta t}}\, e^{im(\delta x_1)^2/2\hbar(\delta t)}e^{-iV(x_1)\delta t/\hbar} \\ &= \sqrt{\frac{m}{i2\pi\hbar\delta t}}\, e^{im\dot{x}_1^2(\delta t)/2\hbar}e^{-iV(x_1)\delta t/\hbar} \\ &= \sqrt{\frac{m}{i2\pi\hbar\delta t}}\, e^{iL(x_1,\dot{x}_1)\delta t/\hbar},\end{aligned} \tag{20.19}$$

where we have taken

$$\dot{x} = \frac{\delta x}{\delta t} \tag{20.20}$$

in view of the $\delta t \longrightarrow 0$ limit, and we have identified

$$L(x, \dot{x}) = \frac{1}{2} m \dot{x}^2 - V(x) \tag{20.21}$$

as the classical Lagrangian.

20.1.3 Functional Integral

Now collecting all N matrix elements (20.19) and putting them into the propagator (20.11), we have

$$\begin{aligned} K(x,t;x_0,t_0) &= \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) e^{iL(x_{N-1},\dot{x}_{N-1})\delta t/\hbar} e^{iL(x_{N-2},\dot{x}_{N-2})\delta t/\hbar} \cdots e^{iL(x_0,\dot{x}_0)\delta t/\hbar} \\ &= \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \left(\prod_{j=0}^{N-1} e^{iL(x_j,\dot{x}_j)\delta t/\hbar}\right). \end{aligned} \tag{20.22}$$

Defining the path-integral differential

$$Dx := \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \prod_{j=1}^{N-1} dx_j, \tag{20.23}$$
(functional-integral differential)

we can simplify the notation in Eq. (20.22) and recombine the exponentials to write

$$K(x,t;x_0,t_0) = \int Dx \, \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt\, L(x,\dot{x})\right]. \tag{20.24}$$
(Feynman propagator)

This is a **functional integral** or **path integral**, as we are integrating the **action functional**

$$S[x(t)] = \int_{t_0}^{t} dt'\, L[x(t'), \dot{x}(t')], \tag{20.25}$$
(action functional)

in terms of which the propagator has the even simpler form

$$K(x,t;x_0,t_0) = \int Dx \, \exp\left(\frac{i}{\hbar} S[x]\right), \tag{20.26}$$
(Feynman propagator)

noting that the integration is with respect to the *trajectory* (or function) $x(t)$. Recall that we discussed the related functional *differentiation* in Section 8.2.1, and we discussed the Lagrangian and classical variational principles in Section 8.2.2. Note that the normalization factor in (20.23) is highly divergent, but it is typically hidden inside the differential here since it typically drops out of calculations on the way to a physical result.

20.1.4 Phase-Space Path Integral

An alternate form of the path integral, before carrying out the momentum integrals, comes about by rewriting the first line of Eq. (20.17) in terms of $\dot{x}_1 = \delta x_1/\delta t$ and the Hamiltonian $H(x,p)$ as

$$\langle x_2 | e^{-iH\delta t/\hbar} | x_1 \rangle = \frac{1}{2\pi\hbar} \int dp_1 \, e^{ip_1\dot{x}_1\delta t/\hbar} e^{-iH(x_1,p_1)\delta t/\hbar}. \tag{20.27}$$

Then collecting all the matrix elements, we can now write the **phase-space path integral**[4]

$$K(x,t;x_0,t_0) = \int Dx\, Dp\, \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt\, \left[p\dot{x} - H(x,p)\right]\right],$$

(phase-space propagator) (20.28)

where the integration measure in phase space is

$$Dx\, Dp := \frac{1}{(2\pi\hbar)^{(N-1)/2}}\left(\prod_{j=1}^{N-1} dx_j\right)\left(\prod_{j=0}^{N-1} dp_j\right).$$

(functional-integration measure) (20.29)

This looks similar to before, except that we now have the phase-space action—the integral of $p\dot{x} - H$—in the propagator instead of the Lagrangian action, and the integration is over both positions and momenta. (Note that like in the classical application of the phase-space action, as in Section 8.2.2, the endpoints of $x(t)$ are pinned to x_0 and x, while the momentum endpoints are not similarly restricted.) At this point, in the case of a Hamiltonian quadratic in p, the integral is Gaussian and can be carried out analytically as above. The stationary-phase condition on the p integral gives the classical Hamilton equation

$$\partial_p \left[p\dot{x} - H(x,p)\right] = \dot{x} - \partial_p H(x,p) = 0, \tag{20.30}$$

which can be used to eliminate p in favor of $\dot{x}$ in $p\dot{x} - H(x,p)$ to reproduce the Lagrangian, and thus the Lagrangian form of the propagator (20.24).

Note that while the positions and momenta were introduced in similar ways, there is a fundamental asymmetry between them.[5] In the infinitesimal propagator from x_j to x_{j+1}, we regard the position as moving smoothly between these two coordinates in a time δt. However, the momentum is a *constant* p_j over the same time interval. So in this **time-slicing** construction (the construction with finite N), the path position follows a continuous, polygonal path, while the momentum is discontinuous and piecewise constant.

20.1.5 Example: Free-Particle Propagator

To evaluate the propagator (20.24) for the free particle, $L(x,\dot{x}) = m\dot{x}^2/2$, it is easiest to go back to the discrete form (20.22), which we can write explicitly as

$$\begin{aligned} K(x,t;x_0,t_0) &= \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right)\left(\prod_{j=0}^{N-1} e^{im\dot{x}_j^2\delta t/2\hbar}\right) \\ &= \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right)\left(\prod_{j=0}^{N-1} e^{im(x_{j+1}-x_j)^2/2\hbar\delta t}\right). \end{aligned} \tag{20.31}$$

Now let

$$x_j = \bar{x}_j + \delta_j, \tag{20.32}$$

where $\bar{x}_j$ is the straight-line path from x_0 to $x = x_N$,

$$\bar{x}_j = x_0 + \left(\frac{x - x_0}{t - t_0}\right)(t_j - t_0) =: x_0 + \bar{v}(t_j - t_0), \tag{20.33}$$

where $\bar{v} = (x - x_0)/(t - t_0)$ is the mean velocity of the path, and the δ_j are the "fluctuations" about this average. In particular, note that

$$\delta_0 = \delta_N = 0. \tag{20.34}$$

[4]This general form of the path integral in phase space was given by H. Davies, "Hamiltonian approach to the method of summation over Feynman histories," *Mathematical Proceedings of the Cambridge Philosophical Society* **59**, 147 (1963) (doi: 10.1017/S0305004100002097).

[5]H. Davies, *op. cit.*; Claude Garrod, "Hamiltonian Path-Integral Methods," *Reviews of Modern Physics* **38**, 483 (1966) (doi: 10.1103/RevModPhys.38.483); M. S. Marinov, "Path integrals in quantum theory: An outlook of basic concepts," *Physics Reports* **60**, 1 (1980) (doi: 10.1016/0370-1573(80)90111-8).

For the increments in the path integral, we may write

$$x_{j+1} - x_j = (\bar{x}_{j+1} + \delta_{j+1}) - (\bar{x}_j + \delta_j) = (\delta_{j+1} - \delta_j) + \bar{v}\delta t, \tag{20.35}$$

and thus the Gaussian factors in the path integral (20.31) involve squared differences of the form

$$(x_{j+1} - x_j)^2 = (\delta_{j+1} - \delta_j + \bar{v}\delta t)^2 = (\delta_{j+1} - \delta_j)^2 + 2(\delta_{j+1} - \delta_j)\bar{v}\delta t + \bar{v}^2\delta t^2. \tag{20.36}$$

This means that the path integral (20.31) splits into a product of three components, each involving Gaussian functions of each of these terms. For example, the last term leads to the product

$$\prod_{j=0}^{N-1} e^{im\bar{v}^2\delta t/2\hbar} = e^{im\bar{v}^2 N\delta t/2\hbar} = e^{im\bar{v}^2(t-t_0)/2\hbar} = e^{im(x-x_0)^2/2\hbar(t-t_0)}, \tag{20.37}$$

while the second term

$$\prod_{j=0}^{N-1} e^{im(\delta_{j+1}-\delta_j)\bar{v}/\hbar} = e^{im(\delta_N-\delta_0)\bar{v}/\hbar} = 1, \tag{20.38}$$

in view of Eq. (20.34). Collecting these with the remaining product leads to the path integral

$$K(x,t;x_0,t_0) = \Big(\frac{m}{i2\pi\hbar\delta t}\Big)^{N/2} e^{im(x-x_0)^2/2\hbar(t-t_0)} \int \left(\prod_{j=1}^{N-1} d\delta_j\right)\left(\prod_{j=0}^{N-1} e^{im(\delta_{j+1}-\delta_j)^2/2\hbar\delta t}\right), \tag{20.39}$$

where we should keep in mind that $\delta_0 = \delta_N = 0$ in the $j = 0$ and $j = N-1$ factors. Also, note that since the coordinate transformation from x_j to δ_j was just a coordinate shift, there is no Jacobian factor to worry about. In making the coordinate transformation here, we have reduced the problem of computing $K(x,t;x_0,t_0)$ to that of a periodic path integral of the form $K(0,t;0,t_0)$.

To finish up the path integral, we can begin by examining the δ_j integral, which involves the terms

$$\begin{aligned}
(\delta_2 - \delta_1)^2 + (\delta_1 - \delta_0)^2 &= \delta_2^2 - 2\delta_2\delta_1 + 2\delta_1^2 \\
(\delta_{j+1} - \delta_j)^2 + (\delta_j - \delta_{j-1})^2 &= \delta_{j+1}^2 - 2\delta_{j+1}\delta_j + 2\delta_j^2 - 2\delta_j\delta_{j-1} + \delta_{j-1}^2 \\
(\delta_N - \delta_{N-1})^2 + (\delta_{N-1} - \delta_{N-2})^2 &= 2\delta_{N-1}^2 - 2\delta_{N-1}\delta_{N-2} + \delta_{N-2}^2
\end{aligned} \tag{20.40}$$

If we add up these contributions, we have an integral of the form

$$K(x,t;x_0,t_0) = \Big(\frac{m}{i2\pi\hbar\delta t}\Big)^{N/2} e^{im(x-x_0)^2/2\hbar(t-t_0)} \int \left(\prod_{j=1}^{N-1} d\delta_j\right) \exp\left[\frac{im}{2\hbar\delta t}\left(\delta_\alpha M_{\alpha\beta}\delta_\beta\right)\right], \tag{20.41}$$

where now $M_{\alpha\beta}$ is a matrix of dimension $(N-1)\times(N-1)$, and has twos along the diagonal, and -1's along the adjacent diagonals:

$$(M_{\alpha\beta}) = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix}. \tag{20.42}$$

This matrix has determinant

$$\det(M_{\alpha\beta}) = N, \tag{20.43}$$

as we will now show. Let $\det_n$ denote the determinant of the $n \times n$ version of the same matrix. Then expanding the determinant in terms of minors along the first row gives the recursion relation

$$\det_n = 2\det_{n-1} - \det_{n-2}. \tag{20.44}$$

With the initial values $\det_1 = 2$ and $\det_2 = 3$, we can see that the recursion and initial conditions are satisfied by

$$\det_n = n + 1. \tag{20.45}$$

This establishes the determinant (20.43).

Now to finish up with the path integral, we can use the Gaussian integration formula

$$\int d^N z \, \exp\left[-z_\alpha S_{\alpha\beta} z_\beta\right] = \frac{\pi^{N/2}}{\sqrt{\det(S_{\alpha\beta})}}, \tag{20.46}$$

for an $N \times N$ matrix $S_{\alpha\beta}$ [see, for example, the normalized Gaussian function in Eq. (4.165)], and adapt it for complex Gaussians as in Eq. (20.18). The result is

$$\begin{aligned} K(x,t;x_0,t_0) &= \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} e^{im(x-x_0)^2/2\hbar(t-t_0)} \left(\frac{i2\pi\hbar\delta t}{m}\right)^{(N-1)/2} \frac{1}{\sqrt{N}} \\ &= \sqrt{\frac{m}{i2\pi\hbar N\delta t}}\, e^{im(x-x_0)^2/2\hbar(t-t_0)} \end{aligned} \tag{20.47}$$

or finally,

$$K(x,t;x_0,t_0) = \sqrt{\frac{m}{i2\pi\hbar(t-t_0)}} \exp\left[\frac{im(x-x_0)^2}{2\hbar(t-t_0)}\right] \tag{20.48}$$

(free-particle propagator)

Note that the removal of the mean $\bar{x}(t)$ from the path $x(t)$ to focus on the fluctuations $\delta(t)$ here is the same idea as the Brownian-bridge construction $B(t) = W(t) - tW(1)$ [Eq. (17.287)], where the (linear) mean drift is subtracted away to produce a periodic path.

20.2 Classical Limit

One nice feature of the path integral is that it gives a clear path to the classical limit, at least in a heuristic argument.[6] Returning to the propagator (20.26),

$$K(x,t;x_0,t_0) = \int Dx \, \exp\left(\frac{i}{\hbar} S[x]\right), \tag{20.49}$$

note that as $\hbar \longrightarrow 0$, the exponential becomes highly oscillatory. Except, that is, where the rest of the exponent, $S[x]$, vanishes. Thus, we might expect the oscillatory paths to roughly cancel, and the integral should be dominated by the "stationary" paths, for which slight variations don't change the action. Dropping all but these stationary paths is known as the **stationary-phase approximation**, and amounts to restricting the path integral to *classical* paths satisfying

$$\delta S[x] = 0, \tag{20.50}$$

(stationary-phase condition)

which is just the variational principle that generates the Euler-Lagrange equation. That is, when $\hbar$ is small compared to the action of the system, the path integral is dominated by the contribution from the classical path. Paths that deviate from the classical path by even a small amount on the classical scale are suppressed because they oscillate wildly and cancel other, slightly deviant paths. When $\hbar$ is larger relative to the action, the paths have more leeway to "explore" around the classical path without suffering the fate of cancellation.

[6] Richard P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, 1965), Section 2-3, p. 29.

20.2.1 Semiclassical Propagator

With this picture in mind, we can go on and develop an approximate form for the propagator in the semiclassical regime.[7] Since we expect the classical path x_c to be important, we will consider small deviations δx from it:

$$x(t) = x_\mathrm{c}(t) + \delta x(t). \tag{20.51}$$

Then expanding the action in a functional Taylor series around the classical path, we find

$$S[x] = S[x_\mathrm{c}] + \delta S[x_\mathrm{c}] + \frac{1}{2}\delta^2 S[x_\mathrm{c}] + \cdots \tag{20.52}$$
(series expansion of the action)

up to second order. Note that by definition, $\delta S[x_\mathrm{c}] = 0$, but we leave the first-order term here to illustrate the general pattern.

Let's take a moment to examine the notation here. Given the action functional (20.25),

$$S[x(t)] = \int_{t_0}^{t} dt'\, L[x(t'), \dot{x}(t')], \tag{20.53}$$

recall that we define the **first variation** $\delta S[x]$ of the functional as

$$\delta S[x; \delta x] := \lim_{\epsilon\to 0} \frac{S[x + \epsilon\,\delta x] - S[x]}{\epsilon} = \partial_\epsilon S[x + \epsilon\,\delta x]\Big|_{\epsilon=0}, \tag{20.54}$$
(first variation)

in analogy to the usual derivative, but the perturbation to the argument is a function $\delta x(t)$. Note that this variation is, roughly speaking, proportional to δx, and hence why we can use it in the series expansion. Then the **second variation** is the variation of the first variation:

$$\begin{aligned}\delta^2 S[x; \delta x; \delta x'] &:= \lim_{\epsilon\to 0} \frac{\delta S[x + \epsilon\,\delta x; \delta x'] - \delta S[x; \delta x']}{\epsilon} \\ &= \partial_\epsilon \delta S[x + \epsilon\,\delta x; \delta x']\Big|_{\epsilon=0} \\ &= \partial_\epsilon \partial_{\epsilon'} S[x + \epsilon\,\delta x + \epsilon'\,\delta x']\Big|_{\epsilon=\epsilon'=0}.\end{aligned} \tag{20.55}$$
(second variation)

This pattern of definitions can obviously continue indefinitely for higher-order variations.

The variations can then serve as the definition for functional derivatives. In terms of the first variation, the **first functional derivative** $\delta S/\delta x$ is defined such that the inner product with the perturbation δx gives the first variation:

$$\left\langle \frac{\delta S}{\delta x},\, \delta x \right\rangle := \int_{t_0}^{t} dt'\, \frac{\delta S}{\delta x(t')}\, \delta x(t') := \delta S[x; \delta x]. \tag{20.56}$$
(first functional derivative)

The **second functional derivative** is similarly defined in terms of the second variation as

$$\left\langle \delta x,\, \frac{\delta^2 S}{\delta x^2}\, \delta x \right\rangle := \int_{t_0}^{t} dt' \int_{t_0}^{t} dt''\, \delta x(t')\, \frac{\delta^2 S}{\delta x(t')\, \delta x(t'')}\, \delta x(t'') := \delta^2 S[x; \delta x; \delta x']. \tag{20.57}$$
(second functional derivative)

Note that these are all reasonably straightforward generalizations of the same derivatives for a scalar-valued function of a vector $S(x_j)$, where the first derivative is a vector of partial derivatives $\partial S/\partial x_j$, and the second derivative is a matrix of derivatives $\partial^2 S/\partial x_j \partial x_k$.

[7] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, 1981), Chapter 13; Hagen Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 5th ed. (World Scientific, 2009), Chapter 4.

In terms of the functional derivatives, we can write the series expansion (20.52) for the action as

$$S[x] = S[x_c] + \int_{t_0}^{t} dt' \left.\frac{\delta S}{\delta x(t')}\right|_{x=x_c} \delta x(t') + \frac{1}{2}\int_{t_0}^{t} dt' \int_{t_0}^{t} dt'' \left.\frac{\delta^2 S}{\delta x(t')\,\delta x(t'')}\right|_{x=x_c} \delta x(t')\,\delta x(t'') + \cdots$$

(series expansion of the action) (20.58)

as an alternative to Eq. (20.52). Again, $\delta S/\delta x$ vanishes when evaluated at x_c, and again this just generalizes the discrete form of the Taylor series.

Inserting this latter expansion into the path integral (20.49), we note again that the first-order term is zero, and we also note that the second-order term gives a Gaussian factor in the fluctuations $\delta x(t)$:

$$K_{sc}(x,t;x_0,t_0) = e^{iS[x_c]/\hbar} \int Dx \, \exp\left(\frac{i}{2\hbar}\int dt' \int dt''\, \delta x(t) \left.\frac{\delta^2 S}{\delta x(t')\,\delta x(t'')}\right|_{x=x_c} \delta x(t'')\right). \tag{20.59}$$

We will be careful and go back to the discrete (time-sliced) form for the path integral to evaluate it, as in Eq. (20.22):

$$K_{sc}(x,t;x_0,t_0) = \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} e^{iS[x_c]/\hbar} \int \left(\prod_{j=1}^{N-1} d\delta x_j\right) \exp\left(\frac{i}{2\hbar}\sum_{j=1}^{N-1}\sum_{j'=1}^{N-1} \delta x_j \left.\frac{\partial^2 S}{\partial x_j\,\partial x_{j'}}\right|_{x=x_c} \delta x_{j'}\right). \tag{20.60}$$

Note that the sums in the exponential only run from 1 to $N-1$ because we regard the endpoints x_0 and x_N to be fixed. Then using the Gaussian-integral formula (20.46),

$$K_{sc}(x,t;x_0,t_0) = \sqrt{\frac{m}{i2\pi\hbar\delta t}} \left(\frac{m}{\delta t}\right)^{(N-1)/2} \det{}^{-1/2}\left(\frac{\partial^2 S}{\partial x_j\,\partial x_{j'}}\right) \exp\left(\frac{i}{\hbar}S[x_c]\right), \tag{20.61}$$

where the determinant is over an $(N-1)\times(N-1)$ matrix. In continuous language, we can write this result in terms of a **functional determinant** as

$$K_{sc}(x,t;x_0,t_0) = \sqrt{\frac{m}{i2\pi\hbar\delta t}}\, \det{}^{-1/2}\left[\frac{\delta t}{m}\left(\frac{\delta^2 S}{\delta x^2}\right)\right] \exp\left(\frac{i}{\hbar}S[x_c]\right),$$

(semiclassical propagator) (20.62)

where we may take the determinant to be defined by the discrete expression (20.61). Note that in deriving this propagator, we have considered lowest-order quantum fluctuations about the classical trajectory, in what amounts to making a Gaussian approximation for the fluctuations. While the exponential gives the quantum phase factor along the classical path, the determinant gives the correction to the amplitude as the size of the Gaussian fluctuations stretches or compresses in phase space.

For the action (20.53), we can write out the first action variation

$$\begin{aligned}\delta S[x;\delta x] &= \int_{t_0}^{t} dt' \left(\frac{\partial L}{\partial x}\,\delta x + \frac{\partial L}{\partial \dot{x}}\,\delta \dot{x}\right) \\ &= \int_{t_0}^{t} dt' \left(\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}}\right)\delta x\end{aligned} \tag{20.63}$$

and the second variation

$$\begin{aligned}\delta^2 S[x;\delta x] &= \int_{t_0}^{t} dt' \left(\frac{\partial^2 L}{\partial x^2}\,\delta x\,\delta x + 2\frac{\partial L}{\partial x \partial \dot{x}}\,\delta x\,\delta \dot{x} + \frac{\partial^2 L}{\partial \dot{x}^2}\,\delta \dot{x}\,\delta \dot{x}\right), \\ &= \int_{t_0}^{t} dt' \int_{t_0}^{t} dt''\, \delta(t'-t'') \\ &\quad \times \left(\frac{\partial^2 L}{\partial x(t')\,\partial x(t'')}\,\delta x(t')\,\delta x(t'') + 2\frac{\partial L}{\partial x(t')\,\partial \dot{x}(t'')}\,\delta x(t')\,\delta \dot{x}(t'') + \frac{\partial^2 L}{\partial \dot{x}(t')\,\partial \dot{x}(t'')}\,\delta \dot{x}(t')\,\delta \dot{x}(t'')\right),\end{aligned} \tag{20.64}$$

and thus the functional derivatives

$$\frac{\delta S[x]}{\delta x(t)} = \frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}} \tag{20.65}$$

and

$$\frac{\delta^2 S[x]}{\delta x(t)\,\delta x(t')} = \delta(t-t')\left(\frac{\partial^2 L}{\partial x^2(t)} + 2\frac{\partial L}{\partial x(t)\,\partial \dot{x}(t)}\frac{d}{dt'}\right) - \frac{d}{dt}\delta(t-t')\frac{\partial^2 L}{\partial \dot{x}^2(t)}\frac{d}{dt'}. \tag{20.66}$$

In this last expression, note that all derivative operators d/dt operate on everything to the right.

For a particle Lagrangian (20.21),

$$L(x,\dot{x}) = \frac{1}{2}m\dot{x}^2 - V(x), \tag{20.67}$$

the first functional derivative is

$$\frac{\delta S[x]}{\delta x(t)} = -V'(x) - m\ddot{x}, \tag{20.68}$$

while the second derivative becomes

$$\frac{\delta^2 S[x]}{\delta x(t)\,\delta x(t')} = -\delta(t-t')\,V''(x) - m\frac{d}{dt}\delta(t-t')\frac{d}{dt'}, \tag{20.69}$$

in which case the functional determinant in the propagator (20.62) becomes

$$\begin{aligned}\det{}^{-1/2}\left[\frac{\delta t}{m}\left(\frac{\delta^2 S}{\delta x^2}\right)\right] &= \delta t^{-(N-1)/2}\det{}^{-1/2}\left(-\frac{d}{dt}\delta(t-t')\frac{d}{dt'} - \delta(t-t')\,\frac{V''(x_{\rm c})}{m}\right)\\ &= \delta t^{-(N-1)}\det{}^{-1/2}\left(-\partial_t^2 - \frac{V''(x_{\rm c})}{m}\right),\end{aligned} \tag{20.70}$$

where we have used the discrete form of the delta function, which is the identity times δt^{-1}. In fact, we have already evaluated the last determinant here in the case $V(x) = 0$ for the free particle. By noting that the second-order, finite-difference operator for the second derivative can be written

$$\Delta_t^{(2)}\psi(t) := \frac{\psi(t+\delta t) - 2\psi(t) + \psi(t-\delta t)}{\delta t^2}, \tag{20.71}$$

such that

$$\partial_t^2\psi(t) = \Delta_t^{(2)}\psi(t) + O(\delta t^3), \tag{20.72}$$

the matrix $\mathbf{M}$ in Eq. (20.42) is the discrete representation for the operator $-\delta t^2\partial_t^2$. Recalling from Eq. (20.43) that $\det\mathbf{M} = N$, then

$$\det\left(-\partial_t^2\right) = \delta t^{-2(N-1)}N, \tag{20.73}$$

and this means, for example, that we can write the propagator (20.62) in continuous notation in terms of a ''renormalized'' determinant:

$$K_{\rm sc}(x,t;x_0,t_0) = \sqrt{\frac{m}{i2\pi\hbar(t-t_0)}}\,\sqrt{\frac{\det(-\partial_t^2)}{\det[-\partial_t^2 - V''(x_{\rm c})/m]}}\,\exp\left(\frac{i}{\hbar}S[x_{\rm c}]\right), \tag{20.74}$$

(semiclassical propagator)

where we used $t - t_0 = N\delta t$.

20.2.1.1 Gel'fand–Yaglom Method

To evaluate the functional determinant further here, we can extend the recursion method that we used in Section 20.1.5 for the free-particle determinant. In place of the matrix (20.42) for the operator $-\delta t\partial_t^2$, we can compute the determinant we need in terms of a similar $(N-1)\times(N-1)$ matrix:

$$\frac{\det\left[-\partial_t^2 - V''(x_c)/m\right]}{\det(-\partial_t^2)} = \frac{1}{N}\det\begin{bmatrix} 2-\delta t^2\omega_{N-1}^2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2-\delta t^2\omega_{N-2}^2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2-\delta t^2\omega_{N-3}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2-\delta t^2\omega_2^2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2-\delta t^2\omega_1^2 \end{bmatrix}, \tag{20.75}$$

where

$$\omega_j^2 := \frac{V''(x_{c,j})}{m} = \frac{V''[x_c(t_j)]}{m} \tag{20.76}$$

defines an effective, time-dependent, harmonic-oscillator frequency based on the curvature of the potential along the classical path. As in the free-particle case, the $n\times n$ version of the determinant satisfies the recursion relation

$$\det_n = \left[2-\delta t^2\omega_n^2\right]\det_{n-1} - \det_{n-2}, \tag{20.77}$$

with initial values

$$\begin{aligned} \det_1 &= 2-\delta t^2\omega_1^2 \\ \det_2 &= (2-\delta t^2\omega_2^2)(2-\delta t^2\omega_1^2) - 1. \end{aligned} \tag{20.78}$$

To cast this recursion into continuous form, note that we can rewrite Eq. (20.77) as

$$\frac{\det_n - 2\det_{n-1} + \det_{n-2}}{\delta t^2} + \omega_n^2\det_{n-1} = 0. \tag{20.79}$$

Defining

$$D_n := \delta t\,\det_n, \tag{20.80}$$

the recursion is the same,

$$\frac{D_n - 2D_{n-1} + D_{n-2}}{\delta t^2} - \omega_n^2 D_{n-1} = 0. \tag{20.81}$$

but we can recast the initial conditions as

$$\begin{aligned} D_1 &= \delta t(2-\delta t^2\omega_1^2) \\ \frac{D_2 - D_1}{\delta t} &= (2-\delta t^2\omega_2^2)(2-\delta t^2\omega_1^2) - 3 + \delta t^2\omega_1^2. \end{aligned} \tag{20.82}$$

In the continuum limit $\delta t \longrightarrow 0$, $D(t)$ then satisfies the differential equation

$$\begin{aligned} \left[\partial_t^2 + \omega^2(t)\right]D(t) &= 0 \\ D(0) &= 0 \\ \dot{D}(0) &= 1. \end{aligned}$$

(continuous representation of the determinant) (20.83)

The determinant then is simply $D(t)/\delta t$, where t here is the final time of the propagator. This recasting of the functional determinant into the solution of an ordinary differential equation is specific to determinants of Schrödinger-type operators $-\partial_x^2 + V(x)$ in one dimension, and this is the **Gel'fand–Yaglom method** for computing functional determinants.[8]

[8] I. M. Gel'fand and A. M. Yaglom, "Integration in Functional Spaces and its Applications in Quantum Physics," *Journal Mathematical Physics* **1**, 48 (1960) (doi: 10.1063/1.1703636); Hagen Kleinert, *op. cit.*, Section 2.4; M. Chaichian and A. Demichev, *Path Integrals in Physics, Volume I: Stochastic Processes in Quantum Mechanics* (Institute of Physics, 2001), p. 168.

20.2.1.2 Van Vleck–Morette Determinant

Now we will use the Gel'fand–Yaglom method to establish a simple form for the semiclassical functional determinant.[9] Using the classical equation of motion $m\ddot{x}_\mathrm{c} = -V'(x_\mathrm{c})$ and differentiating with respect to time leads to

$$\left[\partial_t^2 + \frac{V''(x_\mathrm{c})}{m}\right]\dot{x}_\mathrm{c} = 0. \tag{20.84}$$

This in turn implies that $\dot{x}_\mathrm{c}(t)$ is a solution to the differential equation (20.83). To emphasize this, we will write the solution

$$D_1(t) = \dot{x}_\mathrm{c}(t). \tag{20.85}$$

Since we have a second-order differential equation, there are two independent solutions D_1 and D_2, and the Wronskian determinant

$$W := \det\begin{bmatrix} D_1 & D_2 \\ \dot{D}_1 & \dot{D}_2 \end{bmatrix} = D_1\dot{D}_2 - D_2\dot{D}_1 \tag{20.86}$$

is constant, because the coefficient of ∂_t^2 is constant and the coefficient of ∂_t vanishes. Dividing through by D_1^2,

$$\frac{\dot{D}_2}{D_1} - \frac{D_2}{D_1^2}\dot{D}_1 = \frac{W}{D_1^2}, \tag{20.87}$$

and noting that the left-hand side is the derivative of D_2/D_1, we can integrate both sides to obtain

$$\frac{D_2(t)}{D_1(t)} - \frac{D_2(t_0)}{D_1(t_0)} = \int_{t_0}^{t} dt'\,\frac{W}{D_1^2(t')}, \tag{20.88}$$

or to simplify,

$$D_2(t) = \frac{D_2(t_0)}{D_1(t_0)}D_1(t) + W D_1(t)\int_{t_0}^{t}\frac{dt'}{D_1^2(t')}. \tag{20.89}$$

The general solution $D(t)$ can be written as a superposition $c_1D_1 + c_2D_2$, or

$$D(t) = c_1\dot{x}_\mathrm{c}(t) + c_2\dot{x}_\mathrm{c}(t)\int_{t_0}^{t}\frac{dt'}{\dot{x}_\mathrm{c}^2(t')}, \tag{20.90}$$

where we have absorbed some of the t_0-dependent constants in the coefficients. The initial conditions from (20.83) give $c_1 = 0$ from $D(t_0) = 0$, and $c_2 = \dot{x}_\mathrm{c}(t_0)$ from $\dot{D}(t_0) = 1$. Thus, we have the solution

$$D(t) = \dot{x}_\mathrm{c}(t)\,\dot{x}_\mathrm{c}(t_0)\int_{t_0}^{t}\frac{dt'}{\dot{x}_\mathrm{c}^2(t')} \tag{20.91}$$

to represent the functional derivative.

We will continue by transforming this into a simpler form. First, recall that a classical particle in one dimension, by virtue of the conserved energy

$$E = \frac{1}{2}m\dot{x}_\mathrm{c}^2 + V(x_\mathrm{c}) = \frac{p_\mathrm{c}^2}{2m} + V(x_\mathrm{c}), \tag{20.92}$$

can always be solved in the sense of obtaining an integral for $t(x)$:

$$t - t_0 = \int_{x_0}^{x} dx'\,\frac{m}{\sqrt{2m[E - V(x')]}} = \int_{x_0}^{x} dx'\,\frac{m}{p(x')}. \tag{20.93}$$

Differentiating this gives

$$\frac{\partial t}{\partial E} = -\int_{x_0}^{x} dx'\,\frac{m^2}{\big(2m[E - V(x')]\big)^{3/2}} = -\int_{x_0}^{x} dx'\,\frac{m^2}{p^3(x')}. \tag{20.94}$$

[9] Here we are mainly following M. Chaichian and A. Demichev, *op. cit.*, Section 2.2.3; see also Hagen Kleinert, *op. cit.*, Sections 2.4.5 and 4.3.

Then we can use this result to simplify the integral

$$\begin{aligned}\int_{t_0}^{t}\frac{dt'}{\dot{x}_{\rm c}^2(t')} &= m^2\int_{t_0}^{t}\frac{dt'}{p_{\rm c}^2(t')}\\ &= m^3\int_{t_0}^{t}\frac{\dot{x}_{\rm c}\,dt'}{p_{\rm c}^3(t')}\\ &= m^3\int_{x_0}^{x}\frac{dx_{\rm c}}{p_{\rm c}^3(x_{\rm c})}\\ &= -m\frac{\partial t}{\partial E}.\end{aligned} \tag{20.95}$$

Now to calculate $\partial E\partial t$. Starting with the phase-space action, we have

$$\begin{aligned}S[x_{\rm c}] &= \int_{t_0}^{t}dt'\,\Big[p_{\rm c}(t')\,\dot{x}_{\rm c}(t') - H(x_{\rm c},p_{\rm c})\Big]\\ &= \int_{x_0}^{x}dx_{\rm c}\,p_{\rm c}(x_{\rm c}) - (t-t_0)E.\end{aligned} \tag{20.96}$$

Differentiating with respect to the endpoint x,

$$\begin{aligned}\frac{\partial S[x_{\rm c}]}{\partial x} &= p_{\rm c}(x) + \int_{x_0}^{x}dx_{\rm c}\,\frac{\partial p_{\rm c}}{\partial x} - (t-t_0)\frac{\partial E}{\partial x}\\ &= p_{\rm c}(x) + \left[\int_{x_0}^{x}dx_{\rm c}\,\frac{\partial p_{\rm c}}{\partial E} - (t-t_0)\right]\frac{\partial E}{\partial x},\end{aligned} \tag{20.97}$$

where we pulled $\partial E/\partial x$ out of the integral, since we regard E as only depending on boundary conditions (e.g., on x_0 and $\dot{x}_0$, or x_0 and x). Then

$$\frac{\partial p_{\rm c}}{\partial E} = \left(\frac{\partial E}{\partial p_{\rm c}}\right)^{-1} = \left(\frac{\partial p_{\rm c}}{m}\right)^{-1} = \frac{1}{\dot{x}_{\rm c}}, \tag{20.98}$$

and so the second term in Eq. (20.97) vanishes, so that

$$\frac{\partial S[x_{\rm c}]}{\partial x} = p_{\rm c}(x). \tag{20.99}$$

Differentiating with respect to the initial coordinate,

$$\frac{\partial^2 S[x_{\rm c}]}{\partial x_0\partial x} = \frac{\partial p_{\rm c}(x)}{\partial x_0} = \frac{\partial}{\partial x_0}\sqrt{2m[E-V(x)]} = \frac{m}{p_{\rm c}(x)}\frac{\partial E}{\partial x_0}. \tag{20.100}$$

To find $\partial E/\partial x_0$, we can differentiate Eq. (20.96) with respect to x_0 to obtain

$$\frac{\partial S[x_{\rm c}]}{\partial x_0} = -p_{\rm c}(x_0) - (t-t_0)\frac{\partial E}{\partial x_0}, \tag{20.101}$$

and then differentiate with respect to t to obtain

$$\frac{\partial E}{\partial x_0} = -\frac{\partial^2 S[x_{\rm c}]}{\partial t\partial x_0}. \tag{20.102}$$

From Eq. (20.96), we can differentiate with respect to x_0

$$\begin{aligned}\frac{\partial S[x_{\rm c}]}{\partial x_0} &= -p_{\rm c}(x_0) + \int_{x_0}^{x}dx_{\rm c}\,\frac{\partial p_{\rm c}}{\partial x_0} - (t-t_0)\frac{\partial E}{\partial x_0}\\ &= -p_{\rm c}(x_0) + \left[\int_{x_0}^{x}dx_{\rm c}\,\frac{\partial p_{\rm c}}{\partial E} - (t-t_0)\right]\frac{\partial E}{\partial x_0}\\ &= -p_{\rm c}(x_0),\end{aligned} \tag{20.103}$$

in analogy to Eq. (20.99). Thus, Eq. (20.102) becomes

$$\frac{\partial E}{\partial x_0} = \frac{\partial p_\mathrm{c}(x_0)}{\partial t} = \frac{\partial p_\mathrm{c}(x_0)}{\partial E}\frac{\partial E}{\partial t} = \frac{m}{p_\mathrm{c}(x_0)}\frac{\partial E}{\partial t}, \tag{20.104}$$

and putting this in Eq. (20.100) we have

$$\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x} = \frac{m^2}{p_\mathrm{c}(x_0)p_\mathrm{c}(x)}\frac{\partial E}{\partial t}. \tag{20.105}$$

Solving for $\partial E/\partial t$,

$$\frac{\partial E}{\partial t} = \dot{x}_\mathrm{c}(x_0)\dot{x}_\mathrm{c}(x)\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x}. \tag{20.106}$$

Putting this into Eq. (20.95),

$$\int_{t_0}^{t} \frac{dt'}{\dot{x}_\mathrm{c}^2(t')} = -m\left(\dot{x}_\mathrm{c}(x_0)\dot{x}_\mathrm{c}(x)\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x}\right)^{-1}, \tag{20.107}$$

so the determinant solution $D(t)$ in Eq. (20.91) finally becomes

$$D(t) = -m\left(\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x}\right)^{-1} \tag{20.108}$$

The original determinant that we needed in Eq. (20.75) becomes

$$\frac{\det\left[-\partial_t^2 - V''(x_\mathrm{c})/m\right]}{\det(-\partial_t^2)} = \frac{1}{N\delta t}D(t) = \frac{m}{(t-t_0)}\left(-\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x}\right)^{-1} \tag{20.109}$$

At long last, the semiclassical propagator (20.74) simplifies to

$$K_\mathrm{sc}(x,t;x_0,t_0) = \frac{1}{\sqrt{2\pi i\hbar}}\sqrt{-\frac{\partial^2 S[x_\mathrm{c}]}{\partial x_0 \partial x}}\,\exp\left(\frac{i}{\hbar}S[x_\mathrm{c}]\right),$$

(semiclassical propagator) (20.110)

so that the determinant reduces to a derivative of the classical action at the propagator endpoints.

More generally, in multiple dimensions, this result generalizes to

$$K_\mathrm{sc}(\mathbf{x},t;\mathbf{x}_0,t_0) = \frac{1}{\sqrt{2\pi i\hbar}}\det{}^{1/2}\left[-\frac{\partial^2 S[\mathbf{x}_\mathrm{c}]}{\partial \mathbf{x}_0 \partial \mathbf{x}}\right]\,\exp\left(\frac{i}{\hbar}S[\mathbf{x}_\mathrm{c}]\right),$$

(semiclassical propagator) (20.111)

The determinant here,

$$\det\left[-\frac{\partial^2 S[\mathbf{x}_\mathrm{c}]}{\partial \mathbf{x}_0 \partial \mathbf{x}}\right] = \det\left[-\frac{\partial \mathbf{p}_0}{\partial \mathbf{x}}\right] = \det\left[-\frac{\partial \mathbf{p}}{\partial \mathbf{x}_0}\right],$$

(Van Vleck–Morette determinant) (20.112)

is called the **Van Vleck–Morette determinant**.[10] This determinant has an interpretation as follows. In the variational spirit of the path integral, the classical trajectory $\mathbf{x}_\mathrm{c}$ is specified in terms of its endpoints $\mathbf{x}_0$ and $\mathbf{x}$. Alternately, it can be specified in terms of the initial conditions $(\mathbf{x}_0, \mathbf{p}_0)$ or the end conditions $(\mathbf{x}, \mathbf{p})$. The Van Vleck–Morette determinant is the determinant of the coordinate transformation between the variational and initial/end specifications. Since the phase-space area of a small bundle of paths around

[10] The determinant was introduced in the semiclassical context by J. H. Van Vleck, *Proceedings of the National Academy of Sciences* **14**, 178 (1928) (doi: 10.1073/pnas.14.2.178); and introduced in the context of path integration by C. DeWitt-Morette, *Physical Review* **81**, 848 (1951) (doi: 10.1103/PhysRev.81.848). See also, for example, Bryce DeWitt, *The Global Approach to Quantum Field Theory*, vol. 1 (Oxford, 2003), Chapter 14.

the classical path is invariant along the path (classically, at least), roughly speaking the determinant maps this area into position space, to represent the effect of local convergence or divergence of classical trajectories on the amplitude. (In optics terminology, this is akin to using Gaussian beams instead of rays for tracing optical systems, and the spreading or converging of the beam waist influences the intensity at the beam center.)

We are also glossing over what happens if the functional determinant in Eq. (20.62) vanishes, such that the Van Vleck–Morette determinant diverges. This corresponds to the formation of **caustics** in the classical trajectories, and requires more careful treatment.[11]

20.2.2 Example: Semiclassical Propagator for the Harmonic Oscillator

As an example of the semiclassical propagator, we will evaluate this explicitly for a harmonic oscillator of frequency ω. Note that for this case, the semiclassical propagator (20.62) is *exact*, because it was based on a truncation at the second derivative, and higher derivatives will vanish anyway for the harmonic oscillator, which has a quadratic Lagrangian.

20.2.2.1 Classical Action

Thus, starting with the classical action,

$$S[x_{\rm c}] = \int_{t_0}^{t} dt' \left(\frac{m\dot{x}^2}{2} - \frac{m\omega^2 x^2}{2} \right), \tag{20.113}$$

we will integrate by parts on the first term to obtain

$$S[x_{\rm c}] = \frac{mx\dot{x}}{2}\bigg|_{t_0}^{t} - \frac{m}{2}\int_{t_0}^{t} dt'\, x\big(\ddot{x} + \omega^2 x\big). \tag{20.114}$$

The second factor in the integrand vanishes under the classical equations of motion, so we only have the boundary term

$$S[x_{\rm c}] = \frac{m}{2} x\dot{x}\big|_{t_0}^{t}. \tag{20.115}$$

We will then use the classical trajectory

$$x_{\rm c}(t) = x_0 \cos\omega(t-t_0) + \frac{\dot{x}_0}{\omega}\sin\omega(t-t_0) \tag{20.116}$$

to evaluate the action integral, with the result

$$\begin{aligned} S[x_{\rm c}] &= \frac{m}{2}\left[x_0\dot{x}_0\Big(\cos^2\omega(t-t_0) - \sin^2\omega(t-t_0)\Big) + \left(\frac{\dot{x}_0^2}{\omega} - \omega x_0^2\right)\cos\omega(t-t_0)\sin\omega(t-t_0)\right] - \frac{m}{2}x_0\dot{x}_0 \\ &= \frac{m}{2}\left[\left(\frac{\dot{x}_0^2}{\omega} - \omega x_0^2\right)\cos\omega(t-t_0)\sin\omega(t-t_0) - 2x_0\dot{x}_0\sin^2\omega(t-t_0)\right]. \end{aligned} \tag{20.117}$$

Now rearranging the general solution (20.113) to obtain $\dot{x}_0$ at the final spacetime coordinates (x,t),

$$\dot{x}_0 = \frac{\omega(x - x_0\cos\omega t)}{\sin\omega t} \tag{20.118}$$

we can eliminate the initial derivative. After a bit of algebra we find

$$S[x_{\rm c}] = \frac{m\omega}{2\sin\omega(t-t_0)}\Big[\big(x_0^2 + x^2\big)\cos\omega(t-t_0) - 2x_0x\Big]$$

(harmonic-oscillator action) (20.119)

for the classical action in terms of only the time interval and the end coordinates $x_0 = x_{\rm c}(t_0)$ and $x = x_{\rm c}(t)$.

[11] M. V. Berry and K. E. Mount, "Semiclassical approximations in wave mechanics," *Reports on Progress in Physics* **35**, 315 (1972) (doi: 10.1088/0034-4885/35/1/306).

20.2.2.2 Determinant

Then to obtain the Van Vleck–Morette determinant for the harmonic oscillator, we simply differentiate the action:

$$\frac{\partial^2 S[x_c]}{\partial x_0 \partial x} = -\frac{m\omega}{\sin\omega(t-t_0)}.$$

(Van Vleck–Morette determinant, harmonic-oscillator) (20.120)

Then assembling the parts of the semiclassical propagator (20.110), we have

$$K(x,t;x_0,t_0) = \sqrt{\frac{m\omega}{2\pi i\hbar \sin\omega(t-t_0)}}\ \exp\left(\frac{im\omega}{2\hbar\sin\omega(t-t_0)}\left[\left(x_0^2+x^2\right)\cos\omega(t-t_0) - 2x_0x\right]\right).$$

(propagator, harmonic oscillator) (20.121)

Again, though we derived this expression via the semiclassical propagator, this is identical to the fully quantum propagator. Any higher-order corrections to the path integral go as higher powers of $\hbar^{-1}$, and as functional derivatives of the action beyond second order. It is also easy to check that this expression reduces correctly to the free-particle propagator (20.48) in the limit $\omega \longrightarrow 0$. In deriving that result, we also used the classical (straight-line) path to simplify the calculation.

20.3 Monte-Carlo Methods in Quantum Statistical Mechanics

One nice connection of the path-integral formulation of quantum mechanics is to statistical mechanics,[12] as we will now review. This connection comes by recalling the definition of the partition function

$$Z := \mathrm{Tr}\left[e^{-\beta H}\right], \tag{20.122}$$

(partition function)

where $\beta := 1/k_B T$ for temperature T and Boltzmann constant k_B. We can then compare this to the propagator from combining Eqs. (20.1) and (20.2):

$$K(\beta,t;\alpha,t_0) = \langle\beta|e^{-iH(x,p)(t-t_0)/\hbar}|\alpha\rangle. \tag{20.123}$$

In particular, choosing the initial and final states to be the same, and then choosing *imaginary* times $t_0 = 0$ and $t = -i\hbar\beta$, we have

$$K(\alpha,-i\hbar\beta;\alpha,0) = \langle\alpha|e^{-\beta H}|\alpha\rangle. \tag{20.124}$$

This imaginary-time replacement is called a **Wick rotation** of the propagator. Summing over all states $|\alpha\rangle$ gives the trace we require to complete the partition function, so we identify

$$Z = \sum_\alpha K(\alpha,-i\hbar\beta;\alpha,0),$$

(imaginary-time propagator as partition function) (20.125)

or in the position representation,

$$Z = \int dx\, K(x,-i\hbar\beta;x,0) = \int dx\, \langle x,-i\hbar\beta|x,0\rangle.$$

(imaginary-time propagator as partition function) (20.126)

In particular, from the path integral (20.24), we can make the same replacement to obtain

$$Z = \int dx \int Dx\, \exp\left[\frac{i}{\hbar}\int_0^t dt\, L(x,\dot{x})\right]_{t\to -i\hbar\beta}.$$

(partition function as path integral) (20.127)

[12] Richard P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, 1965), Chapter 10.

We can absorb the dx integral into the path integral as follows. Recall that in the integration path measure (20.23), we had path coordinates $x_1 \dots x_{N-1}$ for the integration, with initial and final coordinates x_0 and $x \equiv x_N$. For the purposes of the diffusion integral (20.127), recall that we should identify the endpoints

$$x = x_0 = x_N, \tag{20.128}$$

which is the subject of the last integration dx. Alternately, for reasons we will come to, we can enforce this by keeping the integral over x, calling it an integral over x_0, and then introducing an integral over x_N, so long as we also introduce a $\delta(x_N - x_0)$ in the integrand to enforce $x_N = x_0$. Thus, we can write

$$Z = \int \tilde{D}x\, \delta[x(\hbar\beta) - x_0] \exp\left[\frac{i}{\hbar}\int_0^t dt\, L(x,\dot{x})\right]_{t\to -i\hbar\beta},$$

(partition function as path integral) (20.129)

where we have modified the integration measure to read

$$\tilde{D}x := \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \prod_{j=0}^{N} dx_j,$$

(functional-integral differential for periodic paths) (20.130)

and in the continuum limit, we have identified the endpoint $x_N \equiv x(t) \longrightarrow x(\hbar\beta)$.

To see what all this means, note that the Hamiltonian (20.12) implies the Schrödinger equation

$$i\hbar\partial_t\psi = -\frac{\hbar^2}{2m}\partial_x^2\psi + V(x)\psi, \tag{20.131}$$

which under the same imaginary-time replacement becomes

$$\partial_\beta\psi = \frac{\hbar^2}{2m}\partial_x^2\psi - V(x)\psi, \tag{20.132}$$

(diffusion equation in imaginary time)

which is a diffusion equation in the imaginary time β, with a damping term (corresponding to a space-dependent cooling in a heat equation) due to $V(x)$. Thus, we expect the paths in Eq. (20.127) to correspond to diffusive paths that as a whole satisfy the diffusion equation (20.132). We will develop this notion further in the next section, but this is also the content of the Feynman–Kac formula, as we developed in the form of Eqs. (17.447) and (17.448) in the context of stochastic calculus.

20.3.1 Path Integral as Ensemble Average

Now we will explore in more detail the paths in the diffusive path integral (20.127). Writing out the Lagrangian explicitly in the path integral, we have

$$\begin{aligned} Z &= \int \tilde{D}x\, \delta[x(\hbar\beta) - x_0] \exp\left[\frac{i}{\hbar}\int_0^t dt \left(\frac{1}{2}m\dot{x}^2 - V(x)\right)\right]_{t\to -i\hbar\beta} \\ &= \int \tilde{D}x\, \delta[x(\hbar\beta) - x_0] \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta} \left(\frac{m}{2}(\partial_{\tilde{\beta}}x)^2 + V(x)\right)\right], \end{aligned} \tag{20.133}$$

where $\tilde{\beta} := \hbar\beta$. Now the idea is to split the two terms in the exponential, giving different interpretations to each one.

Suppose that $f(x)$ is a unit-normalized, nonnegative distribution:

$$\int dx\, f(x) = 1. \tag{20.134}$$

Then we may write any integral involving the product of $f(x)$ with another function as an average,

$$\int dx\, f(x)\, g(x) = \langle\langle g(x) \rangle\rangle_{f(x)}, \tag{20.135}$$

where the expectation value is an ensemble average chosen from the "probability" distribution $f(x)$. We can apply this to the path integral (20.133)

$$Z = \int \tilde{D}x\, \delta[x(\hbar\beta) - x_0] \, \exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, (\partial_{\tilde{\beta}} x)^2\right] \, \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right], \tag{20.136}$$

where, after appropriate normalization, the first exponential factor (the kinetic factor) will play the role of the probability measure $f(x)$.

This probability measure will be normalized by introducing a normalization constant η, such that the resulting path integral is normalized. Then the generalization of Eq. (20.134) is

$$\int \tilde{D}'x\, f[x(t)] = \frac{1}{\eta}\int \tilde{D}'x \, \exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, (\partial_{\tilde{\beta}} x)^2\right] = 1, \tag{20.137}$$

where $\tilde{D}'x$ is the same as $\tilde{D}x$, but omits the integration over the endpoint $x(\hbar\beta)$ (or x_N), because this integration is already taken care of by the δ function. In discrete form, this reads

$$\begin{aligned} \eta &= \int \tilde{D}'x \, \exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, (\partial_{\tilde{\beta}} x)^2\right] \\ &= \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \int \prod_{j=0}^{N-1} dx_j \, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_{j+1} - x_j)^2\right], \end{aligned} \tag{20.138}$$

where $\delta\beta = \beta/N$. Note that this has the form of $N-1$ Gaussian convolutions with one final integral over the result. We can proceed by carrying out first the x_0 integral, which is a Gaussian integral of the form

$$\int dx_0 \, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_1 - x_0)^2\right] = \sqrt{\frac{2\pi\hbar^2\delta\beta}{m}}, \tag{20.139}$$

where we have simply recentered the x_0 integration by $x_0 \longrightarrow x_0 - x_1$ and performed the Gaussian integral. Then Eq. (20.138) becomes

$$\eta = \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \sqrt{\frac{2\pi\hbar^2\delta\beta}{m}} \int \prod_{j=1}^{N-1} dx_j \, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_{j+1} - x_j)^2\right]. \tag{20.140}$$

Iterating this procedure $N-1$ more times, we have

$$\eta = \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \left(\frac{2\pi\hbar^2\delta\beta}{m}\right)^{N/2} = 1, \tag{20.141}$$

and we see that we *already* have a normalized integral from keeping the momentum-integration factors (had we dropped these factors, we could easily use this normalization procedure to restore them). Then applying Eq. (20.135) to the path integral (20.136),

$$Z = \int dx_0 \left\langle\!\!\left\langle \delta[x(\hbar\beta) - x_0] \, \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right] \right\rangle\!\!\right\rangle, \tag{20.142}$$

where we changed the remaining x_N integration over to an integration over x_0 via the δ function, and the double brackets here refer to an ensemble average over $x(\tilde{\beta})$ chosen from the probability functional $f[x(t)]$ defined in Eq. (20.137).

Now before proceeding, we can interpret things a bit. The probability functional $f[x(t)]$ that defines the ensemble average is, in its finite representation in Eq. (20.138), a product of Gaussian factors of the form

$$\exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_{j+1}-x_j)^2\right]. \tag{20.143}$$

This factor is a Gaussian probability density for the separation of neighboring points x_{j+1} and x_j, and says that a pair of such points will only be chosen with significant probability if $|x_{j+1}-x_j|$ is small, of the order of the standard deviation $\sqrt{\hbar^2\delta\beta/m}$ or less. Thus, the points x_j define a *continuous* path in the $N\longrightarrow\infty$ limit, of random, Gaussian steps of variance $\hbar^2\delta\beta/m$. This is just a Wiener process (or Brownian motion). However, a standard Wiener path $W(t)$ takes a step of variance dt in a time interval dt, so we can identify $x(\tilde{\beta})$ as a scaled, displaced Wiener process:

$$x(\tilde{\beta}) = x_0 + \sqrt{\frac{\hbar}{m}}\,W(\tilde{\beta}), \tag{20.144}$$

since $dW(\tilde{\beta})$ has an rms average of $\sqrt{\hbar\delta\beta}$. Furthermore, the δ function in the expectation value ties the endpoint $x(\hbar\beta)$ of the path to the initial point x_0. This is just a Brownian bridge tied down at "time" $\hbar\beta$ to its initial position. We have in fact already worked out the conversion of the δ function to a bridge in Eq. (17.561), with result

$$\left\langle\!\!\left\langle \delta[W(T)]\,F[W(t)] \right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi T}}\left\langle\!\!\left\langle F[B_T(t)] \right\rangle\!\!\right\rangle, \tag{20.145}$$

where F is some functional, and $B_T(t)$ is a Brownian bridge that is "tied down," $B_T(t=T)=0$, at time $t=T$. To adapt this to our paths $x(\tilde{\beta})$, we can note that the tie-down time is $\hbar\beta$, but this should really be a factor $\hbar/m$ longer in view of Eq. (20.144). Thus, the path integral (20.142) becomes

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}}\int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,V(x)\right] \right\rangle\!\!\right\rangle_{x(\tilde{\beta})=x_0+\sqrt{\hbar/m}\,B_{\hbar\beta}(\tilde{\beta})},$$

(partition function as ensemble average over Brownian bridges) (20.146)

where now, the ensemble average is over Wiener loops or Brownian bridges that start and end at x_0, where we sum over all possible x_0. The exponential factor involves an integral of the potential V evaluated along the stochastic loop. Note that the overal β-dependent normalization factor can be dropped, as overall factors in partition functions are ignorable. Also, note that since we have defined the Brownian bridges or stochastic loops to be periodic in $\tilde{\beta}$, with period $\hbar\beta$, we may regard the "imaginary time" $\tilde{\beta}$ itself to be periodic (i.e., representing a compact time dimension), with the same period.

20.3.1.1 Alternate Normalization of the Periodic Path

It is constructive to consider alternate ways to compute the normalization of the path integral (20.136). We already covered one method in Section 20.1.5, where we evaluated a determinant as part of a direct matrix-Gaussian integration to integrate a periodic path integral. Another way is to simply normalize against the free-particle partition function, obtained by simply setting $V=0$ in Eq. (20.136):

$$\begin{aligned} Z_0 &= \int \tilde{D}x\,\delta[x(\hbar\beta)-x_0]\,\exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,(\partial_{\tilde{\beta}}x)^2\right] \\ &= \int \tilde{D}'x\,\exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,(\partial_{\tilde{\beta}}x)^2\right]. \end{aligned} \tag{20.147}$$

This second form looks superficially the same as the integral (20.138) that we computed before, but because the x_N integral changed the δ function into a pinned endpoint on the path, this integral also includes the contribution that we obtained from the δ function in Eq. (20.145). Rather than deal with the path integral,

we will go back to the original partition function (20.122), writing out the free-space Hamiltonian and the momentum-space trace:

$$Z_0 = \int dx \, \langle x| \exp\left[-\frac{\beta p^2}{2m}\right] |x\rangle. \tag{20.148}$$

To evaluate this, we insert a momentum-space identity, to reduce the momentum operators in the exponential to eigenvalues:

$$\begin{aligned} Z_0 &= \int dx \int dp \, \langle x|p\rangle\langle p| \exp\left[-\frac{\beta p^2}{2m}\right] |x\rangle \\ &= \frac{1}{2\pi\hbar} \int dx \int dp \, \exp\left[-\frac{\beta p^2}{2m}\right] \\ &= \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx. \end{aligned} \tag{20.149}$$

The remaining position integral is a divergent volume factor, which is telling us that we shouldn't normalize with respect to Z_0, but rather to the free-partition-function *density* $Z_0(x)$, which is just Z_0 without the extra integration factor:

$$Z_0(x) = \sqrt{\frac{m}{2\pi\hbar^2\beta}}. \tag{20.150}$$

This gives the proper normalization of the kinetic part of the partition function (20.136), including the δ function, but without the integral over x_0. Thus, in converting Eq. (20.136) to an ensemble average, we associate a factor of $1/Z_0(x)$ (i.e., $1/\eta$ in our previous notation) with the probability functional $f[x(t)]$, leaving a factor of $Z_0(x)$ in the remaining ensemble average. The result is

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right] \right\rangle\!\!\right\rangle_{x(\hbar\beta)=x_0}, \tag{20.151}$$

(ensemble average via alternate normalization)

which agrees exactly with Eq. (20.146), implementing the normalization due to the change to an ensemble average and the loop-closure δ function all in one go. However, we don't get the clear distinction of the factor here as being due to the pinning of the loop end.

20.3.1.2 YANPP (Yet Another Normalization of the Periodic Path)

Finally, we will compute the normalization yet *another* way, illustrating how to perform the explicit, iterated convolution with the δ function up front (this method is more along the lines of Section 20.1.5, but with direct evaluation of the individual integrals). We need to compute the free-partition-function density, which we can infer from Eqs. (20.147):

$$Z_0(x) = \int \tilde{D}'x \, \delta[x(\hbar\beta) - x_0] \, \exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, (\partial_{\tilde{\beta}} x)^2\right]. \tag{20.152}$$

In discrete form, this expression becomes

$$Z_0(x) = \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \int \prod_{j=0}^{N-1} dx_j \, \delta(x_N - x_0) \, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_{j+1} - x_j)^2\right], \tag{20.153}$$

Note that this is the same as η in Eqs. (20.138), but now with the loop-closure δ-function included. Carrying out the x_0 integral implements the δ function, leaving a somewhat more complicated integrand than before:

$$Z_0(x) = \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2} \int \left\{ \prod_{j=1}^{N-1} dx_j \, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_{j+1} - x_j)^2\right] \right\} \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_1 - x_N)^2\right]. \tag{20.154}$$

We can then carry out the x_1 integral

$$\begin{aligned}\int dx_1\, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_2-x_1)^2\right]\exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_1-x_N)^2\right]&\\ =\int dx_1\, \exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_2-x_1)^2\right]\exp\left[-\frac{m}{2\hbar^2\delta\beta}(x_1-x_N)^2\right]&\\ =\int dx_1\, \exp\left[-\frac{m}{\hbar^2\delta\beta}x_1^2\right]\exp\left[-\frac{m}{4\hbar^2\delta\beta}(x_2-x_N)^2\right]&\\ =\sqrt{\frac{\pi\hbar^2\delta\beta}{m}}\exp\left[-\frac{m}{4\hbar^2\delta\beta}(x_2-x_N)^2\right],&\end{aligned} \tag{20.155}$$

where we complete the square via

$$\begin{aligned}(x_2-x_1)^2+(x_1-x_N)^2&=2x_1^2-2x_1(x_2+x_N)+x_2^2+x_N^2\\ &=2[x_1-(x_2+x_N)/2]^2-(x_2+x_N)^2/2+x_2^2+x_N^2\\ &=2[x_1-(x_2+x_N)/2]^2+(x_2-x_N)^2/2.\end{aligned} \tag{20.156}$$

For the x_2 integral, we then complete a square $(x_3-x_2)^2+(x_2-x_N)^2/2$, so let's guess that the general case will be a square of the form

$$\begin{aligned}(x_{j+1}-x_j)^2+(x_j-x_N)^2/j&=(1+1/j)x_j^2-2x_j(x_{j+1}+x_N/j)+x_{j+1}^2+x_N^2/j\\ &=[(j+1)/j][x_j-(x_{j+1}+x_N/j)j/(j+1)]^2\\ &\qquad-(x_{j+1}+x_N/j)^2j/(j+1)+x_{j+1}^2+x_N^2/j\\ &=[(j+1)/j][x_j-(x_{j+1}+x_N/j)j/(j+1)]^2+(x_{j+1}-x_N)^2/(j+1).\end{aligned} \tag{20.157}$$

Therefore, the jth integral will contribute a factor of $\sqrt{j2\pi\hbar^2\delta\beta/(j+1)m}$, and change the ''outside exponential'' from an argument of $(x_j-x_N)^2/j$ to $(x_{j+1}-x_N)^2/(j+1)$. After the $(N-1)$th integral, all the exponential factors vanish, leaving

$$\begin{aligned}Z_0(x)&=\left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2}\left(\frac{2\pi\hbar^2\delta\beta}{m}\right)^{(N-1)/2}\prod_{j=1}^{N-1}\sqrt{\frac{j}{j+1}}\\ &=\sqrt{\frac{m}{2\pi\hbar^2N\delta\beta}}\\ &=\sqrt{\frac{m}{2\pi\hbar^2\beta}}.\end{aligned} \tag{20.158}$$

This is, of course, the same factor we obtained in Eq. (20.150) in the previous section, but with distinctly more work. In fact, we have just reproduced the Brownian-bridge construction via square-completion of Section 17.7.1, suitably scaled for the present stochastic paths.The normalization of the partition function then proceeds as in the previous section.

20.3.2 Feynman–Kac Formula

The Feynman–Kac formula that we treated before in Section 17.11 is an equivalent formalism for diffusive path integrals, though we derived it before with a quite different approach using Itō calculus. In particular, the form we will consider here is the solution (17.448)

$$f(x,t)=\left\langle\!\!\left\langle f_0[x+W(t)]\,\exp\left(-\int_0^t dt'\,V[x+W(t'),t-t']\right)\right\rangle\!\!\right\rangle \tag{20.159}$$

to the diffusion equation (17.447)

$$\partial_t f = \frac{1}{2}\partial_x^2 f - V(x,t)f. \tag{20.160}$$

To demonstrate the equivalence of this formula with the present formalism, we will need to transform this equation into precisely the same diffusion equation as Eq. (20.132), which we will write as

$$\partial_{\tilde{\beta}}\psi = \frac{1}{2}\partial_{\tilde{x}}^2\psi - \frac{V\left(\tilde{x}\sqrt{\hbar/m}\right)}{\hbar}\psi, \tag{20.161}$$

where $\tilde{\beta} = \hbar\beta$ as usual, and $\tilde{x} = \sqrt{m/\hbar}\,x$. Then the solution of this equation according to Eq. (20.159) is at x_0 [i.e., $f_0(x) = \delta(x - x_0)$]:

$$f(\tilde{x}, \hbar\beta) = \left\langle\!\!\left\langle \delta[\tilde{x} - \tilde{x}_0 + W(\hbar\beta)] \exp\left(-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V\left[\sqrt{\hbar/m}\,\tilde{x}_0 + \sqrt{\hbar/m}\,W(\tilde{\beta})\right]\right)\right\rangle\!\!\right\rangle. \tag{20.162}$$

Note that what we have here is already the diffusive propagator, giving the probability density (amplitude) at $(\tilde{x}, \hbar\beta)$, given that the ensemble started at $(\tilde{x}_0, 0)$. What remains from Eq. (20.125) is to set the final point x equal to the initial point x_0, and then integrate over the initial point to obtain the partition function:

$$Z = \int d\tilde{x}_0\, f(\tilde{x}_0, \hbar\beta) = \int d\tilde{x}_0 \left\langle\!\!\left\langle \delta[W(\hbar\beta)] \exp\left(-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V\left[\sqrt{\hbar/m}\,\tilde{x}_0 + \sqrt{\hbar/m}\,W(\tilde{\beta})\right]\right)\right\rangle\!\!\right\rangle. \tag{20.163}$$

Then applying Eq. (20.145), we have

$$Z = \frac{1}{\sqrt{2\pi\hbar\beta}}\int d\tilde{x}_0 \left\langle\!\!\left\langle \exp\left(-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V\left[\sqrt{\hbar/m}\,\tilde{x}_0 + \sqrt{\hbar/m}\,B_{\hbar\beta}(\tilde{\beta})\right]\right)\right\rangle\!\!\right\rangle, \tag{20.164}$$

and putting $\tilde{x}_0 = \sqrt{m/\hbar}\,x_0$, we find

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}}\int dx_0 \left\langle\!\!\left\langle \exp\left(-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V\left[x_0 + \sqrt{\hbar/m}\,B_{\hbar\beta}(\tilde{\beta})\right]\right)\right\rangle\!\!\right\rangle, \tag{20.165}$$

which is equivalent to the partition-function expression (20.146) that we already derived. Thus, the Feynman–Kac formula reproduces what we derived quite differently from the imaginary-time propagator, and in fact it indicates somewhat more generality in allowing for time dependence of the potential, which is more useful in the Schrödinger propagator rather than the (equilibrium) partition function.

20.3.3 Thermal Density Matrix

To review the utility of the partition function, recall that the canonical partition function is given in terms of the energies E_j of a system as

$$Z := \sum_n e^{-\beta E_n}, \tag{20.166}$$

where the exponential terms in this sum represent the relative probabilities (according to the Boltzmann distribution) of an element of the canonical ensemble to occupy state n. The partition function neatly gives the normalization factor for the Boltzmann probabilities, such that the normalized probability of occupying state n is

$$P(n) = \frac{e^{-\beta E_n}}{Z}, \tag{20.167}$$

and useful quantities such as the thermal energy may be computed as derivatives of Z:

$$-\partial_\beta \log Z = -\frac{\partial_\beta Z}{Z} = \sum_n E_n \frac{e^{-\beta E_n}}{Z} = \sum_n E_n P(n) = \langle E \rangle\,. \tag{20.168}$$

For a quantum system, the definition (20.122) follows by identifying state energies as expectation values $E_n = \langle n|H|n\rangle$ with respect to energy eigenstates $|n\rangle$, such that

$$Z = \sum_n e^{-\beta\langle n|H|n\rangle} = \sum_n \langle n|e^{-\beta H}|n\rangle = \mathrm{Tr}\left[e^{-\beta H}\right]. \tag{20.169}$$

Then note also that in thermal equilibrium, the quantum state, or thermal density operator is closely related, as

$$\rho = \sum_n P(n)|n\rangle\langle n| = \frac{1}{Z}\sum_n e^{-\beta E_n}|n\rangle\langle n| = \frac{1}{Z}\sum_n e^{-\beta H}|n\rangle\langle n|. \tag{20.170}$$

This is an expression of ρ in the energy basis; the basis-independent expression is therefore

$$\rho = \frac{e^{-\beta H}}{Z}. \tag{20.171}$$
(thermal density operator)

In the position representation, the density matrix is then simply

$$\langle x|\rho|x'\rangle = \frac{\langle x|e^{-\beta H}|x'\rangle}{Z}. \tag{20.172}$$
(thermal density operator)

Examining the steps leading up to Eq. (20.125), we see that the exponential matrix element is precisely the Wick-rotated propagator:

$$\rho(x,x')Z = K(x,-i\hbar\beta;x',0),$$
(imaginary-time propagator as thermal density matrix) (20.173)

and thus the propagator gives the thermal state of a system, up to a factor of Z, or simply the unnormalized thermal density matrix. The next step in computing the partition function in terms of the propagator was to carry out the spatial trace, which follows here as the simple normalization of the density operator:

$$Z = Z\mathrm{Tr}\rho = Z\int dx\,\langle x|\rho|x\rangle = \int dx\,\langle x|e^{-\beta H}|x\rangle. \tag{20.174}$$

Thus, all of the above path-integration results apply also to $Z\rho(x,x')$, if we drop the overall spatial integral and interpret the paths as traveling from x' to x in imaginary time from 0 to $\hbar\beta$. Thus, for example, Eq. (20.127) becomes

$$\rho(x,x_0) = \frac{1}{Z}\int Dx\,\exp\left[\frac{i}{\hbar}\int_0^t dt\,L(x,\dot{x})\right]_{t\to -i\hbar\beta},$$
(partition function as path integral) (20.175)

where as in the original propagator, the paths run from x_0 to x. Similarly, the ensemble-averaged expression (20.146) translates to

$$\rho(x,x_0) = \frac{1}{Z}\sqrt{\frac{m}{2\pi\hbar^2\beta}}\,e^{-m(x-x_0)^2/2\hbar^2\beta}\left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,V(x)\right]\right\rangle\!\!\right\rangle_{x(\tilde{\beta})},$$
(partition function as ensemble average over Brownian bridges) (20.176)

where $x(\tilde{\beta})$ travels from x_0 to x as $\tilde{\beta}$ goes from 0 to $\hbar\beta$ with steps of variance $\hbar\delta\tilde{\beta}/m$ in each time step $\delta\tilde{\beta}$.

The extra exponential factor here bears a bit more explanation. Because we want to pin the endpoint of the path at x, and not x_0, we must still carefully handle this pinning. To do this, we can take Eq. (20.175) and introduce an integration over the endpoint x_N as in Eq. (20.129), inserting a δ function to fix the endpoint,

$$\rho(x_{\hbar\beta},x_0) = \frac{1}{Z}\int dx_N\,Dx\,\delta[x_N - x_{\hbar\beta}]\,\exp\left[-\frac{m}{2\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,(\partial_{\tilde{\beta}}x)^2\right]\,\exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\,V(x)\right], \tag{20.177}$$

in mixed continuous-discrete notation, and where we have changed the notation for the endpoint from x to $x_{\hbar\beta}$ to avoid confusion with the paths $x(\tilde{\beta})$ in the path integral. There are N integrals implied by $dx_N\,Dx$, and the normalization factor in Dx is again the appropriate normalization factor for the kinetic term to act as the probability density, so that the analogue of Eq. (20.142) becomes

$$\rho(x_{\hbar\beta}, x_0) = \frac{1}{Z}\left\langle\!\!\left\langle \delta[x(\hbar\beta) - x_{\hbar\beta}] \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right]\right\rangle\!\!\right\rangle. \tag{20.178}$$

Now we can evaluate the contribution of the δ function here by modifying Eq. (20.145) to handle the offset δ function here. Examining the derivation of this relation, we can see that the factor of $1/\sqrt{2\pi t}$ there came from the probability density for $W(t)$, evaluated at $W(t) = 0$. If we instead evaluate this at some other final point W_t, then we have [cf. Eq. (17.592)]

$$\left\langle\!\!\left\langle \delta[W(t) - W_t]\, F[W(t)]\right\rangle\!\!\right\rangle = \frac{e^{-{W_t}^2/2t}}{\sqrt{2\pi t}}\left\langle\!\!\left\langle F[B_t(t)]\right\rangle\!\!\right\rangle, \tag{20.179}$$

which is what we require here. As in Eq. (20.146), to adapt this to the (unpinned) paths $x(t)$, we set $t \longrightarrow \hbar^2\beta/m$, and then count $W(t)$ as the (scaled) separation of $x_{\hbar\beta}$ from the initial point x_0, which yields Eq. (20.176).

20.3.3.1 Example: Free-Particle Thermal State

The simplest application of the formalism here is to work out the thermal density operator for a free particle. Setting $V = 0$ in Eq. (20.176), the ensemble average becomes trivial, and thus

$$\rho(x, x') = \frac{1}{Z}\sqrt{\frac{m}{2\pi\hbar^2\beta}}\, e^{-m(x-x')^2/2\hbar^2\beta}. \tag{20.180}$$
(free particle thermal state)

Note that the partition function here involves a divergent volume integral over a constant, so we won't bother to write it down. Note also that the density operator here agrees with the free-space propagator (20.48) under the replacement $(t - t_0) \longrightarrow -i\hbar\beta$ (if we take away the factor of Z).

20.3.3.2 Example: Partition Function of an Infinite Square Well

Consider an infinite square well, open from $x = 0$ to L,

$$V(x) = \lim_{V_0\to\infty} V_0\left[\Theta(x - L) + \Theta(-x)\right], \tag{20.181}$$

and let's use world lines to compute the partition function (20.146). In that expression, we have an ensemble average over a particular statistic: the exponentiated line integral of $-V(x)$ over each path. If the path touches either wall of the potential, we have $\exp(-\infty) = 0$, and if the path does *not* touch either wall, then we have $\exp(-0) = 1$. Thus, the ensemble average is just counting the probability that each path does *not* touch either wall of the well. In the language of stochastic processes, this can be phrased in terms of the *escape probability* of a Brownian bridge from the interval $(-a, L - a)$ [see Eq. (17.415)]:

$$P_{\text{escape}} = 1 + \sum_{j=-\infty}^{\infty}\left[e^{-2(a+jL)^2} - e^{-2(jL)^2}\right]. \tag{20.182}$$

Then the partition function is

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}}\int dx_0\,\left\{1 - P_{\text{escape}}[x(\tilde{\beta})]\right\}. \tag{20.183}$$

To adapt the escape probability here, we should let $a \longrightarrow x_0$, but since we are dealing with a bridge that is effectively pinned at time $\hbar^2\beta/m$, we should compensate by setting $a \longrightarrow x_0/\hbar\sqrt{\beta/m}$ and $L \longrightarrow L/\hbar\sqrt{\beta/m}$:

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int_0^L dx_0 \sum_{j=-\infty}^{\infty} \left[e^{-2m(jL)^2/\hbar^2\beta} - e^{-2m(x_0+jL)^2/\hbar^2\beta} \right]. \tag{20.184}$$

The integral over the second term vanishes, because it is of the form

$$\int_0^1 dx\, e^{-a(x+j)^2} = \sqrt{\frac{\pi}{4a}} \left\{ \operatorname{erf}\left[(j+1)\sqrt{a}\right] - \operatorname{erf}\left[j\sqrt{a}\right] \right\}, \tag{20.185}$$

which vanishes under an infinite sum over j. (Note that the individual terms technically give rise to divergent sums, which vanish independently if defined in the sense of a Cauchy principal value.) Then the partition function is

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}}\, L \sum_{j=-\infty}^{\infty} e^{-2m(jL)^2/\hbar^2\beta}, \tag{20.186}$$

which is technically correct, but we will take this a bit further.

Starting with the Poisson sum rule in the form[13]

$$\sum_{n=-\infty}^{\infty} \delta(t-n) = \sum_{n=-\infty}^{\infty} e^{i2\pi nt} = \sum_{n=-\infty}^{\infty} \cos(2\pi nt), \tag{20.187}$$

(Poisson sum rule)

we multiply by $\exp(-at^2)$ and integrate over t,

$$\sum_{n=-\infty}^{\infty} e^{-an^2} = \sum_{n=-\infty}^{\infty} \int dt\, e^{-at^2}\, e^{i2\pi nt}, \tag{20.188}$$

so that

$$\sum_{n=-\infty}^{\infty} e^{-an^2} = \sqrt{\frac{\pi}{a}} \sum_{n=-\infty}^{\infty} e^{-n^2\pi^2/a}. \tag{20.189}$$

(Poisson sum rule for Gaussians)

Using this formula, Eq. (20.186) becomes

$$Z = \frac{1}{2} \sum_{j=-\infty}^{\infty} e^{-\pi^2\hbar^2\beta j^2/2mL^2}, \tag{20.190}$$

or finally after simplifying the sum,

$$Z = \sum_{n=1}^{\infty} e^{-\beta n^2\pi^2\hbar^2/2mL^2} = \sum_{n=1}^{\infty} e^{-\beta E_n}, \tag{20.191}$$

(partition function of infinite square well)

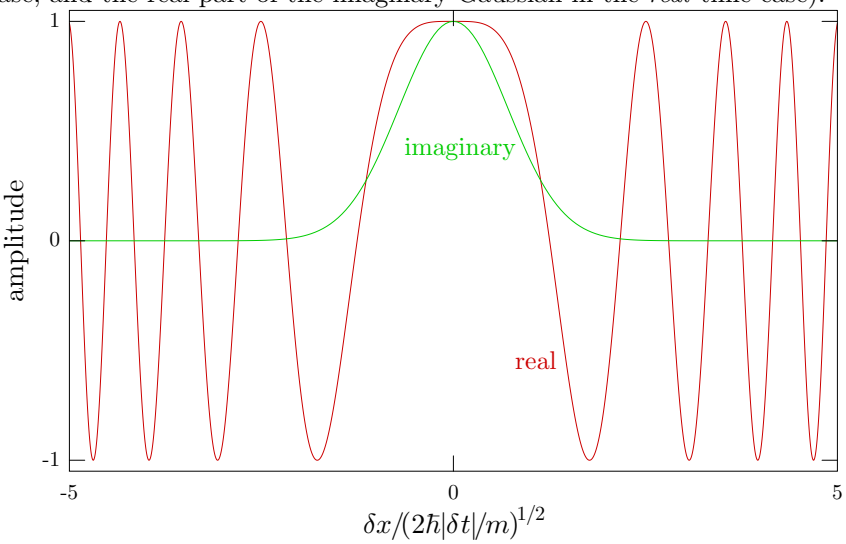
which is what we expect for a partition function for a set of states with energies $E_n = n^2\pi^2\hbar^2/2mL^2$. This is the just another way to derive the eigenenergies for the infinite square well! Note that we just needed the escape probability for a Brownian bridge here; we can also derive the density matrix with this technique, using the more general formula (17.415) for a nonperiodic Brownian bridge.

[13] Daniel A. Steck, *Classical and Modern Optics* (2006), available online at `http://steck.us/teaching`.

20.3.4 Propagator Paths

In developing path integrals in statistical mechanics, due to the imaginary-time identification $t = -i\hbar\beta$ leading to Eq. (20.125), each step in the path integral has a factor of the form $\exp[-m(\delta x)^2/2\hbar|\delta t|]$, where $|\delta t| = \hbar\beta$. In Section 20.3.1, this allowed us to develop the path integral as a Monte-Carlo average over Wiener paths, where adjacent steps in the path were close, with $\delta x \sim \sqrt{\delta t}$.

However, the original quantum propagator has oscillatory factors of the form $\exp[im(\delta x)^2/2\hbar(\delta t)]$ in Eq. (20.17), and it is less obvious that a similar interpretation in terms of diffusive paths is possible there. The two amplitude distributions are contrasted in the plot below (showing the real Gaussian in the *imaginary*-time case, and the real part of the imaginary Gaussian in the *real*-time case).



Even though the factor $\exp[im(\delta x)^2/2\hbar(\delta t)]$ oscillates without decaying, note that it *is* possible to integrate it in the sense of normalizing it, because the tail oscillations increase in frequency [in the same sense that the Fresnel integrals $C(x)$ and $S(x)$ converge to well-defined values as $x \longrightarrow \infty$]. Nevertheless, the undamped, oscillatory character of the step amplitudes cause problems for rigorous handling of the propagator path integral, and there has been a great deal of work in establishing rigorous foundations for the complex path integral.[14]

In a heuristic approach, we can still say that for a Gaussian distribution,

$$f(x) = \frac{1}{\sqrt{\pi}\alpha} e^{-x^2/\alpha^2} \tag{20.192}$$

the convolution of multiple Gaussians for successive steps in the path integral

$$(f * f)(x) = \frac{1}{\sqrt{2\pi}\alpha} e^{-x^2/2\alpha^2}. \tag{20.193}$$

still works whether the phase of α is 0 or $-\pi/4$, so the complex Gaussian still works out as a stable "step" distribution. In the imaginary-time (real Gaussian) case, the Wiener-path interpretation came from the decay of the Gaussian tails for large steps. In the real-time (imaginary Gaussian) case, the "unlikely" paths from the diffusion picture still have large amplitude, but being in the oscillatory tails, they have a rapidly oscillating phase, and tend not to contribute via cancellation with other unlikely paths. This is an argument similar to the one used in Section 20.2.2 for the semiclassical limit of the path integral. In this heuristic way we can still think of the important paths for the propagator as being the Wiener paths.

A reasonably safe and less heuristic approach to the propagator path integral is to realize that from the treatment of path integrals for statistical mechanics in Section 20.3, propagators in imaginary time, $K(x_2, -i\tau_2; x_1, -i\tau_1)$, are mathematically well-defined objects with an interpretation in Wiener paths. The

[14] Sergio Albeverio and Sonia Mazzucchi, *Scholarpedia* **6**, 8832 (2011) (doi: 10.4249/scholarpedia.8832).

quantum-mechanical propagators in real time, $K(x_2, t_2; x_1, t_1)$, can then be defined via analytic continuation of the imaginary-time function. (Similar arguments are possible for analytic continuation of imaginary mass or $\hbar$.) In this sense, even the quantum propagator is calculated in terms of Wiener paths. This gives another way to think about how the important quantum paths are diffusive in nature despite the oscillatory amplitude factors.

20.4 Ordering Issues

Going back to the derivation of the path-integral propagator (20.24), we made the assumption of a particle Hamiltonian (20.12), where the Gaussian integration changed the momentum kinetic term into the velocity kinetic term—but with the opposite sign—in the exponential. The question is, what happens if we have a more general Hamiltonian $H(x, p)$ that *doesn't* separate out into position and momentum terms so easily? That is, what if there is an ordering ambiguity, as in pxp, where different orderings are connected by correction terms given in terms of commutators?

20.4.1 Model Problem

As an example, let us work out path integrals for a one-degree-of-freedom Hamiltonian of the form

$$H(x,p) = p\frac{1}{2mg(x)}p + V(x), \tag{20.194}$$
(variable-mass Hamiltonian)

for some ''space-dependent mass'' or ''metric'' function $g(x)$, and consider the case of various orderings of the kinetic-energy term. The various orderings are equivalent up to commutator terms, which we will indicate. Then, we will continue with a particular ordering (Weyl ordering) and work out the path integral in more detail. We will keep the particular symmetric ordering here on the kinetic-energy term in mind as a reference case. Actually, what turns out to be a more natural ordering is what we will refer to as ''Laplace–Beltrami'' or ''covariant'' ordering,

$$H_{\triangle}(x,p) := \frac{1}{2m}\left(\frac{1}{g^{1/4}}p\frac{1}{\sqrt{g}}p\frac{1}{g^{1/4}}\right) + V(x), \tag{20.195}$$
(Laplace–Beltrami Hamiltonian)

as we will see below [Eq. (20.233)]; despite the more complicated form, the path-integral representations in this ordering are more natural and simple, so this will be our more important reference case.

20.4.1.1 Basic Structure

The reason we need to be somewhat specific about the form of the Hamiltonian is that it will affect some of the basic rules in quantum mechanics (normalization integrals, form of the plane wave, etc.). Before continuing with the path integral, then, we will take a moment to work out the consequences of having a Hamiltonian of the form (20.194), or one of its reorderings.

Starting with the classical Lagrangian

$$L(x,\dot{x}) = \frac{m}{2}g(x)\,\dot{x}^2 - V(x), \tag{20.196}$$
(classical Lagrangian)

we can define the canonical momentum

$$p := \frac{\partial L}{\partial \dot{x}} = mg(x)\,\dot{x} \tag{20.197}$$

and the classical Hamiltonian

$$H(x,p) = p\dot{x} - L = \frac{p^2}{2mg(x)} + V(x). \tag{20.198}$$

In quantizing this Hamiltonian, there is an ambiguity in the ordering of the kinetic-energy term. We can regard the ordering in Eq. (20.194) as an arbitrary choice for the sake of an example, with other orderings obtainable by commuting operators.

The classical Poisson bracket then has the form

$$[f,g]_{\rm P} := \frac{\partial f}{\partial x}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial x}, \tag{20.199}$$

provided (x,p) are canonical coordinates. Thus, the bracket of the canonical coordinates is still $[x,p]_{\rm P}=1$, which means that in quantum mechanics we still have the commutator

$$[x,p] = i\hbar. \tag{20.200}$$

This will define the form of the quantum-mechanical momentum operator.

20.4.1.2 Coordinate Transformation and Curvature

We have to be careful about interpreting the classical system in quantum mechanics, because of the difference between orderings of x and p that does not happen classically, as we noted above.[15] Note in particular that the form of the Lagrangian and Hamiltonian that give rise to the ordering problem are analogous to a curved structure in the phase and state space.[16] The Lagrangian in particular implies the line element

$$ds^2 = g(x)\,dx^2, \tag{20.201}$$

so that in coordinate-free form, the kinetic-energy function $g(x)\dot{x}^2 = (ds/dt)^2 = (ds/dx)^2(dx/dt)^2$. This is equivalent to a flat-space line element

$$ds^2 = dq^2, \tag{20.202}$$

given the nonlinear coordinate transformation

$$\frac{dq}{dx} = \sqrt{g(x)}. \tag{20.203}$$

In these coordinates, the Lagrangian (20.196) becomes

$$L_q(q,\dot{q}) = \frac{m}{2}g(x)\,\dot{q}^2\left(\frac{dx}{dq}\right)^2 - V(q) = \frac{m}{2}\dot{q}^2 - V(q), \tag{20.204}$$

where $V(q)=V[q(x)]$, which leads to the ordinary Hamiltonian

$$H_q(p_q,q) = \frac{p^2}{2m} + V(q) \tag{20.205}$$

with canonical momentum

$$p_q = \frac{\partial L}{\partial \dot{q}} = m\dot{q}. \tag{20.206}$$

In *these* coordinates, we know how to quantize the Hamiltonian: the canonical variables become operators with commutator $[q,p_q]=i\hbar$. The Schrödinger equation in flat-space coordinates is then

$$i\hbar\partial_t\psi = \left(-\frac{\hbar^2}{2m}\partial_q^2 + V(q)\right)\psi. \tag{20.207}$$

[15] Our discussion here is a simplified version of the more general multidimensional case. This point of view of using point-coordinate transformations to understand quantum systems with ordering problems and quantum mechanics on curved manifolds was pioneered by Bryce Seligman DeWitt, ''Point Transformations in Quantum Mechanics,'' *Physical Review* **85**, 653 (1952) (doi: 10.1103/PhysRev.85.653); Bryce S. DeWitt, ''Dynamical Theory in Curved Spaces. I. A Review of the Classical and Quantum Action Principles,'' *Reviews of Modern Physics* **29**, 377 (1957) (doi: 10.1103/RevModPhys.29.377). See also Hagen Kleinert, *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets*, 5th ed. (World Scientific, 2009), Section 1.13; Christian Grosche, ''An Introduction into the Feynman Path Integral,'' arXiv.org preprint (arXiv: hep-th/9302097v1).

[16] at least in 2 or more dimensions; in one dimension the manifold cannot be regarded as ''intrinsically'' curved. The other distinction is that in a truly curved manifold, the transformation that we will do can only be performed *locally* not globally; if there is a global transformation to flat space, then the space is just a flat space in complicated coordinates.

The one-dimensional Laplacian transforms as

$$\partial_q^2 = \partial_q \partial_q = \frac{dx}{dq}\partial_x \left(\frac{dx}{dq}\partial_x\right) = \frac{1}{\sqrt{g}}\partial_x \left(\frac{1}{\sqrt{g}}\partial_x\right), \tag{20.208}$$

and so the Schrödinger equation in the original variables, corresponding to the Lagrangian (20.196), is

$$i\hbar\partial_t\psi = \left(-\frac{\hbar^2}{2m}\triangle + V(x)\right)\psi, \tag{20.209}$$
(Schrödinger equation)

where the curved, one-dimensional Laplacian

$$\triangle := \frac{1}{\sqrt{g}}\partial_x \left(\frac{1}{\sqrt{g}}\partial_x\right) \tag{20.210}$$
(Laplace–Beltrami operator, 1D)

is the one-dimensional form of the **Laplace–Beltrami operator**. Note that this implies a certain natural ordering of the kinetic-energy term in the Hamiltonian, to which we will return below.

The coordinate transformation also induces some other changes in the structure of the Hilbert space. In the flat coordinates, the inner product has the usual form

$$\langle\psi_1|\psi_2\rangle = \int dq\, \psi_1^*(q)\, \psi_2(q), \tag{20.211}$$

which in the context of the curved Hamiltonian transforms to

$$\langle\psi_1|\psi_2\rangle = \int dx\, \sqrt{g(x)}\, \psi_1^*(x)\, \psi_2(x), \tag{20.212}$$
(inner product)

where $\psi(x) := \langle x|\psi\rangle$. To be consistent with this, the identity in the position representation must be

$$\int dx\, \sqrt{g(x)}\, |x\rangle\langle x| = 1. \tag{20.213}$$
(identity in position representation)

By inserting this identity in the inner product

$$\psi(x) = \langle x|\psi\rangle = \int dx'\, \sqrt{g(x')}\, \langle x|x'\rangle\langle x'|\psi\rangle = \int dx'\, \sqrt{g(x')}\, \langle x|x'\rangle\, \psi(x'), \tag{20.214}$$

and comparing to

$$\psi(x) = \int dx'\, \delta(x - x')\, \psi(x'), \tag{20.215}$$

we can conclude

$$\langle x|x'\rangle = \frac{1}{\sqrt{g(x)}}\, \delta(x - x') \tag{20.216}$$
(inner product of position states)

for the orthogonality relation of the position states. In terms of the propagator, which we can rederive by using Eq. (20.212) to insert the identity in

$$\psi(x,t) = \langle x,t|\psi\rangle = \int dx_0\, \sqrt{g(x_0)}\, \langle x,t|x_0,t_0\rangle\langle x_0,t_0|\psi\rangle = \int dx_0\, \sqrt{g(x_0)}\, \langle x,t|x_0,t_0\rangle\psi(x_0,t_0), \tag{20.217}$$

so that we have generalized Eq. (20.7) to

$$\psi(x,t) = \int dx_0\, \sqrt{g(x_0)}\, K(x,t;x_0,t_0)\, \psi(x_0,t_0), \tag{20.218}$$
(evolution via propagator)

where the propagator is

$$K(x,t;x_0,t_0) := \langle x,t|x_0,t_0\rangle = \langle x|e^{-iH(t-t_0)/\hbar}|x_0\rangle \tag{20.219}$$
(propagator)

as it was before in the case where g is independent of x. By extension, in our application of the propagator to statistical mechanics, by transforming Eq. (20.126) we can write

$$Z = \int dx\,\sqrt{g(x)}\,K(x,-i\hbar\beta;x,0) = \int dx_0 \int dx\,\sqrt{g(x)}\,\delta(x-x_0)\,K(x,-i\hbar\beta;x_0,0),$$
(partition function) (20.220)

due to the change in integration measure.

Turning now to momentum, the momentum operator is defined in part by the commutator rule (20.200) which we note is satisfied for the usual momentum operator, modified by inserting factors of g,

$$p = \frac{\hbar}{i} g^{-\alpha}(x)\partial_x g^{\alpha}(x), \tag{20.221}$$

as we can quickly check:

$$xp = \frac{\hbar}{i} g^{-\alpha}(x) x \partial_x g^{\alpha}(x) = \frac{\hbar}{i} g^{-\alpha}(x)(\partial_x x - 1) g^{\alpha}(x) = px + i\hbar. \tag{20.222}$$

However, the momentum operator must be Hermitian with respect to the inner product (20.212),

$$\langle\psi_1|p|\psi_2\rangle = \langle\psi_1|p^\dagger|\psi_2\rangle, \tag{20.223}$$

which as an inner product reads

$$\langle\psi_1, p\psi_2\rangle = \langle p\psi_1, \psi_2\rangle\,, \tag{20.224}$$

and in explicit integral form, we have

$$\int dx\, g^{1/2-\alpha}\,\psi_1^*(x)\left[\frac{\hbar}{i}\partial_x g^{\alpha}\psi_2(x)\right] = \int dx\,\left[\frac{\hbar}{i}\partial_x g^{\alpha}\psi_1^*(x)\right] g^{1/2-\alpha}\psi_2(x). \tag{20.225}$$

This is satisfied only when $\alpha = 1/2 - \alpha$, or $\alpha = 1/4$,

$$\begin{aligned} p &= \frac{\hbar}{i}\frac{1}{g^{1/4}(x)}\partial_x g^{1/4}(x) \\ &= \frac{\hbar}{i}\partial_x - \frac{\hbar g'(x)}{4g(x)} \\ &= \frac{\hbar}{i}\partial_x - \frac{\hbar}{4}[\partial_x \log g(x)], \end{aligned} \tag{20.226}$$
(momentum operator)

and so we have the usual momentum operator with a ''curvature correction'' given in terms of a derivative of $g(x)$. From this form of the momentum operator, we can infer the momentum eigenstates in the position representation

$$\langle x|p\rangle = \frac{e^{ipx/\hbar}}{\sqrt{2\pi\hbar}\,g^{1/4}}, \tag{20.227}$$
(momentum eigenstate)

with a factor of $g^{-1/4}$ to ''adapt'' to the corresponding factors in the momentum operator. Then to infer the momentum-representation identity operator, we can consider

$$\langle x|p\rangle\langle p|x'\rangle = \frac{e^{ip(x-x')/\hbar}}{\sqrt{2\pi\hbar}\,g^{1/4}(x)g^{1/4}(x')}, \tag{20.228}$$

whence integration over p yields

$$\int dp\,\langle x|p\rangle\langle p|x'\rangle = \frac{\delta(x-x')}{g^{1/4}(x)g^{1/4}(x')} = \frac{\delta(x-x')}{\sqrt{g(x)}}. \tag{20.229}$$

comparing to Eq. (20.216) then gives

$$\int dp\,|p\rangle\langle p| = 1 \tag{20.230}$$
(momentum completeness)

for completeness in momentum states, and similarly we have

$$\langle p|p'\rangle = \delta(p-p') \tag{20.231}$$
(momentum orthonormality)

for the momentum orthnormality relation.

20.4.1.3 "Natural" Ordering

Recall again that the classical Lagrangian (20.196) led to the Laplace–Beltrami operator (20.210) as the natural generalization of the Laplacian, in the sense of connecting to the usual Laplacian under a coordinate change into the standard form of the Lagrangian. Writing this operator in terms of the corresponding momentum operator (20.226), we find

$$-\hbar^2\triangle = \frac{1}{g^{1/4}}p\frac{1}{\sqrt{g}}p\frac{1}{g^{1/4}}. \tag{20.232}$$
(Laplace–Beltrami operator, 1D)

Thus, the Hamiltonian induced by the Laplace–Beltrami operator is

$$H_\triangle(x,p) := \frac{1}{2m}\left(\frac{1}{g^{1/4}}p\frac{1}{\sqrt{g}}p\frac{1}{g^{1/4}}\right) + V(x), \tag{20.233}$$
(Laplace–Beltrami Hamiltonian)

which in turn yields the Schrödinger equation (20.209). The kinetic energy differs from that of our original (product-ordered) Hamiltonian (20.194) by ordering terms, as we can see by employing the commutation relation $[f(x),p] = i\hbar f'(x)$ multiple times:

$$\begin{aligned}
\frac{1}{g^{1/4}}p\frac{1}{\sqrt{g}}p\frac{1}{g^{1/4}} &= \frac{1}{g^{1/4}}p\frac{1}{g^{3/4}}p + i\hbar\frac{1}{g^{1/4}}p\frac{g'}{4g^{7/4}} \\
&= p\frac{1}{g}p - i\hbar\frac{g'}{4g^2}p + i\hbar\frac{1}{g^{1/4}}p\frac{g'}{4g^{7/4}} \\
&= p\frac{1}{g}p - i\hbar\frac{g'}{4g^2}p + i\hbar p\frac{g'}{4g^2} - (i\hbar)^2\frac{g'}{4g^{5/4}}\frac{g'}{4g^{7/4}} \\
&= p\frac{1}{g}p + \frac{i\hbar}{4}\left[p,\frac{g'}{g^2}\right] + \frac{\hbar^2 g'^2}{16g^3} \\
&= p\frac{1}{g}p + \frac{\hbar^2}{4g^3}\left(gg'' - 2g'^2\right) + \frac{\hbar^2 g'^2}{16g^3},
\end{aligned} \tag{20.234}$$

so that

$$H(x,p) = H_\triangle(x,p) + V_\triangle(x,p), \tag{20.235}$$
(ordering relation)

where

$$V_\triangle(x,p) := -\frac{\hbar^2(gg'' - 2g'^2)}{8mg^3} - \frac{\hbar^2 g'^2}{32mg^3} \tag{20.236}$$
(effective ordering potential)

is an effective potential due to reordering our original Hamiltonian into the Laplace–Beltrami ordering. *Any* ordering is valid as a problem, given that it is imposed by the physics. The Laplace–Beltrami Hamiltonian

is ''special'' in that it is the correct choice for the Schrödinger equation corresponding to the classical Lagrangian, and thus generated the modified structure of the inner product, momentum operators, and so on. However, any other ordering can be regarded as valid in the sense of being the same as $H_\triangle$ plus $\hbar$-dependent ''quantum-correction'' potentials.

20.4.2 Point Transformations of the Path Integral

Now we would like to understand the path integral for a quantum system corresponding to the Lagrangian (20.196), and thus the Hamiltonian $H_\triangle(x,p)$ in Eq. (20.195) or (20.233). At first we will do this beginning from the ''flat-space'' propagator for the Lagrangian

$$L_q(q,\dot{q}) = \frac{m}{2}\dot{q}^2 - V(q), \tag{20.237}$$

and thus the action functional

$$S[q(t)] := \int_{t_0}^{t} dt\, L_q(q,\dot{q}). \tag{20.238}$$

We have already worked out this problem; to review, in terms of the action functional, the propagator in continuous-time notation is [Eq. (20.26)]

$$K(q,t;q_0,t_0) = \int Dq\, \exp\left[\frac{i}{\hbar}S[q(t)]\right], \tag{20.239}$$

or more precisely, in discrete form,

$$K(q,t;q_0,t_0) = \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dq_j\right)\left(\prod_{j=0}^{N-1} e^{i\delta S[q_j]/\hbar}\right), \tag{20.240}$$

where the short-time action is

$$\begin{aligned} \delta S[q_j] &:= \int_{t_j}^{t_{j+1}} dt\, L_q(q,\dot{q}) \\ &\approx L_q(q_j,\dot{q}_j)\,\delta t \\ &= \frac{m}{2\delta t}\delta q_j^2 - V(q_j)\,\delta t \\ &= \frac{m}{2\delta t}\left(q_{j+1}-q_j\right)^2 - V(q_j)\,\delta t. \end{aligned} \tag{20.241}$$

Now we wish to consider the coordinate transformation

$$\frac{dq}{dx} = \sqrt{g(x)} \tag{20.242}$$

that maps us to the quantum system defined by $H_\triangle(x,p)$, and use this transformation to work out the path integral. In terms of the new coordinate x, we can write the short-time action as

$$\delta S[q_j] = \frac{m}{2\delta t}\left[q(x_{j+1}) - q(x_j)\right]^2 - V[q(x_j)]\,\delta t, \tag{20.243}$$

if we regard the old coordinate as the transformation function $q = q(x)$. The idea is that to simplify the transformed kinetic energy in the action, we will have to write it in terms of the coordinate x at some time in the range $[t_j, t_{j+1}]$, expanding the path integral about these points. Here we will carry out this expansion about both t_j and $t_{j+1/2}$, which we will refer to as the **pre-point** and **midpoint** choices.[17] In terms of stochastic calculus, these choices correspond to Itō and Stratonovich calculi, respectively, and we will also see how the different choices and calculi here correspond to different possible orderings in the Hamiltonian.

[17] We are working out a simplified version of the calculation presented by H. O. Girotti and T. J. M. Simões, ''A Generalized Treatment of Point Canonical Transformations in the Path Integral,'' *Il Nuovo Cimento* **74**, 59 (1983) (doi: 10.1007/BF02721685), for the transformation in multiple dimensions about an arbitrary time point.

20.4.2.1 Midpoint (Stratonovich) Expansion of the Path Integral

We will start by expanding the short-time action (20.243) around $\bar{x}_j$ to order δt, regarding δx_j as $O(\sqrt{\delta t})$. Starting with the expansions

$$\begin{aligned} q(x_j) &= q(\bar{x}_j) - \frac{q'(\bar{x}_j)}{2}\delta x_j + \frac{q''(\bar{x}_j)}{8}\delta x_j^2 - \frac{q'''(\bar{x}_j)}{48}\delta x_j^3 \\ q(x_{j+1}) &= q(\bar{x}_j) + \frac{q'(\bar{x}_j)}{2}\delta x_j + \frac{q''(\bar{x}_j)}{8}\delta x_j^2 + \frac{q'''(\bar{x}_j)}{48}\delta x_j^3, \end{aligned} \tag{20.244}$$

the expansion for the short-time action becomes

$$\begin{aligned} \delta S[q_j] &= \frac{m}{2\delta t}\left[q(x_{j+1}) - q(x_j)\right]^2 - V[q(x_j)]\,\delta t \\ &= \frac{m}{2\delta t}\left[q'(\bar{x}_j)\,\delta x_j + \frac{q'''(\bar{x}_j)}{24}\delta x_j^3\right]^2 - V[q(x_j)]\,\delta t \\ &= \frac{m}{2\delta t}q'^2(\bar{x}_j)\,\delta x_j^2 + \frac{mq'(\bar{x}_j)\,q'''(\bar{x}_j)}{24\delta t}\delta x_j^4 - V[q(x_j)]\,\delta t \\ &= \frac{m}{2\delta t}g(\bar{x}_j)\,\delta x_j^2 + \frac{m(2gg'' - g'^2)}{96g\,\delta t}\delta x_j^4 - V[q(x_j)]\,\delta t, \end{aligned} \tag{20.245}$$

where we have used the derivatives

$$q' = \sqrt{g}, \qquad q'' = \frac{g'}{2\sqrt{g}}, \qquad q''' = \frac{g''}{2\sqrt{g}} - \frac{g'^2}{4g^{3/2}} = \frac{2gg'' - g'^2}{4g^{3/2}}. \tag{20.246}$$

Then we can also transform and then expand the integration measure about the midpoints, using

$$\begin{aligned} q'(x_j) &= q'(\bar{x}_j) - \frac{q''(\bar{x}_j)}{2}\,\delta x_j + \frac{q'''(\bar{x}_j)}{8}\,\delta x_j^2 \\ &= q'(\bar{x}_j)\left[1 - \frac{q''(\bar{x}_j)}{2q'(\bar{x}_j)}\,\delta x_j + \frac{q'''(\bar{x}_j)}{8q'(\bar{x}_j)}\,\delta x_j^2\right] \\ &= \sqrt{g(\bar{x}_j)}\left[1 - \frac{g'(\bar{x}_j)}{4g(\bar{x}_j)}\,\delta x_j + \frac{2gg'' - g'^2}{32g^2}\,\delta x_j^2\right], \end{aligned} \tag{20.247}$$

so that the measure becomes

$$\begin{aligned} \prod_{j=1}^{N-1} dq_j &= \prod_{j=1}^{N-1} dx_j\, q'(x_j) \\ &= \frac{1}{q'(x_0)}\left(\prod_{j=1}^{N-1} dx_j\right)\prod_{j=0}^{N-1} q'(x_j) \\ &= \frac{1}{\sqrt{g(x_0)}}\left(\prod_{j=1}^{N-1} dx_j\right)\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\left[1 - \frac{g'(\bar{x}_j)}{4g(\bar{x}_j)}\,\delta x_j + \frac{2gg'' - g'^2}{32g^2}\,\delta x_j^2\right]. \end{aligned} \tag{20.248}$$

Then collecting the various terms from Eqs. (20.245) and (20.248), and inserting them into the propagator (20.240), we find

$$\begin{aligned} K(x,t;x_0,t_0) = \frac{1}{\sqrt{g(x_0)}}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1} dx_j\right) \\ \times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(\bar{x}_j)}{2\delta t}\delta x_j^2 - V[q(\bar{x}_j)]\,\delta t\right)\right]\right. \\ \left.\times\left[1 - \frac{g'(\bar{x}_j)}{4g(\bar{x}_j)}\,\delta x_j + \left(\frac{2gg'' - g'^2}{32g^2}\right)\delta x_j^2 + \frac{im(2gg'' - g'^2)}{96\hbar g\,\delta t}\delta x_j^4\right]\right\}, \end{aligned} \tag{20.249}$$

after expanding out everything in the exponential except for the explicit kinetic and potential energies, so that the other expansion terms appear as moment-type components in quasi-Gaussian integrals. We can deal with the "mean" term of the form δx_j by converting it into a prefactor as follows:

$$\begin{aligned}
\prod_{j=0}^{N-1}\left[1-\frac{g'(\bar{x}_j)}{4g(\bar{x}_j)}\delta x_j+\frac{g'^2}{32g^2}\delta x_j^{\,2}\right] &= \prod_{j=0}^{N-1}\left[1-\frac{g'(x_j)}{4g(x_j)}\delta x_j+\frac{g'^2}{32g^2}\delta x_j^{\,2}-\frac{gg''-g'^2}{8g^2}\delta x_j^{\,2}\right] \\
&= \prod_{j=0}^{N-1}\exp\left[-\frac{g'(x_j)}{4g(x_j)}\delta x_j-\frac{gg''-g'^2}{8g^2}\delta x_j^{\,2}\right] \\
&= \prod_{j=0}^{N-1}\exp\left[-\delta\log g^{1/4}(x_j)\right] \\
&= \prod_{j=0}^{N-1}\left[\frac{g(x_j)}{g(x_{j+1})}\right]^{1/4} \\
&= \left[\frac{g(x_0)}{g(x)}\right]^{1/4}.
\end{aligned} \tag{20.250}$$

Here we used the conversion

$$\begin{aligned}
\frac{g'(\bar{x}_j)}{g(\bar{x}_j)}\delta x_j &= \left(\frac{g'(x_j)+g''(x_j)\,\delta x_j/2}{g(x_j)+g'(x_j)\,\delta x_j/2}\right)\delta x_j \\
&= \frac{g'(x_j)}{g(x_j)}\delta x_j+\frac{gg''-g'^2}{2g^2}\delta x_j^{\,2},
\end{aligned} \tag{20.251}$$

and note that we only track the x_j vs. $\bar{x}_j$ dependence of the coefficients of δx_j, not δx_j^2, since the distinction does not matter in the latter case to $O(\delta t)$. Then the propagator (20.249) simplifies to

$$\begin{aligned}
K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\
\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(\bar{x}_j)}{2\delta t}\delta x_j^{\,2}-V[q(\bar{x}_j)]\,\delta t\right)\right]\right. \\
\left.\times\left[1+\left(\frac{gg''-g'^2}{16g^2}\right)\delta x_j^{\,2}+\frac{im(2gg''-g'^2)}{96\hbar g\,\delta t}\delta x_j^{\,4}\right]\right\}
\end{aligned} \tag{20.252}$$

upon transforming the δx piece. Using the replacements

$$\delta x^2\longrightarrow\frac{i\hbar}{mg(x)}\,\delta t,\qquad \delta x^4\longrightarrow-\frac{3\hbar^2}{m^2g^2(x)}\,\delta t^2, \tag{20.253}$$

to be justified later [Eq. (20.273), Section 20.4.2.3], the propagator becomes

$$\begin{aligned}K(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right)\\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(\bar{x}_j)}{2\delta t}\delta x_j^2 - V[q(\bar{x}_j)]\,\delta t\right)\right]\right.\\ &\qquad\left.\times\left[1+\frac{i\hbar(gg''-g'^2)}{16mg^3}\,\delta t - \frac{i\hbar(2gg''-g'^2)}{32mg^3}\,\delta t\right]\right\}\\ &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right)\\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(\bar{x}_j)}{2\delta t}\delta x_j^2 - V[q(\bar{x}_j)]\,\delta t\right)\right]\left[1-\frac{i\hbar g'^2}{32mg^3}\,\delta t\right]\right\},\end{aligned} \tag{20.254}$$

where the replacements are valid at any point in the time interval, because the difference is beyond order δt. Then if we finish the coordinate transformation of the potential,

$$V[q(\bar{x}_j)] \equiv V(\bar{x}_j), \tag{20.255}$$

and define the usual velocity

$$\dot{x}_j := \frac{\delta x_j}{\delta t} = \frac{x_{j+1}-x_j}{\delta t}, \tag{20.256}$$

(velocity)

then the path integral finally becomes[18]

$$\begin{aligned}K(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right)\\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\,\dot{x}_j^2 - V(\bar{x}_j) - \bar{V}_{\text{eff}}(\bar{x}_j)\right)\right]\right\},\end{aligned} \tag{20.257}$$

(midpoint path integral)

where we exponentiated the last term in the path integral, and we have defined the effective potential

$$\bar{V}_{\text{eff}}(x) := \frac{\hbar^2 g'^2}{32mg^3}. \tag{20.258}$$

(midpoint effective potential)

As in the q variables, we can write this path integral in short-hand continuous notation as

$$\begin{aligned}K(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\int D\left[x\sqrt{g(\bar{x})}\right]\,\exp\left[\frac{i}{\hbar}\bar{S}[x(t)]\right]\\ \bar{S}[\bar{x}(t)] &= \int_{t_0}^{t}dt\,\left[L(\bar{x},\dot{x}) - \bar{V}_{\text{eff}}(\bar{x})\right]\\ L(\bar{x},\dot{x}) &= \frac{m}{2}g(\bar{x})\,\dot{x}^2 - V(\bar{x}),\end{aligned} \tag{20.259}$$

(midpoint effective potential, continuous shorthand)

but it is particularly important to keep the discrete-form (20.257) as the fundamental object, since certain things such as the factors of g must be counted carefully for the prefactor to make sense. Note also that the midpoint path-integral action $\bar{S}(\bar{x},\dot{x})$ does *not* agree with the classical action with only the Lagrangian (20.196), being different by an effective potential that depends on the time-slicing expansion (and ordering assumptions regarding the Hamiltonian).

[18] Note that this expression agrees with the more general one given by Christian Grosche, "An Introduction into the Feynman Path Integral," arXiv.org preprint (arXiv: hep-th/9302097v1), Eq. (2.24); and in M. Chaichian and A. Demichev, *Path Integrals in Physics, Volume I: Stochastic Processes in Quantum Mechanics* (Institute of Physics, 2001), p. 249, Eq. (2.5.19).

20.4.2.2 Pre-Point (Itō) Expansion of the Path Integral

We can again expand the path integral (20.240), but now around the pre-point x_j to obtain a different (but equivalent) expression for the path integral. The recipe is the same as in the midpoint case:

1. Expand the kinetic energy in the action, keeping terms to $O(\delta t)$, regarding δx_j as $O(\sqrt{\delta t})$.
2. Expand the integration measure, to the extent necessary and also keeping terms to $O(\delta t)$.
3. Write all extra terms as a product outside the exponential.
4. Replace powers of δx_j outside the Gaussian path measure with powers of δt.
5. Exponentiate the converted terms.
6. Change velocity-like terms into prefactors if possible/desired.

Starting with the first item, we can expand the post-point coordinate,

$$q(x_{j+1}) = q(x_j) + q'(x_j)\,\delta x_j + \frac{q''(x_j)}{2}\delta x_j^2 + \frac{q'''(x_j)}{6}\delta x_j^3, \tag{20.260}$$

so that the short-time action (20.243) becomes (to order δt)

$$\begin{aligned}
\delta S[q_j] &= \frac{m}{2\delta t}\left[q(x_{j+1}) - q(x_j)\right]^2 - V[q(x_j)]\,\delta t \\
&= \frac{m}{2\delta t}\left[q'(x_j)\,\delta x_j + \frac{q''(x_j)}{2}\delta x_j^2 + \frac{q'''(x_j)}{6}\delta x_j^3\right]^2 - V[q(x_j)]\,\delta t \\
&= \frac{m}{2\delta t}q'^2(x_j)\,\delta x_j^2 + \frac{m}{2\delta t}q'(x_j)q''(x_j)\,\delta x_j^3 + \frac{m}{\delta t}\left[\frac{q''^2(x_j)}{8} + \frac{q'(x_j)q'''(x_j)}{6}\right]\delta x_j^4 - V[q(x_j)]\,\delta t \\
&= \frac{m}{2\delta t}g(x_j)\,\delta x_j^2 + \frac{mg'(x_j)}{4\delta t}\delta x_j^3 + \frac{mg'^2}{32g\,\delta t}\delta x_j^4 + \frac{m(2gg'' - g'^2)}{24g\,\delta t}\delta x_j^4 - V[q(x_j)]\,\delta t \\
&= \frac{m}{2\delta t}g(x_j)\,\delta x_j^2 + \frac{mg'(x_j)}{4\delta t}\delta x_j^3 + \frac{m(8gg'' - g'^2)}{96g\,\delta t}\delta x_j^4 - V[q(x_j)]\,\delta t,
\end{aligned} \tag{20.261}$$

where we have again used the derivatives (20.246). Now we proceed on to the measure, which doesn't need much except for a careful counting of factors:

$$\begin{aligned}
\prod_{j=1}^{N-1} dq_j &= \prod_{j=1}^{N-1} dx_j\, q'(x_j) \\
&= \frac{1}{q'(x_0)}\left(\prod_{j=1}^{N-1} dx_j\right)\prod_{j=0}^{N-1} q'(x_j) \\
&= \frac{1}{\sqrt{g(x_0)}}\left(\prod_{j=1}^{N-1} dx_j\right)\prod_{j=0}^{N-1}\sqrt{g(x_j)}.
\end{aligned} \tag{20.262}$$

Collecting all terms from the action (20.261) and measure (20.262), and expanding out the exponential in the propagator (20.240), we have

$$\begin{aligned}
K(x,t;x_0,t_0) = \frac{1}{\sqrt{g(x_0)}}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1} dx_j\right) \\
\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(x_j)}{2\delta t}\delta x_j^2 - V(x_j)\,\delta t\right)\right]\right. \\
\left.\times\left[1 + \frac{img'(x_j)}{4\hbar\delta t}\delta x_j^3 + \frac{im(8gg'' - g'^2)}{96\hbar g\,\delta t}\delta x_j^4 - \frac{m^2g'^2}{32\hbar^2\delta t^2}\delta x_j^6\right]\right\},
\end{aligned} \tag{20.263}$$

where we have carried out the expansion of the $\delta x^3/\delta t$ part to second order (generating a $\delta x^6/\delta t^2$ term) to keep the expansion consistent, and we have dropped the dependence on q in the potential. Then using the replacements (20.273) and (20.282), the propagator becomes

$$\begin{aligned} K(x,t;x_0,t_0) = \frac{1}{\sqrt{g(x_0)}}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(x_j)}{2\delta t}\delta x_j^{\,2}-V(x_j)\,\delta t\right)\right]\right. \\ \left.\times\left[1-\frac{3g'(x_j)}{4g(x_j)}\delta x_j-\frac{3i\hbar(8gg''-g'^2)}{96mg^3}\delta t+\frac{15i\hbar g'^2}{32mg^3}\delta t\right]\right\}. \end{aligned} \tag{20.264}$$

In preparation to exponentiate the last factor, we can add and subtract a δx^2 term,

$$\begin{aligned} K(x,t;x_0,t_0) &= \frac{1}{\sqrt{g(x_0)}}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(x_j)}{2\delta t}\delta x_j^{\,2}-V(x_j)\,\delta t\right)\right]\right. \\ &\quad\left.\times\left[1-\frac{3g'(x_j)}{4g(x_j)}\delta x_j+\frac{9g'^2}{32g^2}\delta x_j^{\,2}-\frac{9i\hbar g'^2}{32mg^3}\,\delta t-\frac{3i\hbar(8gg''-g'^2)}{96mg^3}\delta t+\frac{15i\hbar g'^2}{32mg^3}\delta t\right]\right\} \\ &= \frac{1}{\sqrt{g(x_0)}}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i}{\hbar}\left(\frac{mg(x_j)}{2\delta t}\delta x_j^{\,2}-V(x_j)\,\delta t\right)\right]\right. \\ &\quad\left.\times\left[1-\frac{3g'(x_j)}{4g(x_j)}\delta x_j+\frac{9g'^2}{32g^2}\delta x_j^{\,2}-\frac{i\hbar(gg''-g'^2)}{4mg^3}\delta t-\frac{i\hbar g'^2}{32mg^3}\delta t\right]\right\}, \end{aligned} \tag{20.265}$$

where we have used the remaining replacement in Eqs. (20.273):

$$\delta x^2 \longrightarrow \frac{i\hbar}{mg(x)}\,\delta t. \tag{20.266}$$

Along the same lines as Eq. (20.250), we can change the δx_j term into a prefactor via

$$\begin{aligned} \prod_{j=0}^{N-1}\left[1-\frac{3g'(x_j)}{4g(x_j)}\delta x_j+\frac{9g'^2}{32g^2}\delta x_j^{\,2}\right] &= \prod_{j=0}^{N-1}\exp\left[-\frac{g'(x_j)}{4g(x_j)}\delta x_j\right] \\ &= \prod_{j=0}^{N-1}\exp\left[-\delta\log g^{3/4}(x_j)+\frac{3(gg''-g'^2)}{8g^2}\delta x_j^{\,2}\right] \\ &= \prod_{j=0}^{N-1}\left[\frac{g(x_j)}{g(x_{j+1})}\right]^{3/4}\exp\left[\frac{3(gg''-g'^2)}{8g^2}\delta x_j^{\,2}\right] \\ &= \left[\frac{g(x_0)}{g(x)}\right]^{3/4}\prod_{j=0}^{N-1}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{3\hbar^2(gg''-g'^2)}{8mg^3}\right)\right]. \end{aligned} \tag{20.267}$$

Carrying this out in the propagator and exponentiating the remaining parts of the polynomial factor, we have the desired form[19]

$$K(x,t;x_0,t_0) = g^{-3/4}(x)\, g^{1/4}(x_0) \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \times \left\{ \prod_{j=0}^{N-1} \sqrt{g(x_j)}\, \exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2} g(x_j)\, \dot{x}_j^2 - V(x_j) - V_<(x_j)\right)\right] \right\},$$

(pre-point propagator) (20.268)

where the pre-point-ordering effective potential from the expansion is

$$V_<(x) := \frac{\hbar^2 g'^2}{32mg^3} - \frac{\hbar^2(gg'' - g'^2)}{8mg^3},$$

(20.269)
(pre-point effective potential)

and the velocity is as defined in Eq. (20.256). In continuous-notation shorthand, we can write this as

$$\begin{aligned} K(x,t;x_0,t_0) &= g^{1/4}(x_0)\, g^{-3/4}(x) \int D\left[x\sqrt{g(x)}\right]\, \exp\left[\frac{i}{\hbar} S_<[x(t)]\right] \\ S_<[x(t)] &= \int_{t_0}^{t} dt\ [L(x,\dot{x}) - V_<(x)] \\ L(x,\dot{x}) &= \frac{m}{2} g(x)\, \dot{x}^2 - V(x), \end{aligned}$$

(pre-point effective potential, continuous shorthand) (20.270)

where x is to be read as the pre-point, and again the more precise discrete form (20.268) is important to keep in mind for proper counting of factors. Note also that the action must be modified with an extra effective potential, which is different from the potential that came out of the midpoint discretization above.

20.4.2.3 "Moment" Relations

Now to justify the ''moment'' formulae of the quasi-Gaussian integrals that we have used. To handle even powers, we will quickly derive a form of the standard Gaussian moment theorem by integrating the product of a Gaussian and an arbitrary power once by parts:

$$\begin{aligned} \langle x^n \rangle &= \frac{1}{\sqrt{2\pi}\sigma} \int dx\, x^n\, e^{-(x-\mu)^2/2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int dx\, (x+\mu)^n\, e^{-x^2/2\sigma^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int dx\, (x+\mu)^{n-1} x\, e^{-x^2/2\sigma^2} + \frac{\mu}{\sqrt{2\pi}\sigma} \int dx\, (x+\mu)^{n-1}\, e^{-x^2/2\sigma^2} \\ &= \frac{(n-1)\sigma^2}{\sqrt{2\pi}\sigma} \int dx\, (x+\mu)^{n-2}\, e^{-x^2/2\sigma^2} + \mu \langle x^{n-1} \rangle \\ &= (n-1)\sigma^2 \langle x^{n-2} \rangle + \mu \langle x^{n-1} \rangle\,. \end{aligned}$$

(20.271)

In the path integral, we can think of the coordinate as δx, with δx^2 appearing in the Gaussian weight, so we can set $\mu = 0$ in the above recursion, obtaining

$$\langle x^n \rangle = (n-1)\, \sigma^2 \langle x^{n-2} \rangle\,. \tag{20.272}$$

[19] Note that this form generalized to multiple dimensions is quoted by Christian Grosche, ''An Introduction into the Feynman Path Integral,'' arXiv.org preprint (arXiv: hep-th/9302097v1), Eq. (1.7). Note, however, that the form quoted there appears somewhat different, as the quoted effective potential there vanishes in one dimension (where the Ricci scalar $R = 0$), and there is no prefactor. That same form is quoted by M. Chaichian and A. Demichev, *Path Integrals in Physics, Volume I: Stochastic Processes in Quantum Mechanics* (Institute of Physics, 2001), p. 246, Eq. (2.5.7).

This means that we can replace δx^n by $(n-1)[i\hbar\delta t/mg(x)]\,\delta x^{n-2}$, which implies the replacements

$$\delta x^2 \longrightarrow \frac{i\hbar}{mg(x)}\,\delta t, \qquad \delta x^4 \longrightarrow -\frac{3\hbar^2}{m^2g^2(x)}\,\delta t^2, \qquad \delta x^6 \longrightarrow -\frac{15i\hbar^3}{m^3g^3(x)}\,\delta t^3. \tag{20.273}$$

(even-order rules)

Technically, the variance-weight g is a function of the position, but here the width of the Gaussian is small $[O(\delta x)]$, and the rest of the integrand (beyond the normalized Gaussian weight) varies slowly over this range—any corrections to the moments due to this slow variation is ignorable at order δt. However, this is *not* true for the odd "moments," which would otherwise vanish, because the leading contribution comes from this variation. We can handle this more carefully as follows. First, we will define the "effective expectation"

$$E_{\sigma(x)}[x^n] := \int dx\, x^n\, \frac{1}{\sqrt{2\pi}\sigma(x)}\, e^{-x^2/2\sigma^2(x)}, \tag{20.274}$$

which differs from a Gaussian expectation because the integral is not a normalized Gaussian distribution if $\sigma(x)$ is not constant. In view of our discussion above, we have set the mean $\mu = 0$. Then using the expansions

$$\begin{aligned} \sigma^{-1}(x) &= \sigma^{-1}(0)\left[1 - \frac{\sigma'(0)}{\sigma(0)}\,x + O(x^2)\right] \\ \frac{x^2}{\sigma^2(x)} &= \frac{x^2}{\sigma^2(0)}\left[1 - \frac{2\sigma'(0)}{\sigma(0)}x + O(x^2)\right], \end{aligned} \tag{20.275}$$

we can write the nth effective moment (for $n > 2$) as

$$\begin{aligned} E_{\sigma(x)}[x^n] &\approx \frac{1}{\sqrt{2\pi}\sigma(0)}\int dx\, x^n \left[1 - \frac{\sigma'(0)}{\sigma(0)}x\right]\left[1 + \frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\, e^{-x^2/2\sigma^2(0)} \\ &\approx \frac{1}{\sqrt{2\pi}\sigma(0)}\int dx\, x^n \left[1 - \frac{\sigma'(0)}{\sigma(0)}x + \frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\, e^{-x^2/2\sigma^2(0)}, \end{aligned} \tag{20.276}$$

where we are working to order x in the expansions, keeping in mind that the contribution of the integral comes from x of the order σ, so we are counting σ and its derivative as being of the same order as x. Then writing the effective moment in terms of moments with respect to a Gaussian probability density of mean zero and variance $\sigma^2(0)$,

$$\begin{aligned} E_{\sigma(x)}[x^n] &= \langle x^n\rangle - \frac{\sigma'(0)}{\sigma(0)}\langle x^{n+1}\rangle + \frac{\sigma'(0)}{\sigma^3(0)}\langle x^{n+3}\rangle \\ &= \sigma^2(0)\left[(n-1)\langle x^{n-2}\rangle - n\frac{\sigma'(0)}{\sigma(0)}\langle x^{n-1}\rangle + (n+2)\frac{\sigma'(0)}{\sigma^3(0)}\langle x^{n+1}\rangle\right], \end{aligned} \tag{20.277}$$

where we have effectively integrated by parts by using the moment formula (20.272). Then the strategy is to match the coefficients of the three terms. Separating the n and the 2 in the last term and again using Eq. (20.272), we find that we can merge this last piece with the middle term:

$$\begin{aligned} E_{\sigma(x)}[x^n] &= \left\langle \sigma^2(0)\, x^{n-2}\left[(n-1) - n\frac{\sigma'(0)}{\sigma(0)}x + (n+2)\frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\right\rangle \\ &= \left\langle \sigma^2(0)\, x^{n-2}\left[(n-1) - n\frac{\sigma'(0)}{\sigma(0)}x + n\frac{\sigma'(0)}{\sigma^3(0)}x^3\right] + 2\frac{\sigma'(0)}{\sigma(0)}x^{n+1}\right\rangle \\ &= \left\langle \sigma^2(0)\, x^{n-2}\left[(n-1) - n\frac{\sigma'(0)}{\sigma(0)}x + n\frac{\sigma'(0)}{\sigma^3(0)}x^3\right] + 2n\sigma^2(0)\frac{\sigma'(0)}{\sigma(0)}x^{n-1}\right\rangle \\ &= \left\langle \sigma^2(0)\, x^{n-2}\left[(n-1) + n\frac{\sigma'(0)}{\sigma(0)}x + n\frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\right\rangle \\ &= n\left\langle \sigma^2(0)\, x^{n-2}\left[1 + \frac{\sigma'(0)}{\sigma(0)}x + \frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\right\rangle - \sigma^2(0)\,\langle x^{n-2}\rangle\,. \end{aligned} \tag{20.278}$$

Now using the expansion

$$\sigma^2(x) = \sigma^2(0)\left[1 + \frac{2\sigma'(0)}{\sigma(0)}\,x + O(x^2)\right], \tag{20.279}$$

we can restore the x-dependence of the σ^2 coefficient:

$$\begin{aligned} E_{\sigma(x)}[x^n] &\approx n\left\langle \sigma^2(0)\left[1 + \frac{2\sigma'(0)}{\sigma(0)}x\right] x^{n-2}\left[1 + \frac{\sigma'(0)}{\sigma(0)}x - \frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\right\rangle - \sigma^2(0)\left\langle x^{n-2}\right\rangle \\ &\approx n\left\langle \sigma^2(x)\,x^{n-2}\left[1 + \frac{\sigma'(0)}{\sigma(0)}x - \frac{\sigma'(0)}{\sigma^3(0)}x^3\right]\right\rangle - \sigma^2(0)\left\langle x^{n-2}\right\rangle \\ &\approx n\,E_{\sigma(x)}\left[\sigma^2(x)\,x^{n-2}\right] - \sigma^2(0)\left\langle x^{n-2}\right\rangle. \end{aligned} \tag{20.280}$$

In the last step, we used Eq. (20.276) to restore the effecive expectation, and we have consistently kept our expansions to order x. Thus, we have the rule for effective moments ($n > 2$)

$$E_{\sigma(x)}[x^n] = n\,E_{\sigma(x)}\left[\sigma^2(x)\,x^{n-2}\right] - \sigma^2(0)\left\langle x^{n-2}\right\rangle, \tag{20.281}$$

where the equality holds asymptotically for small $\sigma(x)$.

In the even-n case, this moment rule reduces to Eq. (20.272), since the difference between the effective and normal expectations is negligible at order δt. In the odd-n case, the last term vanishes, and this justifies the replacement rule

$$\delta x^3 \longrightarrow \frac{3i\hbar}{mg(x)}\,\delta t\,\delta x \tag{20.282}$$

(odd-order rule)

under the quasi-Gaussian integral, which is the only odd "moment" that we will need.

20.4.2.4 Midpoint–Pre-point Conversion

Now that we have derived two equivalent forms—Eqs. (20.257)–(20.259) for the midpoint form and Eqs. (20.268)–(20.270) for the pre-point form—we should check their equivalence explicitly. Starting with the mid-point path integral (20.257),

$$\begin{aligned} K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ \times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\,\dot{x}_j^{\,2} - V(\bar{x}_j) - \bar{V}_{\text{eff}}(\bar{x}_j)\right)\right]\right\}, \end{aligned} \tag{20.283}$$

we will then need to expand out the $g(\bar{x}_j)$ in the exponential as

$$g(\bar{x}_j) = g(x_j) + \frac{g'(x_j)}{2}\delta x_j + \frac{g''(x_j)}{8}\delta x_j^{\,2}, \tag{20.284}$$

and the measure factor as

$$\sqrt{g(\bar{x}_j)} = \sqrt{g(x_j)}\left[1 + \frac{g'(x_j)}{4g(x_j)}\delta x_j + \frac{g''(x_j)}{16g(x_j)}\delta x_j^{\,2} - \frac{g'^2(x_j)}{32g^2(x_j)}\delta x_j^{\,2}\right], \tag{20.285}$$

Writing out the expanded terms outside the exponential, we have

$$K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_{\text{eff}}(x_j)\right)\right] \times\left[1+\frac{img'(x_j)}{4\hbar\delta t}\delta x_j^{\,3}+\frac{img''}{16\hbar\delta t}\delta x_j^{\,4}-\frac{m^2g'^2}{32\hbar^2\delta t^2}\delta x_j^{\,6} +\frac{g'(x_j)}{4g(x_j)}\delta x_j+\frac{g''}{16g}\delta x_j^{\,2}-\frac{g'^2}{32g^2}\delta x_j^{\,2}+\frac{img'^2}{16\hbar g\delta t}\delta x_j^{\,4}\right]\right\}, \tag{20.286}$$

where we have consistently multiplied out and expanded to order δt as usual, and all quantities are evaluated at the pre-point x_j except as noted. Then using the replacements (20.273) and (20.282), we find

$$\begin{aligned} K(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_{\text{eff}}(x_j)\right)\right] \right. \\ &\quad\times\left[1-\frac{3g'(x_j)}{4g(x_j)}\delta x_j-\frac{3i\hbar g''}{16mg^2}\delta t+\frac{15i\hbar g'^2}{32mg^3}\delta t \right. \\ &\quad\left.\left.+\frac{g'(x_j)}{4g(x_j)}\delta x_j+\frac{i\hbar g''}{16mg^2}\delta t-\frac{i\hbar g'^2}{32mg^3}\delta t-\frac{3i\hbar g'^2}{16mg^3}\delta t\right]\right\} \\ &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_{\text{eff}}(x_j)\right)\right] \right. \\ &\quad\left.\times\left[1-\frac{g'(x_j)}{2g(x_j)}\delta x_j-\frac{i\hbar g''}{8mg^2}\delta t+\frac{i\hbar g'^2}{4mg^3}\delta t\right]\right\}, \end{aligned} \tag{20.287}$$

Then preparing to re-exponentiate, we add and subtract the δx^2 term,

$$\begin{aligned} K(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_{\text{eff}}(x_j)\right)\right] \right. \\ &\quad\left.\times\left[1-\frac{g'(x_j)}{2g(x_j)}\delta x_j+\frac{g'^2}{8g^2}\delta x_j^{\,2}-\frac{i\hbar g'^2}{8mg^3}\delta t-\frac{i\hbar g''}{8mg^2}\delta t+\frac{i\hbar g'^2}{4mg^3}\delta t\right]\right\} \\ &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\quad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_{\text{eff}}(x_j)\right)\right] \right. \\ &\quad\left.\times\left[1-\frac{g'(x_j)}{2g(x_j)}\delta x_j+\frac{g'^2}{8g^2}\delta x_j^{\,2}-\frac{i\hbar(gg''-g'^2)}{8mg^3}\delta t\right]\right\}. \end{aligned} \tag{20.288}$$

As in Eq. (20.267), we can handle the δx_j term as

$$\begin{aligned}\prod_{j=0}^{N-1}\left[1-\frac{g'(x_j)}{2g(x_j)}\delta x_j+\frac{g'^2}{8g^2}\delta x_j^{\,2}\right] &= \prod_{j=0}^{N-1}\exp\left[-\frac{g'(x_j)}{2g(x_j)}\delta x_j\right]\\ &= \prod_{j=0}^{N-1}\exp\left[-\delta\log\sqrt{g(x_j)}+\frac{gg''-g'^2}{4g^2}\delta x_j^{\,2}\right]\\ &= \prod_{j=0}^{N-1}\left[\frac{g(x_j)}{g(x_{j+1})}\right]^{1/2}\exp\left[\frac{gg''-g'^2}{4g^2}\delta x_j^{\,2}\right]\\ &= \left[\frac{g(x_0)}{g(x)}\right]^{1/2}\prod_{j=0}^{N-1}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{\hbar^2(gg''-g'^2)}{4mg^3}\right)\right].\end{aligned} \tag{20.289}$$

Implementing this and exponentiating the remaining polynomial terms, we have

$$\begin{aligned}K(x,t;x_0,t_0) = g^{-3/4}(x)\,g^{1/4}(x_0)\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right)\\ \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(x_j)\,\dot{x}_j^{\,2}-V(x_j)-\bar{V}_<(x_j)\right)\right]\right\},\end{aligned} \tag{20.290}$$

where the effective pre-point potential we derived from the mid-point propagator is

$$\bar{V}_<(x) := \bar{V}_{\rm eff}(x_j)+\frac{\hbar^2(gg''-g'^2)}{8mg^3}-\frac{\hbar^2(gg''-g'^2)}{4mg^3} = \bar{V}_{\rm eff}(x_j)-\frac{\hbar^2(gg''-g'^2)}{8mg^3}. \tag{20.291}$$

Putting in the midpoint effective potential (20.258), we find

$$\bar{V}_<(x) := \frac{\hbar^2 g'^2}{32mg^3}-\frac{\hbar^2(gg''-g'^2)}{8mg^3}, \tag{20.292}$$

which is exactly the effective potential $V_<(x)$ from Eq. (20.269) that we derived from the direct pre-point expansion.

20.4.2.5 Discussion: Integration Under Different Path Measures

To summarize the results of this section so far: beginning with the flat-space propagator (20.240), we have derived two *different* propagators—Eqs. (20.257) and (20.268)—by expanding about different points in the interval between the jth and $(j+1)$th time slices. We have also explicitly shown that they are *equivalent*, in that one can be derived from the other by changing the expansion point.

In the language of stochastic differential equations, these two path integrals correspond to Stratonovich (midpoint) and Itō (prepoint) calculi, at least in the imaginary-time forms of the propagators. In a broad sense, a different choice of calculus will correspond to a different family of paths. More explicitly, the midpoint propagator (20.257) generates imaginary-time paths according to the Stratonovich SDE

$$dx = \sqrt{\frac{\hbar}{mg(\bar{x})}}\,dW \equiv \sqrt{\frac{\hbar}{mg(x)}}\circ dW, \tag{20.293}$$

while the prepoint propagator generates paths via the Itō SDE

$$dx = \sqrt{\frac{\hbar}{mg(x)}}\,dt. \tag{20.294}$$

Note that we have not written a drift term, as in both propagators we have explicitly eliminated velocities and recast them in terms of prefactors and potentials. We know from the theory of SDEs in Section 17.4.2

that these two SDEs do not produce equivalent solutions. For example, casting the SDE (20.293) into Itō form would produce a drift term:

$$dx = -\frac{\hbar g'(x)}{4mg^2(x)}\,dt + \sqrt{\frac{\hbar}{mg(x)}}\,dW. \tag{20.295}$$

Again, this should be compared to the no-drift SDE (20.294) to understand why the paths are different. Then the reason that we obtain the quantum effective potentials in the propagators—and in particular *different* effective potentials for the same system but different choice of expansion point—is to compensate for these differences in paths. Since the paths sample space differently, the potential weights them differently, essentially reweighting the path integral for the new set of paths. That is, paths under one path measure that less likely under the other measure must be ''penalized'' with a reduced amplitude to compensate for the difference in measure.

We can also interpret the correction potentials at a more ''microscopic'' level. In arriving at the path integrals, we made the coordinate transformation defined by $g(x)$ in Eq. (20.203). When this is a nonlinear transformation, it maps the flat-space Gaussian probability densities for the steps into distorted Gaussian distributions. But the kinetic-energy factor (Gaussian in $\dot{x}$) is derived by expanding about the zero-slope point of the distorted Gaussian, which does not necessarily coincide with the mean of the distribution. This introduces an effective drift in the paths, which must be compensated for by the extra potential terms.

20.4.2.6 Cameron–Martin–Girsanov Transformation and Drift Potentials

Another common example of reweighting path integrals for different measures in the mathematical theory of SDEs is based on a change in the drift term of the paths. This is called a **Girsanov transformation**, or a **Cameron–Martin transformation**, or a **Cameron–Martin–Girsanov transformation** of measure.[20] Roughly speaking, this transformation applies when comparing driftless paths

$$x(t) = W(t) \tag{20.296}$$

to paths with a drift:

$$\tilde{x}(t) = x(t) + \int_0^t dt'\, a[\tilde{x}(t')]. \tag{20.297}$$

Note that this last expression can be written in differential form as

$$d\tilde{x}(t) = dx(t) + a[\tilde{x}(t)]\,dt = a[\tilde{x}(t)]\,dt + dW(t). \tag{20.298}$$

A path integral corresponding to the drifting trajectories in Eqs. (20.297) and (20.298) is then

$$I(x,t;x_0,0) = \int Dx\, \exp\Big[-\int_0^t dt'\,\frac{1}{2}\Big(\dot{x}(t') - a[x(t')]\Big)^2\Big], \tag{20.299}$$
(model path integral)

where we have dropped the twiddle on $\tilde{x}$. We can interpret the part of the measure inside the exponential here in discrete form as

$$\frac{1}{2}\Big(\dot{x}(t') - a[x(t')]\Big)^2\,dt = \frac{1}{2}\left(\frac{\delta x}{\delta t} - a(x)\right)^2 \delta t = \frac{(\delta x - a\,\delta t)^2}{2\delta t}, \tag{20.300}$$

so that each step in the walk still has variance δt, but now also a mean of $a(x)\,\delta t$.

Then the Cameron–Martin–Girsanov transformation boils down to expanding the quadratic in the exponential as follows:

$$I(x,t;x_0,0) = \int Dx\, \exp\Big[-\int_0^t dt'\,\frac{\dot{x}^2(t')}{2}\Big]\, \exp\Big[\int_0^t dx(t')\,a(x) - \frac{1}{2}\int_0^t dt'\,a^2(x)\Big]$$
(transformed path integral) (20.301)

[20] Andrei N. Borodin and Paavo Salminen, *Handbook of Brownian Motion—Facts and Formulae*, 2nd ed. (Birkhäuser, 2002), pp. 49-50.

Now the first exponential factor is the appropriate measure for trajectories *without* drift, as in Eq. (20.296). The second exponential factor acts as a ''potential'' that corrects for switching to the different path measure, and is called a **Radon–Nikodym derivative**. [The ''derivative'' here refers to being a ratio of the two path measures, in analogy to transforming between probability distributions $f_y(y)$ and $f_x(x)$ by setting $f_x(x) = f_y(y)(dy/dx)$ given a coordinate transformation $y(x)$.] This potential factor penalizes the paths that don't (randomly) move in the same way as the drifting paths, but rewards the ones that (randomly) *do* move in the right way.

However, the correction factor in Eq. (20.301) does not correspond directly to a potential as we have discussed it, in view of the stochastic integral [the first integral in this factor, with respect to $dx(t')$]. However, we can conveniently change this to a conventional integral as follows.[21] Let $A(x)$ be an antiderivative of $a(x)$; that is,

$$A'(x) = a(x). \tag{20.302}$$

Then

$$\begin{aligned} dA(x) &= A'(x)\,dx + \frac{1}{2}A''(x)\,(dx)^2 \\ &= a(x)\,dx + \frac{1}{2}a'(x)\,dt, \end{aligned} \tag{20.303}$$

where we used the Itō rule $dx^2 = dW^2 = dt$. Solving for $a\,dx$ and integrating gives

$$\int_0^t a(x)\,dx(t') = A[x(t)] - A[x(0)] - \frac{1}{2}\int_0^t a'[x(t')]\,dt'. \tag{20.304}$$

Thus, Eq. (20.301) becomes

$$I(x,t;x_0,0) = \int Dx\,\exp\biggl[-\int_0^t dt'\,\frac{\dot{x}^2(t')}{2}\biggr]\,\exp\biggl[-\frac{1}{2}\int_0^t dt'\,\Bigl(a'(x)+a^2(x)\Bigr)\biggr]\,\exp\Bigl[A[x(t)]-A[x(0)]\Bigr].$$

(transformed integral, potential form) (20.305)

Now the second exponential factor has the form of an effective potential

$$V_{\text{eff}}(x) = \frac{1}{2}\bigl[a'(x)+a^2(x)\bigr], \tag{20.306}$$

(effective drift potential)

and the last exponential factor represents an endpoint correction [which vanishes, for example, in the case of closed paths with $x(t) = x(0)$].

To wrap up this discussion of drift potentials, we should reexamine subtleties in some of the steps in the above treatment that we glossed over, which will connect with the discussion on operator orderings to follow below. In particular, note that the first integral in the last exponential factor of Eq. (20.301) is a stochastic integral, and thus we must specify a calculus to evaluate the integral. Given the expansion of Eq. (20.303), we are clearly interpreting this as an Itō integral [i.e., $a(x)$ in the path integral (20.299) is evaluated at the prepoint $x(t')$]. More precisely, given the path integral (20.305), it is equivalent to the path integral (20.299) provided we interpret the latter integral in the prepoint sense. In the path-integral language, the difference between, for example, a prepoint and midpoint interpretation comes about by introducing a correction of order dx to $a(x)$ in Eq. (20.301). The stochastic integral has a differential of dx, so the first-order correction must be kept in this term (while these corrections are negligible in the last integral). In the language of operator orderings below, such a term arises in the path integral from a term of the form $p\,a(x)$ in the corresponding quantum Hamiltonian, which can be ordered in different ways.

20.4.3 Operator Orderings

We have so far constructed the path integrals for the system corresponding to the Laplace–Beltrami Hamiltonian (20.195) by starting with a simpler Hamiltonian with no ordering issues, and then effecting a coordinate change that transforms both the Hamiltonian and the path integral into the desired form. However, we should

[21] Barry Simon, *Functional Integration and Quantum Physics* (Academic Press, 1979), pp. 172-3 (ISBN: 0126442509).

also be able to follow a modified version of the basic path-integral construction of Section 20.1, using the modified rules from Section 20.4.1.1 corresponding to the Laplace–Beltrami Hamiltonian (20.195). This requires some care due to the operator-ordering issues, and we will see that different operator orderings naturally give rise to different choices of stochastic calculus, or time-expansion points (midpoint vs. prepoint) from the coordinate-transformation derivations of the path integrals in the previous section.

20.4.3.1 Itō Calculus and p-x Ordering

As our first ordering example, consider the "p-x"-ordered version of (20.194),

$$H_{px}(x,p) = p^2 \frac{1}{2mg(x)} + V(x),$$

(variable-mass Hamiltonian, Itō ordering) (20.307)

will all momenta to the left and all positions to the right, which is called **anti-standard ordering**, or which we will call **Itō ordering** or **pre-point ordering**, for reasons we will soon see. By commuting one momentum, we have

$$H_{px}(x,p) = p\frac{1}{2mg(x)}p + i\hbar p\frac{g'}{2mg^2} + V(x) = H(x,p) + i\hbar p\frac{g'}{2mg^2}, \tag{20.308}$$

so that the product-ordered Hamiltonian is related to the Itō-ordered Hamiltonian via a commutator term linear in p. We can then relate this Hamiltonian to the Laplace–Beltrami Hamiltonian via Eq. (20.235), so that

$$H_{px}(x,p) = H_{\triangle}(x,p) + V_{\triangle}(x) + i\hbar p\frac{g'}{2mg^2}, \tag{20.309}$$

where the ordering potential $V_{\triangle}$ is defined in Eq. (20.236).

Now let's revisit the construction of the path integral beginning with Eqs. (20.10) and (20.11):

$$\begin{aligned} K_{px}(x,t;x_0,t_0) &= \langle x|e^{-iH_{px}(x,p)(t-t_0)/\hbar}|x_0\rangle \\ &= \int \left(\prod_{j=1}^{N-1}\sqrt{g(x_j)}\,dx_j\right)\langle x|e^{-iH_{px}\delta t/\hbar}|x_{N-1}\rangle\langle x_{N-1}|e^{-iH_{px}\delta t/\hbar}|x_{N-2}\rangle\cdots\langle x_1|e^{-iH_{px}\delta t/\hbar}|x_0\rangle. \end{aligned} \tag{20.310}$$

The change here—compared to the original path-integral construction of Section 20.1—is due to the modified position identity (20.213), where we have extra factors of $\sqrt{g}$ to keep track of. Then in analogy with Eq. (20.14), we can insert the momentum identity (20.230) into each mini-propagator. Because of the ordering choice, the natural place to insert the identity is to the left of the exponential, as the exponential itself is p-x-ordered, at least to order δt. So we have

$$\begin{aligned} \langle x_2|e^{-iH_{px}\delta t/\hbar}|x_1\rangle &= \int dp_1\,\langle x_2|p_1\rangle\langle p_1|e^{-iH_{px}\delta t/\hbar}|x_1\rangle \\ &= \int dp_1\,\langle x_2|p_1\rangle\langle p_1|x_1\rangle\,e^{-iH_{px}(x_1,p_1)\delta t/\hbar} \\ &= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\,e^{ip_1(x_2-x_1)/\hbar}e^{-iH_{px}(x_1,p_1)\delta t/\hbar} \\ &= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\,e^{i[p_1\dot{x}_1 - H_{px}(x_1,p_1)]\delta t/\hbar}, \end{aligned} \tag{20.311}$$

where as before

$$\delta x_1 := x_2 - x_1, \qquad \dot{x}_1 \equiv \frac{\delta x_1}{\delta t}, \tag{20.312}$$

and we have used the inner product (20.227). This is really the crucial step in the derivation: because of the ordering of the Hamiltonian, there was only one natural place to insert the momentum identity into the mini-propagator. That, as a result, led to the position operators in the Hamiltonian resolving to the eigenvalues

x_1, rather than x_2, and so we expect the result here to be consistent with the pre-point expansion of Section 20.4.2.2. Thus, in constructing path integrals, **p-x ordering in the Hamiltonian leads naturally to Itō calculus**, at least when we think about the equivalent path integral in the diffusion picture. Notice that this matches the intuition from casting SDEs directly into the Fokker–Planck (Kolmogorov forward) equation [see Eqs. (17.129) and (17.139)], which is a diffusion equation (Schrödinger equation in imaginary time). There the derivative operators in the diffusion term appear to the left of the state-dependent diffusion coefficient, just as the momentum operators appear to the left of the "diffusion coefficient" [which appears in the SDE (20.295)] in the Hamiltonian (20.308).

Proceeding to collect all the matrix elements as we did before to derive the phase-space path integral (20.28), we then have

$$\begin{aligned} K_{px}(x,t;x_0,t_0) &= \frac{1}{(2\pi\hbar)^{N/2}} \int \left(\prod_{j=1}^{N-1} \sqrt{g(x_j)}\, dx_j \right) \left\{ \prod_{j=0}^{N-1} dp_j \frac{1}{[g(x_{j+1})g(x_j)]^{1/4}}\, e^{i[p_j \dot{x}_j - H_{px}(x_j,p_j)]\delta t/\hbar} \right\} \\ &= \frac{1}{\sqrt{g(x_0)}\,(2\pi\hbar)^{N/2}} \int \left(\prod_{j=1}^{N-1} dx_j \right) \left\{ \prod_{j=0}^{N-1} dp_j \left[\frac{g(x_j)}{g(x_{j+1})} \right]^{1/4} e^{i[p_j \dot{x}_j - H_{px}(x_j,p_j)]\delta t/\hbar} \right\} \\ &= \frac{1}{[g(x)g(x_0)]^{1/4}\,(2\pi\hbar)^{N/2}} \int \left(\prod_{j=1}^{N-1} dx_j \right) \left\{ \prod_{j=0}^{N-1} dp_j\, e^{i[p_j \dot{x}_j - H_{px}(x_j,p_j)]\delta t/\hbar} \right\}. \end{aligned} \tag{20.313}$$

Thus, the phase-space path integral becomes

$$\begin{aligned} K_{px}(x,t;x_0,t_0) &= \frac{1}{[g(x)g(x_0)]^{1/4}\,(2\pi\hbar)^{N/2}} \int \left(\prod_{j=1}^{N-1} dx_j \right) \left\{ \prod_{j=0}^{N-1} dp_j \exp\left[\frac{i\delta t}{\hbar} \Big[p_j \dot{x}_j - H_{px}(x_j,p_j) \Big] \right] \right\} \\ &=: [g(x)g(x_0)]^{-1/4} \int Dx\, Dp\, \exp\left[\frac{i}{\hbar} \int_{t_0}^{t} dt \Big[p\dot{x} - H_{px}(x,p) \Big] \right] \end{aligned}$$

(phase-space propagator for H_{px}) (20.314)

in discrete and continuous notations.

We can proceed by working out the momentum integrals in Eq. (20.314), recalling that we are working with eigenvalues and thus there is no longer any concern about ordering. The result of each momentum integral is the same as in Eq. (20.19), if we replace m with $mg(x_j)$,

$$\int dp_j \exp\left[\frac{i\delta t}{\hbar} \Big(p_j \dot{x}_j - H_{px}(x_j,p_j) \Big) \right] = \sqrt{\frac{mg(x_j)}{i\delta t}}\, e^{iL(x_j,\dot{x}_j)\delta t/\hbar} \tag{20.315}$$

in terms of the Lagrangian (20.196). Thus, we obtain the propagator

$$K_{px}(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4} \left(\frac{m}{i2\pi\hbar\delta t} \right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j \right) \left\{ \prod_{j=0}^{N-1} \sqrt{g(x_j)} \exp\left[\frac{i\delta t}{\hbar} L(x_j,\dot{x}_j) \right] \right\}$$

(propagator for H_{px}) (20.316)

for the p-x-ordered Hamiltonian.

On the other hand, in view of Eq. (20.309), we can instead compute the propagator for the Hamiltonian $H_\triangle(x,p)$ by replacing $H_{px}(x,p) \longrightarrow H_\triangle(x,p) = H_{px}(x,p) - V_\triangle(x) - i\hbar p g'/2mg^2$ in the propagator (20.314):

$$\begin{aligned} K(x,t;x_0,t_0) = \frac{1}{[g(x_0)g(x)]^{1/4}\,(2\pi\hbar)^{N/2}} \int \left(\prod_{j=1}^{N-1} dx_j \right) \\ \times \left\{ \prod_{j=0}^{N-1} dp_j \exp\left[\frac{i\delta t}{\hbar} \left(p_j \dot{x}_j - H_{px}(x_j,p_j) + V_\triangle(x_j) + i\hbar p_j \frac{g'}{2mg^2} \right) \right] \right\}. \end{aligned} \tag{20.317}$$

The result is the same as for the K_{px} propagator, then, but with a modified velocity and potential:

$$K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}L\left(x_j,\dot{x}_j+i\hbar\frac{g'(x_j)}{2mg^2(x_j)}\right)+\frac{i\delta t}{\hbar}V_\triangle(x_j)\right]\right\}. \tag{20.318}$$

Expanding out the Lagrangian, we have

$$\begin{aligned} L\left(x_j,\dot{x}_j+i\hbar\frac{g'(x_j)}{2mg^2(x_j)}\right) &= \frac{m}{2}g(x_j)\left(\dot{x}_j+i\hbar\frac{g'(x_j)}{2mg^2(x_j)}\right)^2 - V(x_j) \\ &= \frac{m}{2}g(x_j)\,\dot{x}_j^2 - V(x_j) + i\hbar\frac{g'(x_j)}{2g^2(x_j)}\dot{x}_j - \frac{\hbar^2 g'^2}{8mg^3} \\ &= L(x_j,\dot{x}_j) + i\hbar\frac{g'(x_j)}{2g^2(x_j)}\dot{x}_j - \frac{\hbar^2 g'^2}{8mg^3}, \end{aligned} \tag{20.319}$$

and so the propagator becomes

$$K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(L(x_j,\dot{x}_j)+V_\triangle(x_j)-\frac{\hbar^2 g'^2}{8mg^3}\right)-\frac{g'(x_j)}{2g^2(x_j)}\delta x_j\right]\right\}. \tag{20.320}$$

From Eq. (20.289) we have

$$\prod_{j=0}^{N-1}\exp\left[-\frac{g'(x_j)}{2g(x_j)}\delta x_j\right] = \left[\frac{g(x_0)}{g(x)}\right]^{1/2}\prod_{j=0}^{N-1}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{\hbar^2(gg''-g'^2)}{4mg^3}\right)\right], \tag{20.321}$$

and using this to replace the δx_j term in the exponential, we find

$$K(x,t;x_0,t_0) = g^{-3/4}(x)\,g^{1/4}(x_0)\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(L(x_j,\dot{x}_j)+V_\triangle(x_j)+\frac{\hbar^2(gg''-g'^2)}{4mg^3}-\frac{\hbar^2 g'^2}{8mg^3}\right)\right]\right\}. \tag{20.322}$$

From Eq. (20.236) we had

$$V_\triangle(x,p) := -\frac{\hbar^2(gg''-2g'^2)}{8mg^3} - \frac{\hbar^2 g'^2}{32mg^3}, \tag{20.323}$$

so the propagator finally becomes

$$K(x,t;x_0,t_0) = g^{-3/4}(x)\,g^{1/4}(x_0)\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(x_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(L(x_j,\dot{x}_j)+\frac{\hbar^2(gg''-g'^2)}{8mg^3}-\frac{\hbar^2 g'^2}{32mg^3}\right)\right]\right\}. \tag{20.324}$$

(pre-point propagator for $H_\triangle$)

This exactly matches the pre-point (corresponding to integration in the Itō sense) propagator (20.268) that we derived by the point transformation, with the same effective potential (20.269). Again, this corresponds to the Hamiltonian $H_\triangle(x,p)$, and follows in the operator formalism by first reordering this Hamiltonian in anti-standard (p-x) form, and then introducing momentum identities to derive the path integral.

20.4.3.2 Anticipating Calculus and x-p Ordering

By contrast, if we have a Hamiltonian $H_{xp}(x,p)$ that is "x-p"-ordered or in **standard ordering**,

$$H_{xp}(x,p) = \frac{1}{2mg(x)}p^2 + V(x),$$

(variable-mass Hamiltonian, anticipating ordering) (20.325)

with all x's to the left, all p's to the right, then we can modify Eqs. (20.311) to become

$$\begin{aligned}
\langle x_2|e^{-iH_{xp}\delta t/\hbar}|x_1\rangle &= \int dp_1\, \langle x_2|e^{-iH_{xp}\delta t/\hbar}|p_1\rangle\langle p_1|x_1\rangle \\
&= \int dp_1\, \langle x_2|p_1\rangle\langle p_1|x_1\rangle e^{-iH_{xp}(x_2,p_1)\delta t/\hbar} \\
&= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}} \int dp_1\, e^{ip_1(x_2-x_1)/\hbar} e^{-iH_{xp}(x_2,p_1)\delta t/\hbar} \\
&= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}} \int dp_1\, e^{i[p_1\dot{x}_1 - H_{xp}(x_2,p_1)]\delta t/\hbar},
\end{aligned} \tag{20.326}$$

with the difference being that $H(x,p)$ is now evaluated at the *final* point x_2 in the interval $[x_1, x_2]$.

The path integral has the same form (20.28) as for p-x ordering, but now in the integration, the x in the Hamiltonian is evaluated at the *final* time in each time interval $[t, t+dt]$. In the imaginary-time diffusion picture, this again corresponds to integrating with respect to diffusive (Wiener) paths, where the integrand is always evaluated at the final point of each time interval, which is called **anticipating calculus**. Thus, for diffusive path integrals, **x-p ordering leads naturally to anticipating calculus**.

The rest of the derivation carries through in the same way, and for example we obtain the post-point propagator

$$\begin{aligned}
K(x,t;x_0,t_0) = g^{1/4}(x)\, g^{-3/4}(x_0) \left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \\
\times \left\{ \prod_{j=0}^{N-1} \sqrt{g(x_{j+1})}\, \exp\left[\frac{i\delta t}{\hbar}\left(L(x_{j+1},\dot{x}_j) + \frac{\hbar^2(gg'' - g'^2)}{8mg^3} - \frac{\hbar^2 g'^2}{32mg^3}\right)\right] \right\}.
\end{aligned}$$

(post-point propagator for $H_{\triangle}$) (20.327)

in place of the pre-point form (20.324), essentially by interchanging $x_j \longleftrightarrow x_{j+1}$ in the integrand and prefactor.

20.4.3.3 Stratonovich Calculus and Weyl Ordering

Perhaps the most useful ordering of the Hamiltonian to consider is Weyl (symmetrized) ordering (see Section 4.3.5). We can write the Weyl-ordered version of the Hamiltonian (20.194) or (20.195) as

$$H_{\rm W}(x,p) = \left(p\frac{1}{2mg(x)}p\right)_{\rm W} + V(x),$$

(variable-mass Hamiltonian, Weyl ordering) (20.328)

where the subscript denotes Weyl ordering. To relate this form to other orderings, we can take advantage of the product rule for Weyl correspondence (4.121)

$$A(x,p) = B\left(x - \frac{\hbar}{2i}\partial_p, p + \frac{\hbar}{2i}\partial_x\right) C(x,p) = C\left(x + \frac{\hbar}{2i}\partial_p, p - \frac{\hbar}{2i}\partial_x\right) B(x,p) \tag{20.329}$$

for an operator product $\hat{A} = \hat{B}\hat{C}$. Applying this rule twice,

$$\begin{aligned}\hat{p}\frac{1}{g(\hat{x})}\hat{p} &\longleftrightarrow \left(p + \frac{\hbar}{2i}\partial_x\right)\left(p - \frac{\hbar}{2i}\partial_x\right)\frac{1}{g(x)} \\ &= p\frac{1}{g(x)}p + \frac{\hbar^2}{4}\left(\frac{1}{g}\right)'' \\ &= p\frac{1}{g(x)}p - \frac{\hbar^2(gg'' - 2g'^2)}{4g^3} \\ &\longleftrightarrow \left(\hat{p}\frac{1}{g(\hat{x})}\hat{p}\right)_{\rm W} - \frac{\hbar^2(gg'' - 2g'^2)}{4g^3}\end{aligned} \tag{20.330}$$

where arrows denote Weyl correpondence between classical functions and operators, and we are being careful for the moment to mark operators with hats. That is, we can write

$$H(x,p) = H_{\rm W}(x,p) + V_{\rm W}(x), \tag{20.331}$$

where the effective ordering potential to switch between product and Weyl orderings is

$$V_{\rm W}(x) := -\frac{\hbar^2(gg'' - 2g'^2)}{8mg^3}. \tag{20.332}$$

Then using Eqs. (20.235) and (20.236), we can also relate the Weyl and Laplace–Beltrami Hamiltonians via

$$\begin{aligned}H_{\rm W}(x,p) &= H(x,p) - V_{\rm W}(x) \\ &= H_{\triangle}(x,p) + V_{\triangle}(x,p) - V_{\rm W}(x) \\ &= H_{\triangle}(x,p) - \frac{\hbar^2 g'^2}{32mg^3},\end{aligned} \tag{20.333}$$

so that the ordering potential here is simpler than either $V_{\triangle}$ or $V_{\rm W}$.

Now suppose we construct the propagator for the Weyl-ordered Hamiltonian by $H_{\rm W}(x,p)$. In adapting Eqs. (20.311) to this case, it is not obvious where to insert the momentum identity, but there is a trick we can use to make this simple.[22] Recall that in terms of the characteristic operator (4.92),

$$\hat{M}(\pi_x, \pi_p) = e^{i(\pi_x\hat{x} + \pi_p\hat{p})/\hbar}. \tag{20.334}$$

an arbitrary Weyl-ordered operator, for example the Hamiltonian, is obtained from the classical function from the Weyl correspondence (4.97)

$$\hat{H}_{\rm W}(\hat{x},\hat{p}) = \frac{1}{(2\pi\hbar)^2}\int_{-\infty}^{\infty}d\pi_x\int_{-\infty}^{\infty}d\pi_p\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dp\; H_{\rm W}(x,p)\,\hat{M}(\pi_x,\pi_p)\,e^{-i(\pi_x x + \pi_p p)/\hbar}, \tag{20.335}$$

(Note that we are again being careful for the moment about distinguishing classical variables from operators.) In this representation, the only operator-dependence in the Hamiltonian comes in the form of the characteristic operator. Consider the analogous matrix element of the characteristic operator:

$$\begin{aligned}\langle x_2|\hat{M}(\pi_x,\pi_p)|x_1\rangle &= \langle x_2|e^{i(\pi_x\hat{x} + \pi_p\hat{p})/\hbar}|x_1\rangle \\ &= \langle x_2|e^{i\pi_x\hat{x}/2\hbar}e^{i\pi_p\hat{p}/\hbar}e^{i\pi_x\hat{x}/2\hbar}|x_1\rangle \\ &= \int dp_1\,\langle x_2|e^{i\pi_x\hat{x}/2\hbar}|p_1\rangle\langle p_1|e^{i\pi_p\hat{p}/\hbar}e^{i\pi_x\hat{x}/2\hbar}|x_1\rangle \\ &= \int dp_1\,\langle x_2|p_1\rangle\langle p_1|x_1\rangle e^{i\pi_x x_2/2\hbar}e^{i\pi_p p_1/\hbar}e^{i\pi_x x_1/2\hbar} \\ &= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\,e^{ip_1(x_2-x_1)/\hbar}e^{i\pi_x x_2/2\hbar}e^{i\pi_p p_1/\hbar}e^{i\pi_x x_1/2\hbar} \\ &= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\,e^{ip_1\dot{x}_1\delta t/\hbar}e^{i\pi_x(x_1+x_2)/2\hbar}e^{i\pi_p p_1/\hbar},\end{aligned} \tag{20.336}$$

[22] see, e.g., Christian Grosche, "An Introduction into the Feynman Path Integral," arXiv.org preprint (arXiv: hep-th/9302097v1).

where we have used a symmetric splitting, similar to Eq. (4.101), which is *exact* here, because the argument of the exponential is linear in x and p. Then the similar matrix element of the Hamiltonian becomes

$$\begin{aligned}
\langle x_2|\hat{H}_{\rm W}(\hat{x},\hat{p})|x_1\rangle &= \frac{1}{(2\pi\hbar)^2}\int d\pi_x \int d\pi_p \int dx \int dp\; H_{\rm W}(x,p)\,\langle x_2|\hat{M}(\pi_x,\pi_p)|x_1\rangle\, e^{-i(\pi_x x+\pi_p p)/\hbar} \\
&= \frac{1}{(2\pi\hbar)^3[g(x_2)g(x_1)]^{1/4}}\int dp_1\, e^{ip_1\dot{x}_1\delta t/\hbar} \\
&\qquad\qquad \times \int d\pi_x \int d\pi_p \int dx \int dp\, H_{\rm W}(x,p)\, e^{i\pi_x(x_1+x_2)/2\hbar} e^{i\pi_p p_1/\hbar} e^{-i(\pi_x x+\pi_p p)/\hbar} \\
&= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_2\, e^{ip_1\dot{x}_1\delta t/\hbar}\int dx\int dp\; H_{\rm W}(x,p)\,\delta\!\left(x-\frac{x_1+x_2}{2}\right)\delta(p-p_1) \\
&= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_2\, e^{ip_1\dot{x}_1\delta t/\hbar} H_{\rm W}(\bar{x}_1,p_1),
\end{aligned} \tag{20.337}$$

where as before we have defined the midpoint

$$\bar{x}_j := \frac{x_{j+1}+x_j}{2}. \tag{20.338}$$

Now since

$$e^{-iH\delta t/\hbar} = 1 - \frac{iH}{\hbar}\delta t + O(\delta t^2), \tag{20.339}$$

with the second-order contributions negligible, the same argument applies to matrix elements of the mini-evolution operators, and thus Eqs. (20.337) become

$$\langle x_2|e^{-iH_{\rm W}\delta t/\hbar}|x_1\rangle = \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\, e^{i[p_1\dot{x}_1 - H_{\rm W}(\bar{x}_1,p_1)]\delta t/\hbar}. \tag{20.340}$$

Again proceeding to collect all the matrix elements as we did before to derive the phase-space path integral (20.313), we have

$$\begin{aligned}
K_{\rm W}(x,t;x_0,t_0) &= \frac{1}{(2\pi\hbar)^{N/2}}\int\left(\prod_{j=1}^{N-1}\sqrt{g(x_j)}\,dx_j\right)\left\{\prod_{j=0}^{N-1} dp_j \frac{1}{[g(x_{j+1})g(x_j)]^{1/4}}\, e^{i[p_j\dot{x}_j - H_{\rm W}(\bar{x}_j,p_j)]\delta t/\hbar}\right\} \\
&= \frac{1}{\sqrt{g(x_0)}\,(2\pi\hbar)^{N/2}}\int\left(\prod_{j=1}^{N-1} dx_j\right)\left\{\prod_{j=0}^{N-1} dp_j \left[\frac{g(x_j)}{g(x_{j+1})}\right]^{1/4} e^{i[p_j\dot{x}_j - H_{\rm W}(\bar{x}_j,p_j)]\delta t/\hbar}\right\} \\
&= \frac{1}{[g(x)g(x_0)]^{1/4}\,(2\pi\hbar)^{N/2}}\int\left(\prod_{j=1}^{N-1} dx_j\right)\left\{\prod_{j=0}^{N-1} dp_j\, e^{i[p_j\dot{x}_j - H_{\rm W}(\bar{x}_j,p_j)]\delta t/\hbar}\right\}.
\end{aligned} \tag{20.341}$$

Thus, the phase-space path integral becomes

$$\begin{aligned}
K_{\rm W}(x,t;x_0,t_0) &= \frac{1}{[g(x)g(x_0)]^{1/4}\,(2\pi\hbar)^{N/2}}\int\left(\prod_{j=1}^{N-1} dx_j\right)\left\{\prod_{j=0}^{N-1} dp_j \exp\left[\frac{i\delta t}{\hbar}\Big[p_j\dot{x}_j - H_{\rm W}(\bar{x}_j,p_j)\Big]\right]\right\} \\
&=: [g(x)g(x_0)]^{-1/4}\int Dx\,Dp\,\exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt\,\Big[p\dot{x} - H_{\rm W}(\bar{x},p)\Big]\right]
\end{aligned}$$

(phase-space propagator for Weyl-ordered Hamiltonian) (20.342)

in discrete and continuous notations. Now the path integral appears with the position in $H(x,p)$ evaluated at the middle time $t+dt/2$ in the integration interval $[t,t+dt]$. Thus, **Weyl ordering leads naturally to Stratonovich calculus** in the sense of a diffusive path integral. We have indicated the spatial dependence in the Hamiltonian by $\bar{x}$ to emphasize this, and again, $H_{\rm W}(x,p)$ here is the classical Hamiltonian obtained from the Weyl-ordered Hamiltonian by replacing position and momentum operators by eigenvalues. This ordering is convenient, as Stratonovich integration proceeds according to the rules of "normal" calculus, and the integral is explicitly symmetric under time-reversal.

Proceeding to the Lagrangian form of the path integral by carrying out the momentum integrals, we can again note that the result of each momentum integral is the same as in Eq. (20.19), if we replace m with $mg(\bar{x}_j)$,

$$\int dp_j \exp\left[\frac{i\delta t}{\hbar}\Big(p_j\dot{x}_j - H_{\rm w}(\bar{x}_j,p_j)\Big)\right] = \sqrt{\frac{mg(\bar{x}_j)}{i\delta t}}\, e^{iL(\bar{x}_j,\dot{x}_j)\delta t/\hbar} \tag{20.343}$$

in terms of the Lagrangian (20.196). Thus, we obtain the propagator

$$K_{\rm w}(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right)\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}L(\bar{x}_j,\dot{x}_j)\right]\right\}$$

(propagator for Weyl-ordered Hamiltonian) (20.344)

for the Weyl-ordered Hamiltonian $H_{\rm w}(x,p)$.

If we instead want the propagator corresponding to $H_{\triangle}(x,p)$, we can use relation (20.333) to make the substitution $H_{\rm w} \longrightarrow H_{\triangle} = H_{\rm w} + \hbar^2 g'^2/32mg^3$ in the path integral (20.342). As such, the modification is much easier here than in the Itō-ordered case, and we directly obtain the propagator

$$K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(L(\bar{x}_j,\dot{x}_j) - \frac{\hbar^2 g'^2}{32mg^3}\right)\right]\right\}.$$

(propagator for $H_{\triangle}(x,p)$) (20.345)

This propagator is the same as the propagator (20.257) that we obtained by expanding the short-time action about the midpoint in the coordinate-transformation derivation.

20.4.3.4 Symmetrized Ordering

Note that the Weyl ordering is different from the other obvious symmetric choice, that of symmetric ordering:

$$H_{\{xp\}} := \frac{H_{px}+H_{xp}}{2} = \frac{1}{4m}\left[p^2,\frac{1}{g(x)}\right]_+ + V(x).$$

(Hamiltonian for symmetrized ordering) (20.346)

Here we used the Itō and anticipating orderings from Eqs. (20.325) and (20.307), respectively. In this case, the Hamiltonian corresponding to the matrix element (20.311) is evaluated at each endpoint and *then* averaged, as we will see shortly. This is closer to Stratonovich calculus than either Itō or anticipating, in that it focuses on the middle of the interval (by averaging the endpoints).

To derive the path integral for this ordering, we will follow the procedure in Section 20.4.3.1. The propagator still has the form of Eq. (20.310), but upon inserting momentum identities, we have matrix elements

$$\begin{aligned}
\langle x_2|e^{-iH_{\{xp\}}\delta t/\hbar}|x_1\rangle &= \langle x_2|e^{-iH_{xp}\delta t/2\hbar}e^{-iH_{px}\delta t/2\hbar}|x_1\rangle \\
&= \int dp_1\,\langle x_2|e^{-iH_{xp}\delta t/2\hbar}|p_1\rangle\langle p_1|e^{-iH_{px}\delta t/2\hbar}|x_1\rangle \\
&= \int dp_1\,\langle x_2|p_1\rangle e^{-iH(x_2,p_1)\delta t/2\hbar}e^{-iH(x_1,p_1)\delta t/2\hbar}\langle p_1|x_1\rangle \\
&= \frac{1}{2\pi\hbar[g(x_2)g(x_1)]^{1/4}}\int dp_1\, e^{i\{p_1\dot{x}_1 - [H(x_1,p_1)+H(x_2,p_1)]/2\}\delta t/\hbar},
\end{aligned} \tag{20.347}$$

where the equalities here are only up to $O(\delta t)$. The derivation then carries through as in the Itō case through the Hamiltonian path integral (20.314). Then carrying out the momentum integrals, the idea is the same as

in Eq. (20.315), but $g(x_j)$ is replaced by

$$\left(\frac{1}{2g(x_j)}+\frac{1}{2g(x_{j+1})}\right)^{-1}=\frac{2g(x_j)g(x_{j+1})}{g(x_j)+g(x_{j+1})}, \tag{20.348}$$

so that (20.315) becomes in this case

$$\int dp_j \exp\left[\frac{i\delta t}{\hbar}\Big(p_j\dot{x}_j - H_{px}(x_j,p_j)\Big)\right]=\sqrt{\frac{m[Mg^{-1}(x_j)]^{-1}}{i\delta t}}\,e^{i\{m[Mg^{-1}(x_j)]^{-1}\dot{x}_j^2-MV(x_j)\}\delta t/\hbar}, \tag{20.349}$$

where we have introduced the **averaging operator** on the time lattice:

$$Mf(x_j):=\frac{f(x_j)+f(x_{j+1})}{2}. \tag{20.350}$$

Thus, we obtain the propagator

$$K_{\{xp\}}(x,t;x_0,t_0)=[g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{[Mg^{-1}(x_j)]^{-1}}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}[Mg^{-1}(x_j)]^{-1}\dot{x}_j^2-MV(x_j)\right)\right]\right\}$$

(propagator for symmetrized Hamiltonian) (20.351)

for the symmetrized Hamiltonian (20.346).

Note that we can also rewrite this path integral in terms of the midpoint

$$\bar{x}_j:=\frac{x_{j+1}+x_j}{2}, \tag{20.352}$$

using the expansions

$$\begin{aligned} g(x_j)&=g(\bar{x}_j)-\frac{g'(\bar{x}_j)}{2}\delta x_j+\frac{g''(\bar{x}_j)}{8}\delta x_j^2\\ g(x_{j+1})&=g(\bar{x}_j)+\frac{g'(\bar{x}_j)}{2}\delta x_j+\frac{g''(\bar{x}_j)}{8}\delta x_j^2. \end{aligned} \tag{20.353}$$

Thus, we will need

$$[Mg^{-1}(x_j)]^{-1}=g(\bar{x}_j)+\left(\frac{gg''-2g'^2}{8g}\right)\delta x_j^2+\frac{g'^2g''}{32g^2}\delta x_j^4, \tag{20.354}$$

up to fourth order, where all instances of g on the right-hand side are evaluated at the midpoint. Expanding the parts of the path integral that need it,

$$\begin{aligned} \sqrt{[Mg^{-1}(x_j)]^{-1}}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}[Mg^{-1}(x_j)]^{-1}\dot{x}_j^2\right)\right]&=\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\dot{x}_j^2\right)\right]\\ &\quad\times\left(1+\frac{gg''-2g'^2}{16g^2}\delta x^2+im\frac{gg''-2g'^2}{16\hbar g}\frac{\delta x^4}{\delta t}\right)\\ &=\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\dot{x}_j^2\right)\right]\\ &\quad\times\left(1+i\hbar\frac{gg''-2g'^2}{16mg^3}\delta t-3i\hbar\frac{gg''-2g'^2}{16mg^3}\delta t\right)\\ &=\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\dot{x}_j^2\right)\right]\left(1-i\hbar\frac{gg''-2g'^2}{8mg^3}\delta t\right), \end{aligned} \tag{20.355}$$

where we used the replacements (20.273) and consistently expanded to order δt. Thus, in midpoint form, the propagator (20.351) becomes

$$K_{\{xp\}}(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\dot{x}_j^2 - V(\bar{x}_j) - \frac{\hbar^2(gg''-2g'^2)}{8mg^3}\right)\right]\right\},$$

(propagator for symmetrized Hamiltonian, midpoint form) (20.356)

where we have also noted that $MV(x_j)$ can be replaced by $V(\bar{x}_j)$ at order δt.

To compare with other Hamiltonians, we can consider reordering the original Hamiltonian (20.362) in symmetrized form, using

$$p\frac{1}{g(x)}p = \frac{1}{2}\left(p^2\frac{1}{g}+\frac{1}{g}p^2\right)+\frac{i\hbar}{2}\left[p,-\frac{g'}{g^2}\right] = \frac{1}{2}\left(p^2\frac{1}{g}+\frac{1}{g}p^2\right)-\frac{\hbar^2}{2}\frac{gg''-2g'^2}{g^3}, \tag{20.357}$$

so that

$$H(x,p) = H_{\{xp\}} - \frac{\hbar^2(gg''-2g'^2)}{4mg^3}. \tag{20.358}$$

Then with Eqs. (20.235) and (20.236), we can compare the Laplace–Beltrami ordering to the symmetrized form as

$$H_{\triangle}(x,p) = H(x,p)+\frac{\hbar^2(gg''-2g'^2)}{8mg^3}+\frac{\hbar^2g'^2}{32mg^3} = H_{\{xp\}}-\frac{\hbar^2(gg''-2g'^2)}{8mg^3}+\frac{\hbar^2g'^2}{32mg^3}. \tag{20.359}$$

Then putting in the extra terms as correction potentials in the propagator (20.316), we can obtain the propagator for $H_{\triangle}(x,p)$ as

$$K(x,t;x_0,t_0) = [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\exp\left[\frac{i\delta t}{\hbar}\left(L(\bar{x}_j,\dot{x}_j) - \frac{\hbar^2g'^2}{32mg^3}\right)\right]\right\}.$$

(symmetrized propagator for $H_{\triangle}(x,p)$) (20.360)

Note that this is identical to the propagator (20.345) that we derived in Weyl ordering. Thus, in terms of stochastic differential equations, both orderings generate trajectories that correspond to Stratonovich calculus.

20.4.3.5 Product Ordering

Yet another symmetric-type ordering for the Hamiltonian (20.194) is called **product ordering**, and the product-ordered analogue of the model Hamiltonian is

$$H_{gpg} := \frac{1}{\sqrt{g(x)}}\frac{p^2}{2m}\frac{1}{\sqrt{g(x)}}+V(x).$$

(Hamiltonian for symmetrized ordering) (20.361)

We will leave it as an exercise to derive the propagator for this Hamiltonian and for the Laplace–Beltrami Hamiltonian $H_{\triangle}(x,p)$ (see Problem 20.1). However, the upshot is that, once written in midpoint form, the latter propagator is identical to the version derived in Weyl and symmetrized orderings. This conclusion also holds for more general path integrals in curved spacetime.[23]

[23] C. Grosche and F. Steiner, *Handbook of Feynman Path Integrals* (Springer, 1998), p. 170.

20.4.3.6 Equivalence Classes of Orderings

An important conclusion from the above path integrals is that although we can define several path integrals that treat the beginning and end time slices (x_j and x_{j+1}) symmetrically, they end up generating identical path integrals when written in terms of the midpoint $\bar{x}_j$, and thus all correspond to Stratonovich calculus with the same effective potential. Because of the need to Taylor-expand the path integral to fourth order in δx, it seems as if there are many more possible path integrals than choices for stochastic calculus (where we only expand to second order). However, we see that different path integrals that correspond to the same calculus turn out to be equivalent. That is, there is *really* only one path integral for each calculus, but many possible ways to write it down.

20.4.4 Normalizing the Weyl-Ordered Path Integral

Now we want to return to our original model Hamiltonian (20.194) for a variable mass,

$$H(x,p) = p\frac{1}{2mg(x)}p + V(x), \tag{20.362}$$
(variable-mass Hamiltonian)

and consider how to develop a normalized path integral for Monte–Carlo calculations in the sense of Section 20.3.1. In previous sections, we considered propagators for the Hamiltonian in Laplace–Beltrami form $H_\triangle(x,p)$, but we need the propagators for $H(x,p)$. From Eq. (20.235), we only need to add the ordering correction $V_\triangle(x,p)$ from Eq. (20.236) to the potential $V(x)$ in the propagator. Hence, from Eqs. (20.257), (20.258), and (20.259), or from Eq. (20.345), we have the propagator

$$\begin{aligned} K_{pgp}(x,t;x_0,t_0) &= [g(x)g(x_0)]^{-1/4}\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\qquad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\,\dot{x}_j^2 - V(\bar{x}_j) + \frac{\hbar^2(gg''-2g'^2)}{8mg^3}\right)\right]\right\} \\ &= [g(x)g(x_0)]^{-1/4}\int D\Big[x\sqrt{g(\bar{x})}\Big]\,\exp\left[\frac{i}{\hbar}\int_{t_0}^{t}dt\left(\frac{m}{2}g(\bar{x})\,\dot{x}^2 - V(\bar{x}) + \frac{\hbar^2(gg''-2g'^2)}{8mg^3}\right)\right] \end{aligned} \tag{20.363}$$
(midpoint path integral)

for the Hamiltonian (20.362) in midpoint form, in both discrete and continuous notations. To have a better-defined problem to normalize in Monte-Carlo form, we will calculated the partition function for the same Hamiltonian. Using Eq. (20.220), we can write the partition function as

$$\begin{aligned} Z_{pgp} &= \int dx_0\int dx\,\sqrt{g(x)}\,\delta(x-x_0)\,K_{pgp}(x,-i\hbar\beta;x_0,0) \\ &= \int dx_0\int dx\,\delta(x-x_0)\left(\frac{m}{i2\pi\hbar\delta t}\right)^{N/2}\int\left(\prod_{j=1}^{N-1}dx_j\right) \\ &\qquad\times\left\{\prod_{j=0}^{N-1}\sqrt{g(\bar{x}_j)}\,\exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2}g(\bar{x}_j)\,\dot{x}_j^2 - V(\bar{x}_j) + \frac{\hbar^2(gg''-2g'^2)}{8mg^3}\right)\right]\right\}, \end{aligned} \tag{20.364}$$

where we are still writing imaginary time in terms of t, rather than β, for the moment. Unfortunately, due to the presence of $g(\bar{x}_j)$, the normalization procedures of Sections 20.3.1–20.3.1.2 do not directly apply here (i.e., the integrals are not Gaussian). However, we can make things simpler by rescaling time on a path-dependent basis. That is, we change to time t', where

$$dt = dt'\,g[\bar{x}(t)]. \tag{20.365}$$
(temporal rescaling for path integral)

The path integral transformed thusly is

$$Z_{pgp} = \int dx_0 \int dx\, \delta(x-x_0) \left(\frac{m}{i2\pi\hbar}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \times \left\{ \prod_{j=0}^{N-1} \frac{1}{\sqrt{\delta t'_j}} \exp\left[\frac{i\delta t'_j}{\hbar}\left(\frac{m}{2}\dot{x}_j^2 - g(\bar{x}_j)\,V(\bar{x}_j) + \frac{\hbar^2(gg''-2g'^2)}{8mg^2}\right)\right]\right\}, \tag{20.366}$$

where the increments $\delta t'_j$ in the new variable are no longer uniform. This is still a source of inconvenience, so rather than choose the time slices to be uniformly spaced in t, we will choose them (again, in a path-dependent way) to be uniformly spaced in t', so that

$$Z_{pgp} = \int dx_0 \int dx\, \delta(x-x_0) \left(\frac{m}{i2\pi\hbar\delta t'}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \times \left\{ \prod_{j=0}^{N-1} \exp\left[\frac{i\delta t'}{\hbar}\left(\frac{m}{2}\dot{x}_j^2 - g(x_j)\,V(x_j) + \frac{\hbar^2(gg''-2g'^2)}{8mg^2}\right)\right]\right\}. \tag{20.367}$$

This has the form of a path integral in flat space, with a potential and effective potential modified by a factor of $g(x_j)$. Note that in the potentials, we have switched from the midpoint to the prepoint, since for a flat-space path integral (additive noise), we need not make such distinctions. Now switching from t' to $-i\hbar\beta$, we have

$$Z_{pgp} = \int dx_0 \int dx\, \delta(x-x_0) \left(\frac{m}{2\pi\hbar\delta\tilde{\beta}}\right)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \times \left\{ \prod_{j=0}^{N-1} \exp\left[-\frac{\delta\tilde{\beta}}{\hbar}\left(\frac{m}{2}(\partial_{\tilde{\beta}}x_j)^2 + g(x_j)\,V(x_j) - \frac{\hbar^2(gg''-2g'^2)}{8mg^2}\right)\right]\right\}, \tag{20.368}$$

in analogy to Eq. (20.133), and we are still using $\tilde{\beta} := \hbar\beta$. Then proceeding with the normalization as in Eq. (20.146), we can write down the normalized form

$$Z_{pgp} = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\left(g(x)\,V(x) - \frac{\hbar^2(gg''-2g'^2)}{8mg^2}\right)\right]\right\rangle\!\!\right\rangle_{x(\tilde{\beta})=x_0+\sqrt{\hbar/m}\,B_{\hbar\beta}(\tilde{\beta})}. \tag{20.369}$$

(normalized partition function)

Note that due to the time rescaling, the paths here are (scaled and translated) Brownian bridges, just as in a flat-space path integral.

In reviewing this derivation, an obvious question arises: why choose Weyl ordering? We could, of course, have picked any other ordering. Choosing symmetrized or product orderings would lead to the same result, as we have seen. However, Itō or anticipating orderings would have led to prepoint or postpoint path integrals, with different effective potentials. The time rescaling would proceed [but with, e.g., $g(x_j)$ instead of $g(\bar{x}_j)$], and once in flat space, the distinction between prepoint, midpoint, and postpoint is lost, and yet we are stuck with a different potential.

The resolution to this apparent problem is to note that it is *only* in a Stratonovich-type path integral that the normalization turns out to be simple. The reason is that, in the language of SDEs, it is only in the Stratonovich case with no drift that a temporal rescaling preserves the closure of loops (Brownian bridges). Itō SDEs, for example, do not in general preserve loop closure, even when equivalent to a driftless Stratonovich SDE (see Section 17.7.4.3). In terms of the path integral, this would mean that in any other ordering or expansion point, the temporal rescaling would upset the path closure, and thus there would be an extra normalization contribution due to the delta function that enforces the path closure. In any Stratonovich path integral, we don't have to worry about computing any such contribution.

20.4.4.1 A Similar Path Integral

As an additional example of a normalized path integral, consider the alternate ''Hamiltonian''

$$H_{\partial g\partial}(x,p) := -\hbar^2\partial_x \frac{1}{2mg(x)}\partial_x + V(x). \tag{20.370}$$
(variable-mass ''Hamiltonian'')

Superficially, this appears to be the same as the Hamiltonian (20.362) for which we just developed the normalized path integral. However, it is *not*, due to the space-dependent mass. Applying the momentum operator from Eq. (20.226), we readily see that

$$H_{\partial g\partial}(x,p) := g^{1/4}(x)\, p \frac{1}{2mg(x)} p\, \frac{1}{g^{1/4}(x)} + V(x),$$
(variable-mass ''Hamiltonian'') (20.371)

which is somewhat odd, as this Hamiltonian is not even Hermitian. Then rearranging the kinetic term,

$$\begin{aligned} g^{1/4}(x)\, p \frac{1}{g(x)} p\, \frac{1}{g^{1/4}(x)} &= p\frac{1}{g^{3/4}} p\, \frac{1}{g^{1/4}} + i\hbar \frac{g'}{g^{7/4}} p\, \frac{1}{4g^{1/4}} \\ &= p\frac{1}{g}p + i\hbar p \frac{g'}{4g^2} + i\hbar \frac{g'}{4g^{7/4}} p\, \frac{1}{g^{1/4}} \\ &= p\frac{1}{g}p + i\hbar p \frac{g'}{4g^2} + i\hbar \frac{g'}{4g^2} p - \hbar \frac{g'^2}{16g^3} \\ &= p\frac{1}{g}p + i\hbar p \frac{g'}{2g^2} - \frac{\hbar^2(gg''-2g'^2)}{4g^3} - \frac{\hbar^2 g'^2}{16g^3}. \end{aligned} \tag{20.372}$$

where again we used the commutator $[f(x),p] = i\hbar f'$, which still holds with the variable mass. Then relating this Hamiltonian to the original model Hamiltonian (20.362), we have

$$H_{\partial g\partial}(x,p) = H(x,p) + i\hbar p \frac{g'}{4mg^2} - \frac{\hbar^2(gg''-2g'^2)}{8mg^3} - \frac{\hbar^2 g'^2}{32mg^3}. \tag{20.373}$$

Relative to the propagator (20.363) for $H(x,p)$, the last two terms here become extra potential terms. The second term here leads to an extra velocity term. To handle this, we can revisit the development of the path integral in Section 20.4.3.1, where a similar potential term $-i\hbar p g'/2mg^2$ arose. From Eq. (20.320), we can see that this potential will contribute an extra potential $\hbar^2 g'^2/32mg^3$. From Eq. (20.321), we see that there is an additional potential $-\hbar^2(gg''-g'^2)/16mg^3$, as well as an overall factor $[g(x)/g(x_0)]^{1/4}$. Collecting these parts and adapting Eq. (20.363), we find the propagator

$$\begin{aligned} K_{\partial g\partial}(x,t;x_0,t_0) = \frac{1}{\sqrt{g(x_0)}} \Big(\frac{m}{i2\pi\hbar\delta t}\Big)^{N/2} \int \left(\prod_{j=1}^{N-1} dx_j\right) \\ \times \left\{ \prod_{j=0}^{N-1} \sqrt{g(\bar{x}_j)}\, \exp\left[\frac{i\delta t}{\hbar}\left(\frac{m}{2} g(\bar{x}_j)\, \dot{x}_j^2 - V(\bar{x}_j) + \frac{\hbar^2(2gg''-3g'^2)}{8mg^3}\right)\right] \right\}. \end{aligned}$$
(midpoint path integral) (20.374)

We can then adapt the derivation of the partition function (20.369), with the result

$$Z_{\partial g\partial} = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\left(g(x)\, V(x) - \frac{\hbar^2(2gg''-3g'^2)}{8mg^2}\right)\right] \right\rangle\!\!\right\rangle_{x(\tilde{\beta})=x_0+\sqrt{\hbar/m}\, B_{\hbar\beta}(\tilde{\beta})}.$$
(normalized partition function) (20.375)

Note that the effective potential is different than in the partition function (20.369). Note also that the path integral and effective potential are perfectly well-defined, despite the odd nature of the Hamiltonian (20.371).

20.5 Exercises

Problem 20.1

(a) Derive the propagator for the Hamiltonian (20.361)

$$H_{gpg} := \frac{1}{\sqrt{g(x)}} \frac{p^2}{2m} \frac{1}{\sqrt{g(x)}} + V(x). \tag{20.376}$$

(b) Convert the path integral from (a) to midpoint form.

(c) Use the path integral from (b) to derive a product-form path integral for the Laplace–Beltrami Hamiltonian $H_{\triangle}(x, p)$, and show that it is equivalent to the analogous path integral derived from Weyl and symmetrized orderings.

Chapter 21

Path-Integral Calculation of Casimir Energies

21.1 Scalar Theory

So far, in Chapters 13 and 14, we have developed two methods for computing atom–surface interactions: explicit mode summation with dipole-approximation atom–field coupling (with the assumption of perfectly conducting boundaries), and the more general Green-tensor formalism, which can handle dielectrics. Both of these approaches are difficult when the geometry of the surface is not highly symmetric, and it is useful to develop methods amenable to numerical computation in such cases.

Here we will develop the path-integral, or **world-line** method for computing more general Casimir forces (including forces between macroscopic bodies, not just the Casimir–Polder effect). The original formulation of this method[1] is for the scalar field, which is unphysical, but a simple way to introduce the method. The method amounts to a Monte–Carlo solution of a path integral, and is thus quite different from the approaches we have covered up to now. In evaluating the path integrals, we will make extensive use of the results on stochastic processes that we have developed in Chapter 17.

21.1.1 Quantum Scalar Field

We will take the scalar field $\phi(\mathbf{r})$ to be defined by the Lagrangian

$$L = \frac{\hbar^2}{2} \int d^d r \left[(\partial_t \phi)^2 - c^2 (\nabla \phi)^2 - \frac{m^2 c^4}{\hbar^2} \phi^2 \right] \tag{21.1}$$

in d spatial dimensions, which includes coupling to a mass m (for the analogous mass coupling in the electromagnetic field, see Problem 8.16). Note that we are in mks units, where ϕ^2 has dimensions of inverse volume, and $\hbar/mc$ is the Compton length. The Lagrangian thus has overall dimensions of square energy. As a model for interactions with material bodies, we can consider a space-dependent mass,

$$m(\mathbf{r}) = m_0 + \delta m(\mathbf{r}), \tag{21.2}$$

where m_0 is the ''asymptotic'' value of the mass function, representing the mass of the field, and the deviation $\delta m(\mathbf{r})$ models the effects of the material bodies. Then as

$$m^2(\mathbf{r}) = m_0^2 + 2 m_0 \delta m(\mathbf{r}) + \delta m^2(\mathbf{r}), \tag{21.3}$$

[1] Holger Gies, Kurt Langfeld, and Laurent Moyaerts, ''Casimir effect on the worldline,'' *Journal of High Energy Physics* **06**, 018 (2003) (doi: 10.1088/1126-6708/2003/06/018). For an overview of preceding work on world lines, see Christian Schubert, ''An Introduction to the Worldline Technique for Quantum Field Theory Calculations,'' *Acta Physica Polonica B* **27**, 3965 (1996), arXiv.org preprint (arXiv: hep-th/9610108v2); C. Schubert, ''Perturbative quantum field theory in the string-inspired formalism,'' *Physics Reports* **355**, 73 (2001) (doi: 10.1016/S0370-1573(01)00013-8), arXiv.org preprint (arXiv: hep-th/0101036v2).

we can define a background "potential" function $V(\mathbf{r})$ such that

$$V(\mathbf{r}) := \frac{1}{2}\Big[2m_0\delta m(\mathbf{r}) + \delta m^2(\mathbf{r})\Big]c^4, \tag{21.4}$$

which also has dimensions of squared energy. In this case, we can write the Lagrangian as

$$L = \frac{1}{2}\int d^3r \left[\hbar^2(\partial_t\phi)^2 - \hbar^2c^2(\nabla\phi)^2 - m_0^2c^4\phi^2 - 2V(\mathbf{r})\,\phi^2\right], \tag{21.5}$$

(model scalar-field Lagrangian)

Again, the potential will ultimately represent the macroscopic bodies of the Casimir effect, but there is no simple correspondence to the usual $\epsilon(\mathbf{r})$ of the electromagnetic field.

The conjugate momentum to ϕ is

$$\pi(\mathbf{r}) := \frac{\delta L}{\delta\dot{\phi}} = \hbar^2\partial_t\phi(\mathbf{r}). \tag{21.6}$$

Thus, the Hamiltonian is

$$\begin{aligned} H &= \pi\dot{\phi} - L \\ &= \frac{1}{2}\int d^3r \left[\frac{\pi^2}{\hbar^2} + \hbar^2c^2(\nabla\phi)^2 + m_0^2c^4\phi^2 + 2V(\mathbf{r})\,\phi^2\right]. \end{aligned} \tag{21.7}$$

Hamilton's equations give the first-order-in-time evolution in terms of the two coupled fields ϕ and π,

$$\begin{aligned} \partial_t\phi &= \pi/\hbar^2 \\ \partial_t\pi &= \hbar^2c^2\nabla^2\phi - m_0^2c^4\phi - 2V(\mathbf{r})\,\phi, \end{aligned} \tag{21.8}$$

while after decoupling these equations, we find the inhomogeneous wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\partial_t^2\right)\phi = \left(\frac{m_0^2c^2}{\hbar^2} + \frac{2}{\hbar^2c^2}V(\mathbf{r})\right)\phi \tag{21.9}$$

(scalar-field wave equation)

for uncoupled but second-order-in-time evolution.

In quantum mechanics, we simply regard the conjugate $\pi(\mathbf{r})$ and $\phi(\mathbf{r})$ fields to be operators. That is, envision a harmonic oscillator at each $\mathbf{r}$ representing the amplitude of the field there. Then the operator $\phi(\mathbf{r})$ acts as the operator $x \propto (a + a^\dagger)$ for that particular oscillator (but note that this x refers to the *temporal* part of the field, giving the local, time-dependent amplitude). Correspondingly, $\pi(\mathbf{r})$ acts as the local conjugate (quadrature) operator $p \propto (a - a^\dagger)$ at that point in space. (More on this in Section 21.1.4.2.)

Normally we will take the background mass m_0 to be zero for a massless scalar field, but we can note here that we can simply absorb any nonzero mass into an overall constant offset in $V(\mathbf{r})$. Henceforth, we will not write an explicit mass m_0, although the treatment here can cover massive scalar fields as well.

21.1.2 Partition Function

We ultimately want the energy associated with the various configurations of $V(\mathbf{r})$, which is specified by the Hamiltonian. A general way to get at the energy is to consider the partition function,

$$Z := \mathrm{Tr}\left[e^{-\beta H}\right]. \tag{21.10}$$

where $\beta := 1/k_{\mathrm{B}}T$ for temperature T and Boltzmann constant k_{B}. Expressing this in terms of the basis $|n\rangle$ of energy eigenstates (of energy E_n),

$$Z = \sum_n \langle n|e^{-\beta H}|n\rangle = \sum_n e^{-\beta E_n}. \tag{21.11}$$

Recall that according to the Boltzmann distribution in statistical mechanics, each term here represents the *relative* probability of occupying state $|n\rangle$, and the partition function is the sum of these relative probabilities, and thus gives the proper probability normalization. That is, the probability of occupying state $|n\rangle$ is $e^{-\beta E_n}/Z$. Then the energy expectation value is

$$E \equiv \langle H \rangle = \sum_n \frac{e^{-\beta E_n}}{Z} E_n = -\frac{1}{Z}\partial_\beta \sum_n e^{-\beta E_n} = -\frac{1}{Z}\partial_\beta Z \tag{21.12}$$

or

$$E = -\partial_\beta \log Z. \tag{21.13}$$
(energy in terms of partition function)

Thus, we will seek to compute $\log Z$, and thereby compute the Casimir energy.

Generally speaking, the Casimir energy is the energy associated with the vacuum (i.e., the effect is due not to real photons, but rather only to virtual photons). In the zero-temperature limit, the vacuum is occupied with unit probability, which we can see from the partition function, where the ground-state-energy term dominates the other ones as $\beta \longrightarrow \infty$:

$$\lim_{\beta\to\infty} Z = e^{-\beta E_0}. \tag{21.14}$$

Thus, differentiating the partition function directly yields the ground-state energy. Of course, we can also derive temperature-dependent corrections to the ground-state energy with this method.

21.1.2.1 Free Energy

When we consider finite-temperature effects (e.g., in Section 21.1.6), rather than considering the energy E, we will need to compute the **Helmholtz free energy**

$$\mathcal{F} := E - TS, \tag{21.15}$$
(Helmholtz free energy)

where T is the temperature and S is the entropy. To see why, recall the first law of thermodynamics,

$$dE = dQ - dW, \tag{21.16}$$

which gives the small change in internal energy E in terms of the a small heat dQ transferred to the system and the work dW done by the system (commonly written $P\,dV$ in terms of pressure and volume). Since the entropy change is $dS = dQ/T$, we can rewrite the first law as

$$-dW = dE - T\,dS = d\mathcal{F}, \tag{21.17}$$

where we assume the differential dF to be at constant temperature. Thus, at constant temperature, we can interpret the free energy F as the relevant Casimir potential, as its change represents the work involved in rearranging the material bodies. (We use $-dW$ here as the work done *on* the system by rearranging the bodies, rather than the work dW done *by* the system.) If we regard the work as being due to moving one of the bodies through a small displacement $d\mathbf{r}$ against a force $\mathbf{F}$, we have

$$dW = \mathbf{F} \cdot d\mathbf{r}, \tag{21.18}$$

and thus

$$\mathbf{F} = -\nabla\mathcal{F}. \tag{21.19}$$
(force and free energy)

That is, the thermodynamic force on the body (at constant temperature) is given by the gradient of the free energy, taken with respect to a coordinate that represents the position of the body.

To put F in terms of the partition function, we can begin with

$$E = \mathcal{F} + TS = \mathcal{F} - T\frac{\partial \mathcal{F}}{\partial T}, \tag{21.20}$$

where the temperature derivative is taken with no work (e.g., constant volume). Then

$$E = -T^2 \partial_T \left(\frac{\mathcal{F}}{T} \right), \tag{21.21}$$

and using Eq. (21.13), with $(-1/kT^2)\partial_\beta = \partial_T$,

$$\mathcal{F} = -k_{\rm B} T \log Z = -\frac{1}{\beta} \log Z. \tag{21.22}$$

(free energy in terms of partition function)

Note that in the zero-temperature limit, we can see from Eq. (21.14) that

$$\lim_{\beta \to \infty} \log Z = -\beta E_0, \tag{21.23}$$

and thus for zero temperature, both $\mathcal{F}$ and E give the ground-state energy E_0. That E and $\mathcal{F}$ coincide at $T = 0$ is also obvious from the definition (21.15).

21.1.2.2 Partition-Function Renormalization

As we have seen, the Casimir energy is divergent, and must be renormalized by subtracting away the energy without coupling between the bodies. We will revisit this issue later when discussing particular examples, but for now we note that we will need a subtraction of the form

$$E - E_0 = -\partial_\beta \log \frac{Z}{Z_0}, \tag{21.24}$$

(renormalized Casimir energy)

where E_0 represents the field, for example, with $V(\mathbf{r}) = 0$ or any other appropriate reference configuration of the field. The main point here is that overall factors in the partition function (such as normalization factors) will cancel in renormalization, and so we are free to drop such factors along the way without changing the final result. Of course, a similar renormalization applies to the free energy in the case of nonzero temperature.

21.1.3 Path Integral of the Quantum Field

Now it is convenient to introduce a basis of eigenstates of the field operator ϕ, which must sum to the identity:

$$\int d\phi \, |\phi\rangle\langle\phi| = 1. \tag{21.25}$$

The labels $\phi(\mathbf{r})$ for these states correspond to something like the set of all classical fields of a localized excitation, of all possible amplitudes. We can then write the partition-function trace in (21.10) in this basis:

$$Z = \int d\phi \, \langle\phi| e^{-\beta H} |\phi\rangle. \tag{21.26}$$

Note that in the notation here, a volume integration is implied,

$$\int d\phi \equiv \int d^d r \, d\phi(\mathbf{r}), \tag{21.27}$$

where d is the number of spatial dimensions (i.e., normally $d = 3$), to obtain a scalar integration result. Now if we split β into N pieces $\Delta\beta := \beta/N$, we can insert identities in between each of the split exponential factors:

$$\begin{aligned} Z &= \int d\phi \, \langle\phi| e^{-\Delta\beta H} \left[\int d\phi_{N-1} |\phi_{N-1}\rangle\langle\phi_{N-1}| \right] \cdots e^{-\Delta\beta H} \left[\int d\phi_2 |\phi_2\rangle\langle\phi_2| \right] e^{-\Delta\beta H} \left[\int d\phi_1 |\phi_1\rangle\langle\phi_1| \right] e^{-\Delta\beta H} |\phi\rangle \\ &= \int d\phi \, \langle\phi| \left\{ \prod_{j=1}^{N-1} \int d\phi_j \left[e^{-\Delta\beta H} |\phi_j\rangle\langle\phi_j| \right] \right\} e^{-\Delta\beta H} |\phi\rangle. \end{aligned} \tag{21.28}$$

This is the same procedure that we used for the quantum-mechanical path integral in Section 20.1.1, but generalized to the quantum-field case. Then if we take

$$\phi_0 \equiv \phi_N \equiv \phi, \tag{21.29}$$

in analogy to the endpoint identification (20.128), we can rewrite the partition function as

$$Z = \int \prod_{j=1}^{N} d\phi_j \, \langle \phi_j | e^{-\Delta\beta H} | \phi_{j-1} \rangle. \tag{21.30}$$

Introducing the shorthand for the path-integral volume element

$$D\phi := \prod_{j=1}^{N} d\phi_j = \prod_{j=0}^{N} d\phi_j \, \delta(\phi_N - \phi_0), \tag{21.31}$$
(field path-integration element)

we then have

$$Z = \int D\phi \prod_{j=1}^{N} \langle \phi_j | e^{-\Delta\beta H} | \phi_{j-1} \rangle, \tag{21.32}$$
(field path integral)

where the path integral is periodic due to the identification $|\phi_N\rangle \equiv |\phi_0\rangle$.

21.1.3.1 Momentum Projectors and Imaginary Time

The path integral (21.32) that we have written down so far can be interpreted as something like the propagation of the field state through many small increments $\Delta\beta$ in imaginary time. This observation motivates the next step in the calculation. An exponential evolution operator generates *first*-order evolution in time, but the scalar-field evolution, as we have seen, is *second*-order in time, or first-order if we consider coupling to the conjugate π field. This is also a sensible step since the Hamiltonian, and thus the evolution operator, involves both fields. Now we write out the evolution operator explicitly using Eq. (21.7) (with $m_0 = 0$):

$$e^{-\Delta\beta H} = \exp\left[-\frac{\Delta\beta}{2} \int d^3r \, \frac{\pi^2}{\hbar^2} \right] \exp\left[-\frac{\Delta\beta}{2} \int d^3r \Big[\hbar^2 c^2 (\nabla\phi)^2 + 2V(\mathbf{r}) \, \phi^2 \Big] \right] + O(\Delta\beta^2). \tag{21.33}$$

Notice that we have split the exponential into the conjugate parts; we are assuming the limit of vanishingly small $\Delta\beta$, so we can ignore the higher-order corrections here. Concentrating on a single factor in Eq. (21.32), we can insert the identity in terms of the basis of momentum-field eigenstates $|\pi\rangle$ for each such factor:

$$\begin{aligned}
&\langle \phi_j | e^{-\Delta\beta H} | \phi_{j-1} \rangle \\
&\quad = \int d\pi_j \langle \phi_j | \pi_j \rangle \langle \pi_j | e^{-\Delta\beta H} | \phi_{j-1} \rangle \\
&\quad = \int d\pi_j \langle \phi_j | \pi_j \rangle \langle \pi_j | \exp\left[-\frac{\Delta\beta}{2\hbar^2} \int d^3r \, \pi^2 \right] \exp\left[-\frac{\Delta\beta}{2} \int d^3r \Big[\hbar^2 c^2 (\nabla\phi)^2 + 2V(\mathbf{r}) \, \phi^2 \Big] \right] | \phi_{j-1} \rangle \\
&\quad = \int d\pi_j \langle \phi_j | \pi_j \rangle \langle \pi_j | \phi_{j-1} \rangle \exp\left[-\frac{\Delta\beta}{2\hbar^2} \int d^3r \, \pi_j^2 \right] \exp\left[-\frac{\Delta\beta}{2} \int d^3r \Big[\hbar^2 c^2 (\nabla\phi_{j-1})^2 + 2V(\mathbf{r}) \, \phi_{j-1}^2 \Big] \right].
\end{aligned} \tag{21.34}$$

Note that the ϕ_{j-1} in the potential term could be just as written in terms of any combination of ϕ_j and ϕ_{j-1}. The inner products correspond to inner products of conjugate-variable eigenstates, e.g., $\langle x | p \rangle = e^{ipx/\hbar} / \sqrt{2\pi\hbar}$, so

$$\begin{aligned}
&\langle \phi_j | e^{-\Delta\beta H} | \phi_{j-1} \rangle \\
&\propto \int d\pi_j \exp\left[\int d^3r \left(\frac{i}{\hbar} \pi_j (\phi_j - \phi_{j-1}) - \frac{\Delta\beta}{2\hbar^2} \pi_j^2 \right) \right] \exp\left[-\frac{\Delta\beta}{2} \int d^3r \Big[\hbar^2 c^2 (\nabla\phi_{j-1})^2 + 2V(\mathbf{r}) \, \phi_{j-1}^2 \Big] \right].
\end{aligned} \tag{21.35}$$

Focusing on the momentum-dependent factor, we can carry out the momentum integral explicitly. First we complete the square,

$$\begin{aligned}\int d\pi_j \exp\left[\int d^3r \left(\frac{i}{\hbar}\pi_j(\phi_j-\phi_{j-1}) - \frac{\Delta\beta}{2\epsilon_0}\pi_j^2\right)\right] \\ = \int d\pi_j \exp\left[-\frac{\Delta\beta}{2\hbar^2}\int d^3r\left(\pi_j - i\frac{\hbar}{\Delta\beta}(\phi_j-\phi_{j-1})\right)^2\right]\exp\left[-\frac{1}{2}\int d^3r\frac{(\phi_j-\phi_{j-1})^2}{\Delta\beta}\right],\end{aligned} \tag{21.36}$$

again keeping only terms to order $\Delta\beta$ in splitting the exponential. We can evaluate the Gaussian integral in analogy to

$$\int_{-\infty+i\eta}^{\infty+i\eta} \exp\left(-\alpha x^2\right)\, dx = \int_{-\infty}^{\infty}\exp\left(-\alpha x^2\right)\, dx = \sqrt{\frac{\pi}{\alpha}} \qquad (\mathrm{Re}[\alpha]>0), \tag{21.37}$$

since shifting the contour in the imaginary direction in the complex plane does not affect the integral's value (due to analyticity of the Gaussian). The π_j integral thus gives an overall factor that is independent of the fields, so we can drop it from the partition function. We can return the remaining quadratic factor, merging it into the remaining exponential in Eq. (21.35):

$$\langle\phi_j|e^{-\Delta\beta H}|\phi_{j-1}\rangle \propto \exp\left[-\frac{\Delta\beta}{2}\int d^3r\left(\frac{(\phi_j-\phi_{j-1})^2}{\Delta\beta^2} + \hbar^2c^2(\nabla\phi_{j-1})^2 + 2V(\mathbf{r})\,\phi_{j-1}^2\right)\right]. \tag{21.38}$$

Now the partition function (21.32) contains the product of many such factors. Taking the limit $\Delta\beta \longrightarrow d\beta$, and taking $\tilde{\beta} := \hbar c\beta$ to be the "Boltzmann length,"

$$Z = \int D\phi\, \exp\left[-\frac{1}{2\hbar c}\int d\tilde{\beta}\, d^3r\left(\hbar^2c^2(\partial_{\tilde{\beta}}\phi)^2 + \hbar^2c^2(\nabla\phi)^2 + 2V(\mathbf{r})\,\phi^2\right)\right].$$

(field path integral, β-integral form) (21.39)

Here, we have defined the derivative corresponding to the $(j-1)$ field:

$$\partial_{\tilde{\beta}}\phi_{j-1} := \frac{\phi_j - \phi_{j-1}}{\Delta\tilde{\beta}}. \tag{21.40}$$

There is no other dependence on ϕ in the kinetic term, so there is no concern with ordering or a "space"-dependent mass.

Note that the partition function now has the form

$$Z = \int D\phi\, \exp\left[\frac{i}{\hbar}\int dt\, L(\phi,\partial_t\phi)\right]_{t\longrightarrow -i\tilde{\beta}/c}, \tag{21.41}$$

where L is the Lagrangian (21.5). This has the form of the (coherent) path-integral propagator, but with the imaginary-time replacement $t \to -i\tilde{\beta}/c = -i\hbar\beta$, which converts the unitary-evolution integral into a diffusive path integral. This is the **Wick rotation** of the path integral, referring to the effective change from a Minkowski to Euclidian metric via the change in minus sign of the temporal term in the quadratic form of the Lagrangian.

21.1.4 Reduction of the Hilbert Space

Because we are dealing with a linear system (i.e., a quadratic Hamiltonian), we can simplify the above path integral greatly. This is fortunate, as a path integral with field integration variables is not necessarily an easy thing to handle.

21.1.4.1 Evaluation of the Gaussian Functional Integral

First, note that we have an exponentiated quadratic form in the partition function,

$$Z = \int D\phi\, \exp\left[-\frac{1}{2\hbar c}\int d\tilde{\beta}\, d^3r\,\phi(\mathbf{r},\tilde{\beta})\left(-\hbar^2c^2\partial_{\tilde{\beta}}^2 - \hbar^2c^2\nabla^2 + 2V(\mathbf{r})\right)\phi(\mathbf{r},\tilde{\beta})\right], \tag{21.42}$$

where we have integrated by parts in the derivative terms. We thus have a (functional) Gaussian integral, which we can always do. To do this, we note that the functional integral is the extension of the finite-dimensional Gaussian integral,

$$\int dz_1 \cdots dz_N \, \exp\left[-\frac{1}{2} z_\alpha \left(S^{-1}\right)_{\alpha\beta} z_\beta\right] = (2\pi)^{N/2} \sqrt{\det(S_{\alpha\beta})}. \tag{21.43}$$

Here, $S_{\alpha\beta} = \langle z_\alpha z_\beta \rangle$, where the expectation value is taken with respect to the normalized Gaussian function, and we can always assume the matrix in such a quadratic form to be symmetric. Thus, it is diagonalized by an orthogonal transformation, and making this transformation, the integral is of the form

$$\int dz_1 \cdots dz_N \, \exp\left[-\frac{1}{2} \sum_\alpha \frac{z_\alpha^2}{\sigma_\alpha^2}\right] = (2\pi)^{N/2} \prod_{\alpha=1}^{N} \sigma_\alpha, \tag{21.44}$$

where the σ_α^2 are the eigenvalues of $S_{\alpha\beta}$, and we have carried out the N independent Gaussian integrals. The eigenvalue product then just becomes the determinant.

Applying this argument to the functional case, we drop the constant factors and thus write

$$Z = \sqrt{\det\left(c^2\left(-\hbar^2\partial_{\tilde{\beta}}^2 - \hbar^2\nabla^2\right) + 2V\right)^{-1}}, \tag{21.45}$$

or

$$\log Z = -\frac{1}{2} \log\det\left(c^2\left(-\hbar^2\partial_{\tilde{\beta}}^2 - \hbar^2\nabla^2\right) + 2V\right). \tag{21.46}$$

Also noting that for any symmetric matrix $\mathbf{A}$,

$$\log\det \mathbf{A} = \mathrm{Tr} \log \mathbf{A}, \tag{21.47}$$

(which is clear in the diagonal case, where the log of the eigenvalue product is the sum of the eigenvalue logarithms, and follows in the general case by an orthogonal transformation), we can finally write

$$\log Z = -\frac{1}{2} \mathrm{Tr} \log\left(c^2\left(-\hbar^2\partial_{\tilde{\beta}}^2 - \hbar^2\nabla^2\right) + 2V\right).$$

(trace-log form of partition function) (21.48)

Thus we have resolved the functional integral into an evaluation of the trace (or equivalently, determinant) of an operator. Recall that in dropping overall factors from Z, this expression is only meaningful in the context of renormalization. Note that it is common in field theory to use units where $\hbar = c = 1$, and not to bother with the factor of 2 in the potential term, where we would find

$$\log Z = -\frac{1}{2} \mathrm{Tr} \log\left(-\partial_{\tilde{\beta}}^2 - \nabla^2 + V\right). \tag{21.49}$$

We would obtain this same form by removing a factor $\hbar^2 c^2$ from the log, and absorbing the factor $2/\hbar^2 c^2$ into the potential (whereupon it would gain the dimensions of inverse squared length). However, we will keep the units as written in (21.48) to better connect to the relativistic particle, and the utility of the factor of 2 will become apparent soon.

Now, observe that the operator in Eq. (21.48) is just the (scaled) operator for the wave equation (21.9),

$$\left(\frac{\partial_t^2}{c^2} - \nabla^2 + \frac{2}{\hbar^2 c^2} V(\mathbf{r})\right) \phi(\mathbf{r}, t) = 0, \tag{21.50}$$

but under the same thermal-time replacement $t \longrightarrow i\tilde{\beta}/c = i\hbar\beta$:

$$\left(-\partial_{\tilde{\beta}}^2 - \nabla^2 + \frac{2}{\hbar^2 c^2} V(\mathbf{r})\right) \phi(\mathbf{r}, \tilde{\beta}) = 0. \tag{21.51}$$

The (classical) Green function $G(\mathbf{r},t;\mathbf{r}',t')$ for the scalar field is defined to be the solution of the wave equation corresponding to a perturbation localized to the spacetime point $(\mathbf{r}',t')$:

$$\left(\hbar^2\partial_t^2 - \hbar^2c^2\nabla^2 + 2V\right)G(\mathbf{r},t;\mathbf{r}',t') = \delta^3(\mathbf{r}-\mathbf{r}')\delta(t-t').$$

(scalar-field Green function) (21.52)

Note that G can be defined up to an arbitrary factor, and we have made one particular choice here. We can think of $G(\mathbf{r},t;\mathbf{r}',t')$ as a matrix element of an operator G, expressed in a basis of space-time states $|\mathbf{r},t\rangle$, so that the defining equation for the Green operator is

$$\left(\hbar^2\partial_t^2 - \hbar^2c^2\nabla^2 + 2V\right)G = 1, \tag{21.53}$$

or

$$G = \frac{1}{\hbar^2\partial_t^2 - \hbar^2c^2\nabla^2 + 2V}. \tag{21.54}$$

Thus, in terms of the Green function, the partition function (21.48) can be written

$$\log Z = \frac{1}{2}\mathrm{Tr}\log\tilde{G},$$

(Green-function form of partition function) (21.55)

where

$$\tilde{G} = G|_{t\longrightarrow i\tilde{\beta}/c} \tag{21.56}$$

is the imaginary-time Green operator.

21.1.4.2 Digression: Second Quantization and Mode Summation

In computing Casimir energies, at zero temperature we are just summing the zero-point energies of all the field modes, or rather the differences in the zero-point energies when comparing different boundary configurations. Here we will connect our expressions thus far to the mode-summation picture to gain some intuition into what we are doing.

To begin, though, we will be more precise about the quantum field modes by explicitly quantizing the scalar field. The time and space dependence of the scalar wave equation (21.9) are separable if we take

$$\phi(\mathbf{r},t) = \left[\phi(\mathbf{r})\,e^{-i\omega t} + \text{H.c.}\right], \tag{21.57}$$

which gives the Helmholtz-like equation

$$\left[-\nabla^2 + \frac{2}{\hbar^2c^2}V(\mathbf{r})\right]f_j(\mathbf{r}) = \frac{\omega_j^2}{c^2}f_j(\mathbf{r}), \tag{21.58}$$

where we use j as a mode index, the eigenvalue of the jth mode is ω_j^2/c^2, and we label the mode functions by f_j, normalized so that

$$\int d^3r\,|f_j(\mathbf{r})|^2 = 1. \tag{21.59}$$

This is the normal-mode decomposition of the scalar field. We can quantize the normal modes by changing the time dependence $e^{-i\omega t}$ of the mode to an annihilation operator, so that we obtain the quantized normal-mode operators

$$\phi_j = \frac{1}{\sqrt{2\hbar\omega_j}}\left[f_j(\mathbf{r})\,a_j + \text{H.c.}\right], \tag{21.60}$$

(quantum normal-mode fields)

where we have chosen the overall constant to make the Hamiltonian come out nicely below. The quantized conjugate normal-mode operators are given by (21.6) as

$$\pi_j = -i\hbar\sqrt{\frac{\hbar\omega_j}{2}}\left[f_j(\mathbf{r})\,a_j - \text{H.c.}\right],$$

(quantum normal-mode momentum fields) (21.61)

with second-quantized field operators

$$\phi = \sum_j \phi_j, \qquad \pi = \sum_j \pi_j. \tag{21.62}$$
(quantized field operators)

Then noting that the eigenvalue equation (21.58) implies

$$\phi_j \left[-\nabla^2 + \frac{2}{\hbar^2 c^2} V(\mathbf{r}) \right] \phi_j = \frac{\omega_j^2}{c^2} \phi_j^2, \tag{21.63}$$

we can put the mode operators (21.62) into the Hamiltonian (21.7) to obtain

$$\begin{aligned} H &= \frac{1}{2} \int d^d r \left[\frac{\pi^2}{\hbar^2} + \phi \Big(-\hbar^2 c^2 \nabla^2 + 2V(\mathbf{r}) \Big) \phi \right] \\ &= \frac{1}{2} \sum_j \int d^d r \left[\frac{\pi_j^2}{\hbar^2} + \hbar^2 c^2 \phi_j \left(-\nabla^2 + \frac{2}{\hbar^2 c^2} V^2(\mathbf{r}) \right) \phi_j \right] \\ &= \frac{1}{2} \sum_j \int d^d r \left[\frac{\pi_j^2}{\hbar^2} + \hbar^2 \omega_j^2 \phi_j^2 \right] \\ &= \frac{1}{2} \sum_j \int d^d r \, \hbar\omega_j \left[a_j^\dagger a_j + a_j a_j^\dagger \right] |f_j(\mathbf{r})|^2 \\ &= \frac{1}{2} \sum_j \hbar\omega_j \left[2 a_j^\dagger a_j + 1 \right], \end{aligned} \tag{21.64}$$

or finally,

$$H = \sum_j \hbar\omega_j \left[a_j^\dagger a_j + \frac{1}{2} \right], \tag{21.65}$$

which is a sum of harmonic oscillators for each normal mode. In deriving this, we used the fact that only terms like $a_j^\dagger a_j$ and $a_j a_j^\dagger$ could contribute to the Hamiltonian, simplifying the algebra, and that $[a, a^\dagger] = 1$ for a harmonic oscillator. In particular, at zero temperature, the field is in the vacuum state, and the field energy is

$$E = \langle 0|H|0\rangle = \sum_j \frac{\hbar\omega_j}{2}, \tag{21.66}$$
(mode summation for vacuum energy)

which is the sum of all zero-point energies of the (classical) normal modes. (Again, this quantity is obviously divergent, and must be regularized by computing the difference in this quantity between two configurations.) Alternately, we can leave the spatial integration in (21.64), to write the Hamiltonian as

$$H = \sum_j \int d^d r \, |f_j(\mathbf{r})|^2 \, \hbar\omega_j \left[a_j^\dagger a_j + \frac{1}{2} \right], \tag{21.67}$$

where the integrand represents an energy *density* of the field. We can define an energy density at zero temperature as the vacuum expectation value of the integrand, or

$$\mathscr{E}(\mathbf{r}) = \sum_j \frac{\hbar\omega_j}{2} |f_j(\mathbf{r})|^2, \tag{21.68}$$
(mode summation for vacuum energy density)

which is again a sum over zero-point energies, weighted by the local value of the corresponding (squared) normalized mode function. The energy density here can provide an intuitive way to think about Casimir–Polder effects, where an atom ''samples'' the local energy density of the field, which has some spatial dependence due to the combination of all the mode functions.

21.1.4.3 Digression Part II: Mode Summation of the Functional Determinant

Recall from Eq. (21.45) or (21.55) that we have a partition function that is basically the determinant of the Green operator:

$$Z = \sqrt{\det \tilde{G}}. \tag{21.69}$$

In terms of the eigenvalues λ_j of $\tilde{G}$, this becomes

$$Z = \sqrt{\prod_j \lambda_j}, \tag{21.70}$$

and then the logarithm changes this into an eigenvalue sum:

$$\log Z = \frac{1}{2} \sum_j \log \lambda_j. \tag{21.71}$$

In the zero-temperature limit, the zero-point energy is proportional to $\log Z$, as we see from Eq. (21.14). However, here we have a sum of *logarithms* of eigenvalues, while the energy mode sum (21.66) is a sum of (square roots of) eigenvalues, without the same logarithm. It is not so obvious that we are still doing the same mode summation, so we will work out the determinant directly in terms of its eigenvalues to recover the former mode sum.

The key here, of course, is that the eigenvalues ω_j of the wave operator are *not* the same as the eigenvalues of the Green operator. We will work with the inverse Green operator in imaginary time,

$$\tilde{G}^{-1} \propto -\partial_{\tilde{\beta}}^2 - \nabla^2 + \frac{2}{\hbar^2 c^2} V, \tag{21.72}$$

which has the form of the wave operator in Eq. (21.58), but with the extra, imaginary-time dimension $\tilde{\beta}$. This is a free, scalar wave in this dimension (in the sense of not coupling to the potential V), but is bounded in extent from 0 to $\hbar\beta$, with periodic boundary conditions, as we will justify in more detail in Section 21.1.6. Thus, denoting the eigenvalues of $\tilde{G}$ by $\lambda_{n,j}$, we can write (up to a constant factor)

$$\frac{c^2}{\lambda_{n,j}} = \left(\frac{2\pi n}{\hbar\beta}\right)^2 + \omega_j^2, \tag{21.73}$$

where the first term on the right-hand side gives the squared frequency that respects the periodic boundary conditions (where $n \in \mathbb{Z}$), and the second term comes from Eq. (21.58). Then using these eigenvalues in Eq. (21.70), we have

$$Z = \prod_{n=-\infty}^{\infty} \prod_j \left[\left(\frac{2\pi n}{\hbar\beta}\right)^2 + \omega_j^2\right]^{-1/2}, \tag{21.74}$$

where we are dropping overall factors of c. Noting that the function is even in n, we can combine the negative and positive n in the product to find

$$Z = \prod_j \frac{1}{\omega_j} \prod_{n=1}^{\infty} \left[\left(\frac{2\pi n}{\hbar\beta}\right)^2 + \omega_j^2\right]^{-1}. \tag{21.75}$$

In working with the eigenvalues of the Green operator (21.72) with the extra dimension, we expect there to be many more eigenvalues than we had from the wave operator, an expectation that is of course confirmed by the extra mode index n here. Therefore, this quantity should be even *more* divergent than the mode sum we seek. We will be careful to normalize against the free extra dimension, or in other words we will normalize the partition function by the partition function of the inverse Green operator $-\partial_{\tilde{\beta}}^2$ to mitigate the

effects of the extra degree of freedom here. Thus, we will work with

$$\frac{Z}{Z(-\partial_{\tilde{\beta}}^2)} = \prod_j \frac{1}{\omega_j} \prod_{n=1}^{\infty} \frac{\left(\frac{2\pi n}{\hbar\beta}\right)^2}{\left(\frac{2\pi n}{\hbar\beta}\right)^2 + \omega_j^2} = \prod_j \left[\omega_j \prod_{n=1}^{\infty} 1 + \left(\frac{\hbar\omega_j\beta}{2\pi n}\right)^2\right]^{-1}, \tag{21.76}$$

noting that we exclude the zero-eigenvalue mode that would otherwise cause problems in the sum. Using the product identity

$$\sinh z = z \prod_{j=1}^{\infty} \left[1 + \left(\frac{z}{\pi j}\right)^2\right], \tag{21.77}$$

we find

$$\frac{Z}{Z(-\partial_{\tilde{\beta}}^2)} = \prod_j \left[\frac{2}{\hbar\beta} \sinh\left(\frac{\hbar\omega_j\beta}{2}\right)\right]^{-1}. \tag{21.78}$$

Then using $1/\sinh x = 2/(e^x - e^{-x}) = 2e^{-x}/(1 - e^{-2x})$, the partition function becomes

$$\frac{Z}{Z(-\partial_{\tilde{\beta}}^2)} = \prod_j \hbar\beta \left[\frac{e^{-\hbar\omega_j\beta/2}}{1 - e^{-\hbar\omega_j\beta}}\right]. \tag{21.79}$$

The bracketed quantity here is the standard form of the partition function for a quantum harmonic oscillator of frequency ω_j, and the overall factors of $\hbar\beta$ correspond to overall energy offsets that are discarded in renormalization. Dropping factors of $2/\hbar\beta$ and computing the log of the partition function, we find

$$\log \frac{Z}{Z(-\partial_{\tilde{\beta}}^2)} = -\sum_j \log \sinh\left(\frac{\hbar\omega_j\beta}{2}\right). \tag{21.80}$$

Using $\partial_x \log\sinh x = \coth x$, we can then differentiate the log of the partition function to find the energy:

$$E = -\partial_\beta \log \frac{Z}{Z(-\partial_{\tilde{\beta}}^2)} = \sum_j \frac{\hbar\omega_j}{2} \coth\left(\frac{\hbar\omega_j\beta}{2}\right). \tag{21.81}$$

(mode-summation energy, thermal state)

In the zero-temperature limit, we use $\lim_{x\to\infty} \coth x = 1$ to obtain

$$E_0 = \sum_j \frac{\hbar\omega_j}{2}, \tag{21.82}$$

which is the same mode sum that we obtained in Eq. (21.66). Eq. (21.81) then gives the generalization of this mode-sum energy for finite temperatures.

21.1.4.4 Integral Representation of the Logarithm

Now to continue with the path-integral construction, we implement a transformation[2] that allows us to transform the partition function into another path integral. We can write the integral identity

$$\tilde{G} = \int_0^\infty d\mathcal{T} \exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right) \tag{21.83}$$

[2]This transformation was used, e.g., by Julian Schwinger, "On Gauge Invariance and Vacuum Polarization," *Physical Review* **82**, 664 (1951) (doi: 10.1103/PhysRev.82.664). Thus, $\mathcal{T}$ is often called the **Schwinger proper time** or the **Fock–Schwinger proper time**, after the earlier work of V. Fock, *Proper time in classical and quantum mechanics Physikalische Zeitschrift der Sowjetunion* **12**, 404 (1937), reprinted in *Selected Works: V. A. Fock*, L. D. Faddev, L. A. Khalfin, and I. V. Komarov, Eds. (Chapman & Hall, 2004), p. 421 (doi: 10.1201/9780203643204.ch10b).

for the Green operator, and then consider the variation of the inverse, $\delta(\tilde{G}^{-1})$, using this integral form:

$$\delta(\tilde{G}^{-1})\tilde{G} = \delta(\tilde{G}^{-1}) \int_0^\infty d\mathcal{T}\, \exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right) = -\delta\left[\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right)\right]. \tag{21.84}$$

It is worth emphasizing here that we are regarding $\tilde{G}^{-1}$ as the "independent variable" here of the variation. The left-hand side is the same as $\delta \log(\tilde{G}^{-1}) = -\delta \log \tilde{G}$, so removing the variations, we have

$$\log \tilde{G} = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right). \tag{21.85}$$

The partition function (21.55) then becomes

$$\log Z = \frac{1}{2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \mathrm{Tr}\left[\exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right)\right]. \tag{21.86}$$

Technically, this integral is divergent at $\mathcal{T} = 0$, but this divergence is cured if we consider any *difference* between two such integrals, as we can see:

$$\lim_{\epsilon \longrightarrow 0} \left[\int_\epsilon^\infty \frac{d\mathcal{T}}{\mathcal{T}} e^{-A\mathcal{T}} - \int_\epsilon^\infty \frac{d\mathcal{T}}{\mathcal{T}} e^{-B\mathcal{T}}\right] = -\log \frac{A}{B}. \tag{21.87}$$

Here, note that the subtraction removes the singularity at $\mathcal{T} = 0$, and the result follows from applying Eq. (21.85). In terms of the partition function (21.86), this means it is now time to explicitly renormalize it, as we discussed in Section 21.1.2.2:

$$\log \frac{Z}{Z_0} = \frac{1}{2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \mathrm{Tr}\left[\exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right) - \exp\left(-\frac{\mathcal{T}}{\tilde{G}_0}\right)\right]. \tag{21.88}$$

Here, Z_0 and $\tilde{G}_0$ are respectively the partition and Green functions corresponding to a background potential V_0. Writing out the Green function explicitly,

$$\log \frac{Z}{Z_0} = \frac{1}{2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \mathrm{Tr}\left\{ \exp\left[-\mathcal{T}\left(c^2\left(-\hbar^2\partial_{\tilde{\beta}}^2 - \hbar^2\nabla^2\right) + 2V\right)\right] - \exp\left[-\mathcal{T}\left(c^2\left(-\hbar^2\partial_{\tilde{\beta}}^2 - \hbar^2\nabla^2\right) + 2V_0\right)\right]\right\}.$$

(renormalized partition function, four-dimensional form) (21.89)

The choice of V_0 depends on the context of the problem, but could correspond to the scalar-field mass or even $V_0 = 0$. (In the case of multiple interacting objects, the background potential typically corresponds to moving all objects to arbitrarily large separation.)

To simplify this expression for $\log Z$ somewhat before proceeding, we can rescale $\mathcal{T} \longrightarrow 2\mathcal{T}/\hbar^2c^2$, with the result

$$\log Z = \frac{1}{2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \mathrm{Tr}\left\{ \exp\left[-\mathcal{T}\left(-\frac{1}{2}\partial_{\tilde{\beta}}^2 - \frac{1}{2}\nabla^2 + \frac{V}{\hbar^2c^2}\right)\right]\right\}.$$

(partition function, four-dimensional form, rescaled Green operator) (21.90)

To keep the expressions simple, we are suppressing the renormalization term, with the understanding that the resulting expressions only make sense as a difference between comparable material configurations.

We can now carry out the $\tilde{\beta}$ integral as follows. First, we note that we can express the trace above in terms of the space-time basis $|\mathbf{r}, \tilde{\beta}\rangle$, e.g., we can write

$$\log Z = \frac{1}{2} \int d^3r\, d\tilde{\beta} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\, \langle \mathbf{r}, \tilde{\beta}| \left\{ \exp\left[-\mathcal{T}\left(-\frac{1}{2}\partial_{\tilde{\beta}}^2 - \frac{1}{2}\nabla^2 + \tilde{V}\right)\right]\right\} |\mathbf{r}, \tilde{\beta}\rangle, \tag{21.91}$$

where $\tilde{V}$ is shorthand for $V/\hbar^2c^2$. If we focus on the $\tilde{\beta}$-dependent factor, we have a Gaussian integral,

$$\begin{aligned}\int_0^{\hbar c\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\exp\left(\frac{\mathcal{T}}{2}\partial_{\tilde{\beta}}^2\right)|\tilde{\beta}\rangle &= \int_0^{\hbar c\beta} d\tilde{\beta}\int d\xi\,\langle\tilde{\beta}|\exp\left(\frac{\mathcal{T}}{2}\partial_{\tilde{\beta}}^2\right)|\xi\rangle\langle\xi|\tilde{\beta}\rangle \\ &= \int_0^{\hbar c\beta} d\tilde{\beta}\int d\xi\,\exp\left(-\frac{\mathcal{T}}{2\hbar^2}\xi^2\right)|\langle\tilde{\beta}|\xi\rangle|^2 \\ &= \frac{1}{2\pi\hbar}\int d\xi\,\exp\left(-\frac{\mathcal{T}}{2\hbar^2}\xi^2\right)\int_0^{\hbar c\beta} d\tilde{\beta} \\ &= \frac{\hbar c\beta}{2\pi\hbar}\sqrt{\frac{2\pi\hbar^2}{\mathcal{T}}} \\ &= \frac{\hbar c\beta}{\sqrt{2\pi\mathcal{T}}},\end{aligned} \tag{21.92}$$

where we have introduced the conjugate "frequency" ξ_j to $\tilde{\beta}_j$, and we have again used the one-dimensional, normalized plane-wave state $\langle x|p\rangle = e^{ipx/\hbar}/\sqrt{2\pi\hbar}$. Then we have the simplified expression

$$\log Z = \frac{\hbar c\beta}{\sqrt{8\pi}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{3/2}}\,\mathrm{Tr}\left\{\exp\left[-\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]\right\}, \tag{21.93}$$

(unrenormalized partition function, three-dimensional form)

or using Eq. (21.24) to obtain the energy, we have

$$E = -\frac{\hbar c}{\sqrt{8\pi}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{3/2}}\,\mathrm{Tr}\left\{\exp\left[-\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]\right\}, \tag{21.94}$$

(unrenormalized scalar-field Casimir energy, three-dimensional form)

where now the trace is only over the spatial states. Again, this expression for the energy must be renormalized by comparison between two sensible configurations to obtain a finite and physically relevant result.

21.1.4.5 Particle Path Integral

Now the idea is to repeat the path-integration construction of Section 21.1.3, but with ordinary coordinate states. This, of course, follows closely the procedure in Section 20.1. First, we continue the trace in terms of the spatial basis $|\mathbf{r}\rangle$:

$$E = -\frac{\hbar c}{\sqrt{8\pi}}\int d^3r\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{3/2}}\,\langle\mathbf{r}|\left\{\exp\left[-\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]\right\}|\mathbf{r}\rangle. \tag{21.95}$$

Dividing the "proper-time" variable $\mathcal{T}$ into N small bits, and inserting space-time identities, we have

$$\begin{aligned}\int d^3r\,\langle\mathbf{r}|\exp\left[-\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]|\mathbf{r}\rangle &= \int d^3r\,\langle\mathbf{r}|\left\{\prod_{j=1}^{N-1}\int d^3r_j\left[e^{-\Delta\mathcal{T}(-\nabla^2/2+\tilde{V})}|\mathbf{r}_j\rangle\langle\mathbf{r}_j|\right]\right\}e^{-\Delta\mathcal{T}(-\nabla^2/2+\tilde{V})}|\mathbf{r}\rangle \\ &= \int\prod_{j=1}^{N} d^3r_j\,\langle\mathbf{r}_j|\exp\left[-\Delta\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]|\mathbf{r}_{j-1}\rangle \\ &= \int_{\mathbf{r}(\mathcal{T})=\mathbf{r}(0)} D\mathbf{r}\prod_{j=1}^{N}\langle\mathbf{r}_j|\exp\left[-\Delta\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]|\mathbf{r}_{j-1}\rangle,\end{aligned} \tag{21.96}$$

where we are again using the notation

$$\mathbf{r}_0 \equiv \mathbf{r}_N \equiv \mathbf{r}, \tag{21.97}$$

and

$$D\mathbf{r} := \prod_{j=1}^{N} d^dr_j = \prod_{j=0}^{N} d^dr_j\,\delta^d(\mathbf{r}_N-\mathbf{r}_0). \tag{21.98}$$

(path-integration element)

Now in each matrix element, we can split the exponential by dropping a negligible, $O(\Delta\mathcal{T}^2)$ term, and then insert a momentum-basis identity:

$$\begin{aligned}\langle \mathbf{r}_j| \exp\left[-\Delta\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]|\mathbf{r}_{j-1}\rangle &= \langle \mathbf{r}_j| e^{\Delta\mathcal{T}\nabla^2/2} e^{-\Delta\mathcal{T}\tilde{V}}|\mathbf{r}_{j-1}\rangle \\ &= \int d^d p_j\, \langle \mathbf{r}_j|\mathbf{p}_j\rangle\langle \mathbf{p}_j| e^{\Delta\mathcal{T}\nabla^2/2} e^{-\Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})}|\mathbf{r}_{j-1}\rangle \\ &= \int d^d p_j \langle \mathbf{r}_j|\mathbf{p}_j\rangle\langle \mathbf{p}_j| e^{-\Delta\mathcal{T}p_j^2/2\hbar^2} e^{-\Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})}|\mathbf{r}_{j-1}\rangle \\ &= \frac{1}{(2\pi\hbar)^d}\int d^d p_j\, e^{i\mathbf{p}_j\cdot(\mathbf{r}_j-\mathbf{r}_{j-1})/\hbar} e^{-\Delta\mathcal{T}p_j^2/2\hbar^2} e^{-\Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})}.\end{aligned} \tag{21.99}$$

Here, we have introduced the conjugate momentum $\mathbf{p}_j$ to $\mathbf{r}_j$, and we have also introduced the number d of spatial dimensions (i.e., typically $d=3$). Completing the squares and carrying out the Gaussian integrals (recalling that imaginary parts of the integration variable can be shifted away),

$$\begin{aligned}&\langle \mathbf{r}_j| \exp\left[-\frac{1}{2}\Delta\mathcal{T}\left(-\nabla^2+\tilde{V}\right)\right]|\mathbf{r}_{j-1}\rangle \\ &\quad= \frac{1}{(2\pi\hbar)^d}\int d^d p_j\, e^{i\mathbf{p}_j\cdot(\mathbf{r}_j-\mathbf{r}_{j-1})/\hbar - \Delta\mathcal{T}p_j^2/2\hbar^2} e^{-\Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})} \\ &\quad= \frac{1}{(2\pi\hbar)^d}\int d^d p_j\, e^{-(\Delta\mathcal{T}/2\hbar^2)[\mathbf{p}_j - i(\mathbf{r}_j-\mathbf{r}_{j-1})\hbar/\Delta\mathcal{T}]^2} e^{-(\mathbf{r}_j-\mathbf{r}_{j-1})^2/2\Delta\mathcal{T}} e^{-\Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})} \\ &\quad= \frac{1}{(2\pi\hbar)^d}\int d^d p_j\, e^{-(\Delta\mathcal{T}/2\hbar^2)\mathbf{p}_j^2} e^{-(\mathbf{r}_j-\mathbf{r}_{j-1})^2/2\Delta\mathcal{T} - \Delta\mathcal{T}\tilde{V}(\mathbf{r}_{j-1})} \\ &\quad= \frac{1}{(2\pi\Delta\mathcal{T})^{d/2}}\exp\left\{-\Delta\mathcal{T}\left[\frac{1}{2}\left(\frac{\mathbf{r}_j-\mathbf{r}_{j-1}}{\Delta\mathcal{T}}\right)^2+\tilde{V}(\mathbf{r}_{j-1})\right]\right\},\end{aligned} \tag{21.100}$$

where we have ignored $O(\Delta\mathcal{T}^2)$ terms. Then assembling all the factors in Eq. (21.96),

$$\begin{aligned}&\int d^d r\, \langle \mathbf{r}| \exp\left[-\mathcal{T}\left(-\frac{1}{2}\nabla^2+\tilde{V}\right)\right]|\mathbf{r}\rangle \\ &\quad= \frac{1}{(2\pi\Delta\mathcal{T})^{Nd/2}}\int_{\mathbf{r}(\mathcal{T})=\mathbf{r}(0)} D\mathbf{r}\, \prod_{j=1}^{N}\exp\left\{-\Delta\mathcal{T}\left[\frac{1}{2}\left(\frac{\mathbf{r}_j-\mathbf{r}_{j-1}}{\Delta\mathcal{T}}\right)^2+\tilde{V}(\mathbf{r}_{j-1})\right]\right\} \\ &\quad= \frac{1}{(2\pi\Delta\mathcal{T})^{Nd/2}}\int_{\mathbf{r}(\mathcal{T})=\mathbf{r}(0)} D\mathbf{r}\, \exp\left[-\int_0^{\mathcal{T}} d\tau\left(\frac{(\partial_\tau\mathbf{r})^2}{2}+\tilde{V}(\mathbf{r})\right)\right].\end{aligned} \tag{21.101}$$

As in the field path integral, the exponentiated integral is a shorthand for the product of many close-to-unity exponential factors. The choice here of $\mathbf{r}_{j-1}$ (instead of $\mathbf{r}_j$, or some combination) is of course not unique, and as we know from Section 20.4, the time-slicing choice here corresponds to a particular choice of stochastic calculus (here, Itō calculus). This turns out to be of no consequence for this problem, because there is no curvature or space-dependent mass—all the explicit space dependence here is in the potential. It is, however, a more important distinction in the electromagnetic case.

Then, for the sake of completeness, we can write out the Casimir energy (21.95) as

$$E = -\frac{\hbar c}{\sqrt{8\pi}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{3/2}}\,\eta(\mathcal{T})\int_{\mathbf{r}(\mathcal{T})=\mathbf{r}(0)} D\mathbf{r}(\tau)\,\exp\left[-\int_0^{\mathcal{T}} d\tau\left(\frac{(\partial_\tau\mathbf{r})^2}{2}+\frac{V(\mathbf{r})}{\hbar^2c^2}\right)\right], \tag{21.102}$$

(scalar-field Casimir energy (unrenormalized))

where we have renamed the divergent factor $\eta(\mathcal{T}) = (2\pi\Delta\mathcal{T})^{-Nd/2}$, and it is important to remember that this expression is only sensible after renormalization by subtracting the analogous background energy in terms of V_0. Notice that the $\mathcal{T}$ and τ (proper-time) variables that we have introduced in fact have dimensions of *squared length*.

21.1.4.6 Monte-Carlo Integration and Stochastic "Loops"

Now we convert this integral into a Monte-Carlo average, suitable for evaluation on a computer. We motivate the basic idea as follows (as in Section 20.3.1): suppose $f(x)$ is a normalized, nonnegative (probability) distribution; then we may rewrite an integral involving $f(x)$ as an expectation value,

$$\int dx\, f(x)\, g(x) = \left\langle\!\!\left\langle g(x) \right\rangle\!\!\right\rangle_{f(x)}, \tag{21.103}$$

where the double-bracket expectation value is an average over the distribution $f(x)$. Computationally, this allows us to throw random deviates x_n, chosen from the distribution $f(x)$, and simply average the values $g(x_n)$ to obtain an estimate for the integral.

In the integral (21.102), we would like to choose the factor

$$\exp\left[-\int_0^{\mathcal{T}} d\tau\, \frac{(\partial_\tau \mathbf{r})^2}{2}\right] \tag{21.104}$$

as the probability distribution. To normalize it, we note the equivalence

$$\int d^3r\, \langle \mathbf{r}| \exp\left[\frac{\mathcal{T}}{2}\nabla^2\right] |\mathbf{r}\rangle = \eta(\mathcal{T}) \int\limits_{\mathbf{r}(\mathcal{T})=\mathbf{r}(0)} D\mathbf{r}\, \exp\left[-\int_0^{\mathcal{T}} d\tau \frac{(\partial_\tau \mathbf{r})^2}{2}\right], \tag{21.105}$$

which follows from Eq. (21.101) by removing the potential. Thus, we can work out the normalization of this factor, following the procedure of Eq. (21.92):

$$\begin{aligned}
\langle \mathbf{r}| \exp\left[\frac{\mathcal{T}}{2}\nabla^2\right] |\mathbf{r}\rangle &= \int d^d p\, \langle \mathbf{r}| \exp\left[\frac{\mathcal{T}}{2}\nabla^2\right] |\mathbf{p}\rangle\langle \mathbf{p}|\mathbf{r}\rangle \\
&= \int d^d p\, \exp\left[-\frac{\mathcal{T}}{2\hbar^2}p^2\right] |\langle \mathbf{r}|\mathbf{p}\rangle|^2 \\
&= \frac{1}{(2\pi\hbar)^d} \int d^d p\, \exp\left[-\frac{\mathcal{T}}{2\hbar^2}p^2\right] \\
&= \frac{1}{(2\pi\hbar)^d} \left(\frac{2\pi\hbar^2}{\mathcal{T}}\right)^{d/2} \\
&= \frac{1}{(2\pi\mathcal{T})^{d/2}}.
\end{aligned} \tag{21.106}$$

Note that we have not computed the remaining integral with respect to $\mathbf{r}$, which would cause this result to diverge; this could be interpretated as something like the energy *density* at any point may be finite, but the total background energy diverges, as we expect. Recalling that this is the same integration variable as $\mathbf{r}_0 = \mathbf{r}_N$ in the path integral, we can then identify the normalization

$$\int D'\mathbf{r}\, \exp\left[-\int_0^{\mathcal{T}} d\tau \frac{(\partial_\tau \mathbf{r})^2}{2}\right] = \frac{1}{\eta(\mathcal{T})} \langle \mathbf{r}| \exp\left[\frac{\mathcal{T}}{2}\nabla^2\right] |\mathbf{r}\rangle = \frac{1}{\eta(\mathcal{T})(2\pi\mathcal{T})^{d/2}} \tag{21.107}$$

and again, we are only integrating over all the intermediate path coordinates:

$$D'\mathbf{r} = \prod_{j=1}^{N-1} d^d r_j. \tag{21.108}$$

Thus, to convert to a Monte-Carlo average, we should replace this part of the integral with an expectation value over this distribution, and tack on the factor on the right-hand side of Eq. (21.106) to compensate for the normalization of the probability distribution. Thus, Eq. (21.102) becomes

$$E = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^d x_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar^2 c^2}\int_0^{\mathcal{T}} d\tau\, V[\mathbf{x}(\tau)]\right] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)},$$

(scalar-field Casimir energy, Monte-Carlo form) (21.109)

where we have introduced the total number of spacetime dimensions $D := 1 + d$ (i.e., such that typically $D = 4$; recall that we normalized the timelike dimension separately, but in contributes to the overall normalization in the same way as the spacelike dimensions). We have also changed the remaining integration over $\mathbf{r}$ to $\mathbf{x}_0$. Here, the expectation value is taken over functions $\mathbf{x}(\tau)$ chosen according to the (unnormalized) Gaussian probability density

$$\exp\left[-\int_0^{\mathcal{T}} d\tau\, \frac{(\partial_\tau \mathbf{x})^2}{2},\right] \tag{21.110}$$

subject to the closure condition $\mathbf{x}(0) = \mathbf{x}(\mathcal{T}) = \mathbf{x}_0$. Recall that this is shorthand for the discrete version, where points in $3N$-dimensional Cartesian space are chosen according to the Gaussian probability density for the *increments*

$$\prod_{j=1}^{N} \exp\left[-\frac{(\mathbf{x}_j - \mathbf{x}_{j-1})^2}{2\Delta\mathcal{T}}\right], \tag{21.111}$$

again where $\mathbf{x}_0 = \mathbf{x}_N = \mathbf{r}$ is the beginning and termination point of the ''path'' here, which is also the spatial integration variable in (21.102).

Now let's take a closer look at the interpretations of Eqs. (21.102) and (21.109). To simplify the discussion, we'll also take $V = 0$ for the moment. In Eq. (21.102), we have a many-dimensional spatial integral (Nd dimensions, in the limit of large N), where the integrand has Gaussian velocity-weighting factors of the form

$$\exp\left[\frac{(\mathbf{x}_j - \mathbf{x}_{j-1})^2}{2\Delta\mathcal{T}}\right]. \tag{21.112}$$

Thus, successive positions (separated by $\Delta\mathcal{T}$ in ''time'') in the discrete form of the parameterized many-dimensional coordinate $\mathbf{x}(\tau)$ must be close together—of the order of $\sqrt{\Delta\mathcal{T}}$ or less—otherwise the contribution of that particular point $\mathbf{x}(\tau)$ to the integral is negligible. As such, only a very small fraction of all possible $\mathbf{x}(\tau)$—essentially, those that correspond to continuous paths in the large-N limit—can possible contribute to the integral.

In Eq. (21.109), the idea is to make the integral much easier to evaluate. We focus only on those paths that whose amplitudes are not destroyed by the factor (21.112), by using precisely these factors to determine which paths to choose (at random). The coordinates $\mathbf{r}_j$ and $\mathbf{r}_{j-1}$ [equivalently, $\mathbf{r}(\tau)$ and $\mathbf{r}(\tau - \Delta\mathcal{T})$] are separated by a random distance of variance $\Delta\mathcal{T}$ (in each direction). Because the first and last points in this stochastic path are identical, and this is a stochastic ''loop.'' The path is a continuous random walk, where the components of $\mathbf{x}(\tau)$ have the form

$$\mathbf{x}(\tau) = \mathbf{x}_0 + \sqrt{\mathcal{T}}\,\mathbf{B}\Big(\frac{\tau}{\mathcal{T}}\Big), \tag{21.113}$$

with $\mathbf{B}(t)$ a **vector Brownian bridge**, having scalar **Brownian bridges** $B_j(t)$ as its d components. Recall from Section 17.7 that Brownian bridges are the continuous limits of Gaussian random walks that return to their initial point, $B(1) = B(0)$. The paths $\mathbf{x}(\tau)$ themselves are the **world lines** in this quantum-Monte-Carlo calculation, being parameterized by the ''proper time'' $\mathcal{T}$. In any case, for the evaluation of the Casimir energy in Eq. (21.109), we must still weight each path according to the Gaussian potential factor for both the configuration potential $V(\mathbf{r})$ and the background potential $V_0(\mathbf{r})$. Rather than discuss these abstractly, we will illustrate the calculation of Casimir potentials in this formalism with a couple of simple examples.

21.1.5 Analytic Evaluation of Scalar Casimir Energies

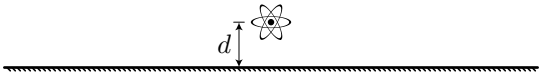
21.1.5.1 Strong-Coupling Limit: Atom–Plane Interaction

To compute the Casimir–Polder interaction of an atom with a conducting plane, we will first note that this theory does not incorporate dispersion, so we will stick to a perfectly conducting plane, and we will not expect to get the crossover from z^{-3} to z^{-4} behavior, as this requires dispersion of the atom (frequency-dependent polarizability). Also, to compute the interaction with the perfectly conducting plane, we will consider the

"strong-coupling limit" of $V(\mathbf{r}) \longrightarrow \infty$ at any location of the conducting body ($V(\mathbf{r}) = 0$ elsewhere outside the body). For concreteness, we may write

$$V(\mathbf{r}) = \chi\Theta(z - d), \tag{21.114}$$

where the coupling parameter $\chi \longrightarrow \infty$, $\Theta(x)$ is the Heaviside function, the boundary of the plane is given by $z = d$, and the atom is situated at the origin so that d is also the atom–mirror separation.



Recall [see Eq. (1.60)] that the dipole potential for an atom interacting via its induced dipole moment with an electric field has the form $V_{\text{dipole}} = -(1/2)\alpha_0 E^2$, where α_0 is the dc polarizability. Again, we are ignoring dispersion here, so we are implicitly assuming that the atomic polarizability satisfies $\alpha(\omega) = \alpha_0$. In this case, we do not have a steady field, so we should interpret the squared electric field as an expectation value in the vacuum [as we did in Section 13.6 for the atomic dipole moment in the near-field Casimir–Polder potential; see also the discussion leading up to Eq. (13.201)]:

$$V_{\text{dipole}} = -\frac{1}{2}\alpha_0 \big\langle E^2 \big\rangle \,. \tag{21.115}$$

Now we must connect this expression to the Casimir energy in Eq. (21.109). First, note that the vacuum energy density of the electromagnetic field at any particular point has the similar form

$$\mathscr{E}(\mathbf{r}) = \epsilon_0 \big\langle E^2(\mathbf{r}) \big\rangle \,. \tag{21.116}$$

We can compare this expression to the energy density that we can deduce from Eq. (21.109), by noting that the spatial integral represents an energy integrated over the energy density, and also taking the renormalization to be against vacuum, $V_0(\mathbf{r}) = 0$:

$$\mathscr{E}(\mathbf{r}) - \mathscr{E}_0(\mathbf{r}) = \frac{\hbar c}{8\pi^2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle 1 - \exp\left[- \int_0^{\mathcal{T}} d\tau\, \tilde{V}[\mathbf{x}(\tau)] \right] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \tag{21.117}$$

As usual, we take $\mathbf{x}(0) = \mathbf{x}(\tau) = \mathbf{r}$, which we in turn take to be the origin, where the atom is located. Combining the above three expressions, we can then write

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{16\pi^2 \epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle 1 - \exp\left[- \int_0^{\mathcal{T}} d\tau\, \tilde{V}[\mathbf{x}(\tau)] \right] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{21.118}$$

for the world-line form of the Casimir–Polder potential. In this form, the potential is completely general, specified in terms of an arbitrary potential $V(\mathbf{r})$ and an atomic location $\mathbf{r}$ as the source point for the paths.

Actually, in the above argument for the interpreting the spatial integrand of the total energy (21.109) as the energy density, we should be more careful: there are many *other* functions whose integral could add up to the right energy. To see this more directly, note that we can trace through the whole derivation again, starting with a Lagrangian density $\mathscr{L}(\mathbf{r})$ that samples the fields at only one point in space,

$$L = \int d^3r\, \mathscr{L}(\mathbf{r}), \tag{21.119}$$

and then carrying through the derivation with $\mathscr{L}$ instead of L. We can fix the Lagrangian density uniquely by starting with a partition-function "density" analogous to Eq. (21.10),

$$Z(\mathbf{r}) := \operatorname{Tr}\left[e^{-\beta \mathscr{H}(\mathbf{r})\, d^3r} \right], \tag{21.120}$$

where $\mathscr{H}(\mathbf{r})$ is the Hamiltonian density, defined by

$$H =: \int d^3r\, \mathscr{H}(\mathbf{r}). \tag{21.121}$$

The conjugate fields are defined with respect to the *full* Hamiltonian H, as the fields must respect *global* boundary conditions. Note that we can decompose the total partition function into the partition density by writing out the trace,

$$\begin{aligned} Z &= \sum_n \langle n|e^{-\beta H}|n\rangle \\ &= \sum_n \langle n| \exp\left(-\beta \int d^3r\, \mathscr{H}(\mathbf{r})\right)|n\rangle \\ &= \sum_n \langle n| \prod_{\mathbf{r}} \exp\left(-\beta \mathscr{H}(\mathbf{r})\, d^3r\right)|n\rangle \\ &= \prod_{\mathbf{r}} \sum_n \langle n| \exp\left(-\beta \mathscr{H}(\mathbf{r})\, d^3r\right)|n\rangle, \end{aligned} \tag{21.122}$$

where the product is over all points in space (as the integral is a sum over all points in space), and the last step follows by inserting identities of the form $\sum_{n'} |n'\rangle\langle n'|$ in terms of energy eigenstates. Then we have

$$\log Z = \int d^3r \log Z(\mathbf{r}), \tag{21.123}$$

where we have dropped an additive constant ($\log d^3r$) on the right-hand side. Because we $\log Z$ to compute the energy, we see that the energy density simply integrates to the total energy, in terms of the logarithms of the corresponding partition functions. Furthermore, $Z(\mathbf{r})$ clearly only represents the field energy at a single point $\mathbf{r}$, so it produces the proper energy density. Everything in the derivation then carries through with $\int d\tilde{\beta}\, d^3r$ replaced by $\int d\tilde{\beta}$ up through Eq. (21.42). At this point, the functional determinants and traces refer only to $\tilde{\beta}$, and no longer to $\mathbf{r}$. Thus, the expression of the trace in Eq. (21.91) in terms of a Euclidean space-time integral is just a time ($\tilde{\beta}$) integral, and the result in deriving the path integral is just the spatial integrand of (21.109), as desired. That is to say, the spatial integral in the total energy (21.109) is exactly the spatial integral that appears in the Hamiltonian (21.7).

Coming back to the planar-mirror case, this expression is extremely simple to interpret in the strong-coupling limit. If the path touches the surface, then the argument in the exponential of Eq. (21.117) diverges negatively, so the exponential vanishes, and the overall contribution of that path to the expectation value is unity. Otherwise, the contribution of a non-touching path simply vanishes (as a direct result of renormalization, since these paths instead contribute to the Lamb shift). Then the z-coordinate of the path is the only one relevant to the calculation, and we must only keep track if its maximum excursion $\sup[z(\tau)]$ takes it past $z = d$. We can write this as an average over Heaviside functions of the paths as

$$\begin{aligned} V_{\text{dipole}} &= -\frac{\hbar c \alpha_0}{16\pi^2\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle \Theta\Big\{ \sup[z(\tau)] - d \Big\} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= -\frac{\hbar c \alpha_0}{16\pi^2\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle \Theta\Big\{ \sqrt{\mathcal{T}} \sup[B(t)] - d \Big\} \right\rangle\!\!\right\rangle_{B(t)}, \end{aligned} \tag{21.124}$$

where we have transformed to the scale-independent, standard Brownian bridge $B(t)$, and $t = \tau/\mathcal{T} \in [0,1]$. Thus, we see that the renormalization simply cuts off the lower (divergent) end of the integral. To evaluate this analytically, we will implement the cutoff separately for each path:

$$\begin{aligned} V_{\text{dipole}} &= -\frac{\hbar \alpha_0}{16\pi^2\epsilon_0} \left\langle\!\!\left\langle \int_{d^2/\sup^2[B(t)]}^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \right\rangle\!\!\right\rangle_{B(t)} \\ &= -\frac{\hbar c \alpha_0}{32\pi^2\epsilon_0} \left\langle\!\!\left\langle \left(\frac{\sup^2[B(t)]}{d^2}\right)^2 \right\rangle\!\!\right\rangle_{B(t)} \\ &= -\frac{\hbar c \alpha_0}{32\pi^2\epsilon_0 d^4} \left\langle\!\!\left\langle \{\sup[B(t)]\}^4 \right\rangle\!\!\right\rangle_{B(t)}. \end{aligned} \tag{21.125}$$

In scaling out the path sizes, we see that the z^{-4} scaling of the Casimir–Polder force is built into the path scale and the $\mathcal{T}^{-3}$ dependence of the integral. The path statistics enter here simply to give an overall constant factor. As it turns out, the maximum, one-sided excursion $x = \sup[B(t)]$ (over $t \in [0,1]$) has a well-defined probability density [Eq. (17.383)]:

$$f(x) = 4xe^{-2x^2} \qquad (x \geq 0). \tag{21.126}$$

The fourth moment of this distribution is 1/2 [from Eq. (17.385)]. Putting in this value for the fourth moment, we find the scalar result

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{64\pi\epsilon_0 \pi d^4} = -\frac{3\hbar c \alpha_0}{(4\pi\epsilon_0)8\pi d^4}\left(\frac{1}{6}\right),$$

(Casimir–Polder potential, perfectly conducting plane, scalar result) (21.127)

The asymptotic expression for the electromagnetic field in the limit of large z was [see Eq. (13.60)]

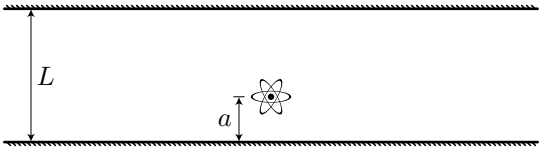
$$V_{\text{CP}} = -\frac{\hbar c \alpha_0}{(4\pi\epsilon_0)8\pi d^4},$$

(Casimir–Polder potential, perfectly conducting plane, far-field electromagnetic result) (21.128)

which is the same except for the factor of 1/6. Again, since we are ignoring dispersion, it is most sensible to compare to the far-field result, where only the dc atomic polarizability contributes. While the scalar result does not quite agree with the electromagnetic Casimir–Polder potential, the scalar result *does* reproduce the contribution of only the TE modes to the total potential; the factor of 1/6 here is the same factor that appears in Eq. (13.181). This is sensible, since TE modes at a planar surface behave as scalar waves, since the polarization is the same for the incident, reflected, and transmitted waves, and effectively the polarization drops out of the calculation. By contrast, the polarization is different for the same three components of TM waves, and the scalar result does not capture this extra complexity.

21.1.5.2 Strong-Coupling Limit: Atomic Interaction with Two Planes

Next, we consider the slightly more complicated case of an atom interacting with two parallel, conducting planes. The atom is situated at $z = 0$, with the barriers at $z = -a$ and $z = L - a$.



We consider only the case of the atom *between* the two planes, as if the atom is not in between them, the potential is simply the one-plane potential due to the nearer surface.

To evaluate the potential here, we need the probability that a stochastic path touches *either* surface. This is equivalent to the problem of calculating the *escape probability* of a Brownian bridge outside the interval defined by $[-a, L-a]$. We have already calculated this before, and the result for a standard Brownian bridge is [Eq. (17.415)]

$$P_{\text{escape}} = e^{-2a^2} + \sum_{j=1}^{\infty}\left[e^{-2(jL-a)^2} + e^{-2(jL+a)^2} - 2e^{-2j^2L^2}\right]. \tag{21.129}$$

However, we have paths $x(\tau)$, which by their definition are equivalent to Brownian bridges $B_{2\mathcal{T}}(t)$ running in time from 0 to $2\mathcal{T}$, for which the escape probability is given by scaling all squared lengths by this "time interval":

$$P_{\text{escape}}(\mathcal{T}) = e^{-2a^2/\mathcal{T}} + \sum_{j=1}^{\infty}\left[e^{-2(jL-a)^2/\mathcal{T}} + e^{-2(jL+a)^2/\mathcal{T}} - 2e^{-2j^2L^2/\mathcal{T}}\right]. \tag{21.130}$$

This is precisely the path average in Eq. (21.118), since the path average is of a function that is unity whenever the bridge escapes (touches either plane), and 0 otherwise. Thus, Eq. (21.118) becomes

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{16\pi^2\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3}\left[e^{-2a^2/\mathcal{T}} + \sum_{j=1}^\infty \left(e^{-2(jL-a)^2/\mathcal{T}} + e^{-2(jL+a)^2/\mathcal{T}} - 2e^{-2j^2L^2/\mathcal{T}}\right)\right], \tag{21.131}$$

and using the integral result

$$\int_0^\infty dx\, \frac{e^{-a/x}}{x^3} = \frac{1}{a^2}, \tag{21.132}$$

we finally have

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0}\left[\frac{1}{a^4} + \sum_{j=1}^\infty \left(\frac{1}{(jL-a)^4} + \frac{1}{(jL+a)^4} - \frac{2}{j^4L^4}\right)\right]$$

(Casimir–Polder potential between two conducting planes) (21.133)

Note that the first term here agrees with the single-plate result (21.127), while the other terms represent "reflected images" due to the other mirror; all the other terms vanish in the single-plane limit $L \longrightarrow \infty$. Note also that this result is invariant under the replacement $a \longrightarrow L - a$, as it must be. The a-independent term here technically doesn't influence the atomic potential, and can be dropped in the renormalization, though we will keep it here for illustrative purposes. Also, note the similarity of the summation structure to the frequency and lifetime shifts of a Lorentz atom between two parallel, planar conductors (Problem 1.3).

Finally, note that we can use the summation formula

$$\sum_{j=1}^\infty \frac{1}{(j+x)^4} = \frac{\psi^{(3)}(1+x)}{6}, \tag{21.134}$$

where $\psi^{(n)}(x)$ is a polygamma function, which is the nth derivative of the digamma function $\psi(x) := \partial_x \Gamma(x)$ [and $\Gamma(x)$ is the gamma function], along with the limiting value $\psi^{(3)}(1) = 6\zeta(4) = \pi^4/15$. The summation formula follows by differentiating the series formula[3]

$$\psi(1+x) = -\gamma + \sum_{j=1}^\infty \frac{x}{j(j+x)} = -\gamma + \sum_{j=1}^\infty \left(\frac{1}{j} - \frac{1}{j+x}\right) \tag{21.135}$$

(which technically is invalid when x is a negative integer, but this won't be a problem here). Thus, we have

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0}\left[\frac{1}{a^4} - \frac{\pi^4}{45L^4} + \frac{1}{6L^4}\Big[\psi^{(3)}(1-a/L) + \psi^{(3)}(1+a/L)\Big]\right]$$

(Casimir–Polder potential between two conducting planes) (21.136)

as an analytic form for the potential of an atom between two planes, within scalar theory. Again, the L^{-4} term only contributes an overall offset to the atomic potential, and can be dropped.

21.1.5.3 Strong-Coupling Limit: Plane–Plane Interaction

As a final example we evaluate directly the Casimir interaction between two parallel planes separated by distance L, this time without any atom involved.



[3] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 259, Eq. (6.3.16).

To do this, we return to Eq. (21.109) for the (unrenormalized) Casimir energy. As a reference, we wish to reference the absolute energy such that the Casimir energy vanishes as $L \longrightarrow \infty$. Thus, we will take $V(\mathbf{x})$ to refer to the two parallel planes separated by L, and $V_0(\mathbf{x})$ to refer to the two parallel planes with arbitrarily large separation. As before, the expectation value of each exponential in Eq. (21.109) is equivalent to $1 - P_{\text{touch}}$, where P_{touch} is the probability of the path $\mathbf{x}(\tau)$ to touch the surface defined by the appropriate potential function $V(\mathbf{x})$. But we have to be careful when computing the touch probabilities here. The touch probability for $V(\mathbf{x})$ is straightforward: P_{touch} refers to the probability to touch *either* plate. However, the touch probability for $V_0(\mathbf{x})$ is a bit trickier. Since we are counting the case of *widely separated* plates, we must count all the touch probabilities that are *equivalent* to the touching path in the $V(\mathbf{x})$ case. Each path that touches $V(\mathbf{x})$ should be compared with *two* background paths: one that is ''near'' each of the now widely separated plates. Thus, the path average with renormalization in Eq. (21.109) is

$$\begin{aligned}\Big[1 - P_{\text{touch}}[V(\mathbf{x})]\Big] - \Big[1 - P_{\text{touch}}[V_0(\mathbf{x})]\Big] &= P_{\text{touch}}[V_0(\mathbf{x})] - P_{\text{touch}}[V(\mathbf{x})] \\ &= P(\text{touch } 1) + P(\text{touch } 2) - P(\text{touch } 1 \vee \text{touch } 2) \\ &= P(\text{touch } 1 \wedge \text{touch } 2).\end{aligned} \tag{21.137}$$

That is, we only count paths that touch *both* planes; paths that touch only one plane (or neither plane) are dropped after renormalization. With the global minus sign in the Casimir energy, the overall energy is *negative* in this renormalization scheme (leading to an attractive two-body force).

Now we will work out the relevant probabilities and compute the energy in two parts, representing the exterior and interior of the pair of planes. First, the exterior. Consider a point a exterior to the pair of planes. The probability for a path starting and ending at a to touch *both* planes is just the probability to touch the more distant plane, which is a distance $L + a$ away. The probability for a standard Brownian bridge to cross a boundary a distance d away is [Eq. (17.380)]

$$P_{\text{cross}}(d) = e^{-2d^2}. \tag{21.138}$$

Again, we are effectively considering Brownian bridges over a time interval $\mathcal{T}$, so the probability for the path $x(\tau)$ to touch the more distant surface is given by scaling d^2 down by $\mathcal{T}$, and then letting $d \longrightarrow L + a$:

$$P(\text{touch } 1 \wedge \text{touch } 2) = e^{-(L+a)^2/\mathcal{T}}. \tag{21.139}$$

The energy associated with the exterior then is just Eq. (21.109), integrated over the extent of each exterior region. These are equivalent, so we just count twice the result for a single region:

$$\begin{aligned}\frac{(E - E_0)_{\text{exterior}}}{A} &= -\frac{\hbar c}{8\pi^2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left[2 \int_0^\infty da\, e^{-2(L+a)^2/\mathcal{T}}\right] \\ &= -\frac{\hbar c}{16\pi^2} \int_0^\infty da\, \frac{1}{(L+a)^4} \\ &= -\frac{\hbar c}{32\pi^2 L^3} \left(\frac{2}{3}\right).\end{aligned} \tag{21.140}$$

We again used the integral (21.132) to evaluate the $\mathcal{T}$ integral. Also, we only integrated one dimension out of the full volume integral in Eq. (21.109), dividing by the cross-sectional area A in lieu of performing the (divergent) transverse integrals. Note that while this appears to be a contribution from the ''exterior'' of the two planes, this is really an artifact of the renormalization. In the exterior region, the contributions in Eqs. (21.137) for touching the near plane and for touching *either* plane exactly cancel. The leftover is the contribution for touching the *far* plane, which is a contribution from subtracting one of the *one*-body energies. This can be viewed as an *interior* contribution, but from when the planes are far apart.

For the interior contribution, we need the probability for a path to touch *both* surfaces. However, it is easier to use the probability to touch *either* surface, since that is the escape probability that we used above,

Eq. (21.129). Choosing an interior point a distance a away from the two surfaces, the escape probability for $x(\tau)$ is precisely what we used before for the atom between two planes, Eq. (21.130). Thus, what we need is

$$\begin{aligned} P(\text{touch 1} \wedge \text{touch 2}) &= P(\text{touch 1}) + P(\text{touch 2}) - P(\text{touch 1} \vee \text{touch 2}) \\ &= e^{-2a^2/\mathcal{T}} + e^{-2(L-a)^2/\mathcal{T}} \\ &\quad - e^{-2a^2/\mathcal{T}} - \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2/\mathcal{T}} + e^{-2(jL+a)^2/\mathcal{T}} - 2e^{-2j^2L^2/\mathcal{T}} \right] \\ &= e^{-2(L-a)^2/\mathcal{T}} - \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2/\mathcal{T}} + e^{-2(jL+a)^2/\mathcal{T}} - 2e^{-2j^2L^2/\mathcal{T}} \right]. \end{aligned} \tag{21.141}$$

Here we have again used the single-barrier crossing probability for a Brownian bridge by rescaling the distance in Eq. (21.138) appropriately. Putting this probability in for the expectation value in Eq. (21.109) and integrating over the interior of the two planes,

$$\frac{(E-E_0)_{\text{interior}}}{A} = -\frac{\hbar}{8\pi^2 c^3} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \int_0^L da \left\{ e^{-2(L-a)^2/\mathcal{T}} - \sum_{j=1}^{\infty} \left[e^{-2(jL-a)^2/\mathcal{T}} + e^{-2(jL+a)^2/\mathcal{T}} - 2e^{-2j^2L^2/\mathcal{T}} \right] \right\}. \tag{21.142}$$

Again carrying out the $\mathcal{T}$ integral first,

$$\begin{aligned} \frac{(E-E_0)_{\text{interior}}}{A} &= -\frac{\hbar c}{32\pi^2} \int_0^L da \left\{ \frac{1}{(L-a)^4} - \sum_{j=1}^{\infty} \left[\frac{1}{(jL-a)^4} + \frac{1}{(jL+a)^4} - \frac{2}{j^4L^4} \right] \right\} \\ &= \frac{\hbar c}{32\pi^2} \int_0^L da \left\{ \sum_{j=2}^{\infty} \frac{1}{(jL-a)^4} + \sum_{j=1}^{\infty} \left[\frac{1}{(jL+a)^4} - \frac{2}{j^4L^4} \right] \right\} \\ &= \frac{\hbar c}{32\pi^2} \left\{ \frac{1}{3L^3} \sum_{j=2}^{\infty} \left(\frac{1}{(j-1)^3} - \frac{1}{j^3} \right) + \sum_{j=1}^{\infty} \left[\frac{1}{3L^3} \left(\frac{1}{j^3} - \frac{1}{(j+1)^3} \right) - \frac{2}{j^4L^3} \right] \right\} \\ &= \frac{\hbar c}{32\pi^2 L^3} \sum_{j=1}^{\infty} \left[\frac{2}{3} \left(\frac{1}{j^3} - \frac{1}{(j+1)^3} \right) - \frac{2}{j^4} \right]. \end{aligned} \tag{21.143}$$

Here, note that the sums over $1/j^3$ and $1/(j+1)^3$ are equivalent, except that the second sum is missing a term of unity. We have already found that the sum over $1/j^4$ is $\pi^4/90$. Thus,

$$\frac{(E-E_0)_{\text{interior}}}{A} = \frac{\hbar c}{32\pi^2 L^3} \left(\frac{2}{3} - \frac{\pi^4}{45} \right). \tag{21.144}$$

Then we obtain the total energy by adding Eqs. (21.144) and (21.140). The common 2/3 term cancels out, leaving only the $\pi^4/45$ term, with the result

$$\boxed{\frac{E-E_0}{A} = -\frac{\pi^2 \hbar c}{1440 L^3}.}$$

(Casimir energy density of two conducting planes, scalar result) (21.145)

Note that this energy appears as an offset in the atom–two-planes potential (21.136). This result is exactly half of the true Casimir energy for two conducting planes,[4]

$$\boxed{\frac{E-E_0}{A} = -\frac{\pi^2 \hbar c}{720 L^3},}$$

(Casimir energy density of two conducting planes, electromagnetic result) (21.146)

suggesting that the two polarizations of the electromagnetic field in this geometry act as two independent scalar fields.

[4] H. B. G. Casimir, "On the attraction between two perfectly conducting plates," *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* **51**, 793 (1948).

21.1.6 World Lines at Finite Temperature

Thus far, we have focused on the zero-temperature limit of the scalar Casimir effect. However, it is not difficult to adapt this theory to the case of nonzero temperature.[5] To do this, we return to the partition-function expression (21.91), and again focus on the $\tilde{\beta}$-dependent factor, but convert it into a path integral rather than simply evaluating it as in Eq. (21.92). First, letting $\Delta\mathcal{T} = \hbar\beta/N$, with $\tilde{\beta}_0 \equiv \tilde{\beta}_N \equiv \tilde{\beta}$,

$$\begin{aligned}\int_0^{\hbar\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\exp\left(\mathcal{T}\partial_{\tilde{\beta}}^2\right)|\tilde{\beta}\rangle &= \int_0^{\hbar\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\left\{\prod_{j=1}^{N-1}\int d\tilde{\beta}_j\left[e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}_j\rangle\langle\tilde{\beta}_j|\right]\right\}e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}\rangle \\ &= \int\prod_{j=1}^{N} d\tilde{\beta}\,\langle\tilde{\beta}_j|e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}_{j-1}\rangle \\ &= \int_{\tilde{\beta}(\mathcal{T})=\tilde{\beta}(0)} D\tilde{\beta}\prod_{j=1}^{N}\langle\tilde{\beta}_j|e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}_{j-1}\rangle.\end{aligned} \tag{21.147}$$

Here we must be somewhat careful with the integration limits. While the endpoint $\tilde{\beta}$ spans 0 to $\hbar c\beta$, the other $\tilde{\beta}_j$ are *unrestricted*. Since they are merely intermediate coordinates we may define as we wish, and in particular we want to be able to introduce the usual momentum eigenstates (plane waves) for these coordinates. Again, in each matrix element, we can insert a momentum-basis identity, where ξ_j is conjugate to $\tilde{\beta}_j$, and then complete the square and carry out the Gaussian integral:

$$\begin{aligned}\langle\tilde{\beta}_j|e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}_{j-1}\rangle &= \int d\xi_j\,\langle\tilde{\beta}_j|\xi_j\rangle\langle\xi_j|e^{\Delta\mathcal{T}\partial_{\tilde{\beta}}^2}|\tilde{\beta}_{j-1}\rangle \\ &= \int d\xi_j\,\langle\tilde{\beta}_j|\xi_j\rangle\langle\xi_j|e^{-\Delta\mathcal{T}\xi_j^2/\hbar^2}|\tilde{\beta}_{j-1}\rangle \\ &= \frac{1}{2\pi\hbar}\int d\xi_j\,e^{i\xi_j(\tilde{\beta}_j-\tilde{\beta}_{j-1})/\hbar}e^{-\Delta\mathcal{T}\xi_j^2/\hbar^2} \\ &= \frac{1}{2\pi\hbar}\int d\xi_j\,e^{-(\Delta\mathcal{T}/\hbar^2)[\xi_j - i(\tilde{\beta}_j-\tilde{\beta}_{j-1})\hbar/2\Delta\mathcal{T}]^2}e^{-(\tilde{\beta}_j-\tilde{\beta}_{j-1})^2/4\Delta\mathcal{T}} \\ &= \frac{1}{\sqrt{4\pi\Delta\mathcal{T}}}\exp\left[-\frac{(\tilde{\beta}_j-\tilde{\beta}_{j-1})^2}{4\Delta\mathcal{T}^2}\right].\end{aligned} \tag{21.148}$$

Again, we should be a bit careful here, as technically this procedure doesn't work for the conjugate momentum to $\tilde{\beta}_N$, because it has a discrete spectrum, corresponding to a bounded "time." We can fix this by introducing $|\xi_{N-1}\rangle$ again for this last matrix element, and obtain the equivalent result. Now putting this matrix element back into Eq. (21.147),

$$\begin{aligned}\int_0^{\hbar c\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\exp\left(\mathcal{T}\partial_{\tilde{\beta}}^2\right)|\tilde{\beta}\rangle &= \frac{1}{(4\pi\Delta\mathcal{T})^{N/2}}\int_{\tilde{\beta}(\mathcal{T})=\tilde{\beta}(0)} D\tilde{\beta}\prod_{j=1}^{N}\exp\left[-\frac{(\tilde{\beta}_j-\tilde{\beta}_{j-1})^2}{4\Delta\mathcal{T}^2}\right] \\ &= \frac{1}{(4\pi\Delta\mathcal{T})^{N/2}}\int_{\tilde{\beta}(\mathcal{T})=\tilde{\beta}(0)} D\tilde{\beta}\,\exp\left[-\int_0^{\mathcal{T}} d\tau\,\frac{(\partial_\tau\tilde{\beta})^2}{4}\right]\end{aligned} \tag{21.149}$$

As we computed in Eq. (21.92) or in Eq. (21.106), the normalization factor for this path integral is

$$\int_0^{\hbar c\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\exp\left(\mathcal{T}\partial_{\tilde{\beta}}^2\right)|\tilde{\beta}\rangle = \frac{\hbar c\beta}{\sqrt{4\pi\mathcal{T}}}. \tag{21.150}$$

Then we have shown that the path average in $\tilde{\beta}$ is normalized such that

$$\left(\frac{\hbar c\beta}{\sqrt{4\pi\mathcal{T}}}\right)^{-1}\frac{1}{(4\pi\Delta\mathcal{T})^{N/2}}\int_{\tilde{\beta}(\mathcal{T})=\tilde{\beta}(0)} D\tilde{\beta}\,\exp\left[-\int_0^{\mathcal{T}} d\tau\,\frac{(\partial_\tau\tilde{\beta})^2}{4}\right] = 1, \tag{21.151}$$

[5] Klaus Klingmüller and Holger Gies, "Geothermal Casimir phenomena," *Journal of Physics A: Mathematical and Theoretical* **41**, 164042 (2008) (doi: 10.1088/1751-8113/41/16/164042).

and so, in switching to a path average, we are making the replacement

$$\int_0^{\hbar c\beta} d\tilde{\beta}\,\langle\tilde{\beta}|\exp\left(\mathcal{T}\partial_{\tilde{\beta}}^2\right)|\tilde{\beta}\rangle \longrightarrow \frac{\hbar c\beta}{\sqrt{4\pi\mathcal{T}}}\Big\langle\ \Big\rangle_{\tilde{\beta}(\tau)}. \tag{21.152}$$

After differentiating $\log Z$ with respect to β to get the energy, the factor of β here disappears, and we obtain the same expression as before, Eq. (21.109), but with a different interpretation: the path average is now with respect to *four*-dimensional paths, where the fourth dimension is the $\tilde{\beta}$-direction (''time'' direction). Of course, the potential doesn't couple to this dimension, so we will evaluate it separately from the three spatial dimensions.

To evaluate the time-dimension path integral, first we should return to the $\tilde{\beta}$ integral, as for example in Eq. (21.150). Recall that $\tilde{\beta}$ entered as the ''imaginary time'' in the partition function (21.39), running from $\tilde{\beta} = 0$ to $\hbar\beta$. The dependence on $\tilde{\beta}$ comes in via the fields $\phi(\mathbf{r}, \tilde{\beta})$, where the endpoint constraint (21.29) reads

$$\phi(\mathbf{r}, \tilde{\beta} = \hbar\beta) = \phi(\mathbf{r}, \tilde{\beta} = 0). \tag{21.153}$$

Since the only dependence on $\tilde{\beta}$ is periodic with period $\hbar c\beta$, we may take $\tilde{\beta}$ itself to be a periodic coordinate, with period $\hbar c\beta$. This helps us to make sense of the path integral in Eq. (21.150), where recall that the intermediate steps in the path integral are defined on an unbounded (aperiodic) coordinate. Thus it is possible for a path beginning at some $\tilde{\beta}$ to wander out of $[0, \hbar\beta]$ as $\tau \longrightarrow \mathcal{T}$, but the next-to-last and last points must be close $[O(\sqrt{\Delta\mathcal{T}})]$ together. This can still work if the path reconnects to an ''image'' of $\tilde{\beta}$ at $\tilde{\beta} + n\hbar c\beta$, for some integer n. Ultimately, this is equivalent to having paths on a cylindrical manifold, where the $\tilde{\beta}$ direction is periodic, and the spatial dimensions are extended. Then paths can either reconnect directly to the same spot, as in the spatial dimensions, or reconnect by winding around the $\tilde{\beta}$-direction. In the zero-temperature limit, the extent of the $\tilde{\beta}$-direction becomes arbitrarily large, reducing space to an extended four-dimensional manifold.

Now from what we know about Brownian bridges (Section 17.7), any bridge $B(t)$ can be ''deformed'' to connect the point 0 to the point c by introducing a constant drift. The relative probability of the deformed bridge, or the weight of the deformed bridge compared to the standard bridge, is

$$\frac{P[W(1) = c]}{P[W(1) = 0]} = \frac{e^{-c^2/2}\,dc/\sqrt{2\pi}}{e^{-0^2/2}\,dc/\sqrt{2\pi}} = e^{-c^2/2}. \tag{21.154}$$

However, we are considering paths that are pinned at time $2\mathcal{T}$ rather than 1, so we really should consider

$$\frac{P[W(2\mathcal{T}) = c]}{P[W(2\mathcal{T}) = 0]} = e^{-c^2/4\mathcal{T}}. \tag{21.155}$$

Thus, if we normalize to the zero-temperature case, we can count all possible paths with winding number n, and weight them by this factor, with $c \longrightarrow n\hbar c\beta$. We will then include the many more possible paths in the nonzero-temperature case, but organized by winding number, with each set mapped onto the zero-temperature paths, and weighted explicitly. Then the path average is *only* with respect to the zero-temperature paths, and except for these probability weights, the $\tilde{\beta}$ part of the path integral goes away. Thus, the path-average expression (21.156) for the (unrenormalized) Casimir energy becomes

$$E = -\frac{\hbar c}{32\pi^2}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3}\left(\sum_{n=-\infty}^{\infty} e^{-n^2\hbar^2c^2\beta^2/4\mathcal{T}}\right)\int d^3r\left\langle \exp\left[-\int_0^{\mathcal{T}} d\tau\,\tilde{V}[\mathbf{x}(\tau)]\right]\right\rangle,$$

(scalar-field Casimir energy, nonzero temperature) (21.156)

where we have not written the renormalization term to keep the expression (relatively) simple, and the new factor is the sum over all winding numbers. Note that this reduces to the zero-temperature case when $\beta \longrightarrow \infty$, so that only the $n = 0$ term survives, and the sum is replaced by unity.

21.1.6.1 Example: Temperature-Dependent Atom–Planar-Conductor Potential

With temperature dependence, we can similarly adapt the atom–surface potential (21.118) to read

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3}\left(\sum_{n=-\infty}^{\infty} e^{-n^2\hbar^2c^2\beta^2/4\mathcal{T}}\right)\left\langle 1-\exp\left[-\int_0^{\mathcal{T}} d\tau\,\tilde{V}[\mathbf{x}(\tau)]\right]\right\rangle_{\mathbf{x}(\tau)}. \tag{21.157}$$

We will evaluate this for a planar surface in the strong-coupling limit, where the atom–surface distance is d. Again, the path average here is the probability of a Brownian bridge pinned at time $2\mathcal{T}$ to touch the surface, which is $\exp(-d^2/\mathcal{T})$, as we argued in Eq. (21.139). Thus,

$$\begin{aligned} V_{\text{dipole}} &= -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3}\sum_{n=-\infty}^{\infty} e^{-n^2\hbar^2c^2\beta^2/4\mathcal{T}}e^{-d^2/\mathcal{T}} \\ &= -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3}\sum_{n=-\infty}^{\infty} e^{-(n^2\hbar^2c^2\beta^2+4d^2)/4\mathcal{T}} \\ &= -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0 d^4}\sum_{n=-\infty}^{\infty}\frac{1}{(1+n^2\hbar^2c^2\beta^2/4d^2)^2} \\ &= -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0 d^4}\left(\frac{2\pi d}{\hbar c\beta}\right)\coth\left(\frac{2\pi d}{\hbar c\beta}\right), \end{aligned} \tag{21.158}$$

where we have again used the integral formula (21.132). Writing out the explicit temperature, our result is

$$V_{\text{dipole}} = -\frac{\hbar c \alpha_0}{(4\pi\epsilon_0)16\pi d^4}\left(\frac{2\pi d k_{\text{B}}T}{\hbar c}\right)\coth\left(\frac{2\pi d k_{\text{B}}T}{\hbar c}\right).$$

(atom–surface potential, nonzero T, scalar result) (21.159)

As $T \longrightarrow 0$, this expression reduces to the former result (21.127), since $\lim_{x\to 0} x\coth x = 1$. As $T \longrightarrow \infty$ we use $\lim_{x\to\infty}\coth x = 1$ to obtain

$$V_{\text{dipole}} = -\frac{k_{\text{B}}T\alpha_0}{(4\pi\epsilon_0)8d^3}$$

(atom–surface potential, high T, scalar result) (21.160)

Interestingly, this is half of the electromagnetic result (14.344), rather than 1/6 as in the zero-temperature limit.

21.2 Worldlines and the Relativistic Scalar Particle

To wrap up the worldline calculation of Casimir energies of scalar particles, we will review some of the general theory of classical and quantum scalar particles. This will clarify some aspects of the form of the worldline path integral, as well as the ‘‘worldline’’ nomenclature for the partition-function path integral.

21.2.1 Action

In general relativity, the trajectories of point particles correspond to geodesic motion. In flat space, we will take this to mean that the **proper time**

$$\tau = \int d\tau \tag{21.161}$$

is extremized, where in one spatial dimension,

$$(d\tau)^2 = (dt)^2 - (dx/c)^2. \tag{21.162}$$

Note that τ is defined up to an arbitrary constant, unless the integral is taken between two events, in which case we interpret τ as a **proper-time interval**. This is the same as the space-time interval s up to a factor of c. Thus,

$$\tau = \int \sqrt{dt^2 - dx^2/c^2} = \int dt \sqrt{1 - \frac{\dot{x}^2}{c^2}}. \tag{21.163}$$

(Usually, this geodesic principle is given in terms of the space-time interval ds, where $ds^2 = -d\tau^2$, in which case $|ds^2|$ enters under the square root to maintain a real integrand for "timelike" particle motion.) We will then take the action for the relativistic particle to be

$$S = \alpha \int dt \sqrt{1 - \frac{\dot{x}^2}{c^2}}, \tag{21.164}$$

where α is some constant (with units of energy) yet to be determined. In addition, we can add in a background potential $V(x,t)$ by tacking it on as in the nonrelativistic action:[6]

$$S = \int dt \left[\alpha \sqrt{1 - \frac{\dot{x}^2}{c^2}} - V(x,t) \right]. \tag{21.165}$$

We can then identify the bracketed quantity as the Lagrangian $L(x, \dot{x})$. Computing the conjugate momentum,

$$p = \frac{\partial L}{\partial \dot{x}} = -\alpha \left(1 - \frac{\dot{x}^2}{c^2} \right)^{-1/2} \frac{\dot{x}}{c^2}. \tag{21.166}$$

For this to coincide with the usual relativistic momentum, we should choose $\alpha = -mc^2$, and thus

$$p = \frac{m\dot{x}}{\sqrt{1 - \dot{x}^2/c^2}}. \tag{21.167}$$

The Euler–Lagrange equation

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = 0 \tag{21.168}$$

then simply yields

$$\dot{p} = -\frac{\partial V}{\partial x}, \tag{21.169}$$

as we expect in the nonrelativistic limit.

Turning to the Hamiltonian, we obtain

$$\begin{aligned} H(x,p) &= p\dot{x} - L \\ &= p\dot{x} + mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}} + V(x,t) \\ &= \frac{1}{mc^2} \sqrt{1 - \frac{\dot{x}^2}{c^2}} \left(p^2c^2 + m^2c^4 \right) + V(x,t). \end{aligned} \tag{21.170}$$

Solving Eq. (21.167) for $\dot{x}$ gives

$$\frac{\dot{x}^2}{c^2} = \frac{p^2c^2}{p^2c^2 + m^2c^4}, \tag{21.171}$$

or

$$1 - \frac{\dot{x}^2}{c^2} = \frac{m^2c^4}{p^2c^2 + m^2c^4}, \tag{21.172}$$

[6] Herbert Goldstein, Charles Poole, and John Safko, *Classical Mechanics*, 3rd ed. (Addison Wesley, 2001), Section 7.9, p. 312.

so that the Hamiltonian (21.170) becomes

$$H(x,p) = \sqrt{p^2c^2 + m^2c^4} + V(x,t), \tag{21.173}$$
(relativistic-particle Hamiltonian)

which is the usual expression for the relativistic energy, with an external potential included. Then to summarize, we have the relativistic action and Lagrangian

$$\begin{aligned} S[x] &= \int dt\, L(x,\dot{x}) \\ L(x,\dot{x}) &= -mc^2\sqrt{1 - \frac{\dot{x}^2}{c^2}} - V(x,t). \end{aligned} \tag{21.174}$$
(relativistic-particle Lagrangian and action)

Note that if T denotes the particle kinetic energy, then we can deduce T from the Hamiltonian by subtracting the rest mass mc^2 and the potential:

$$T = \sqrt{p^2c^2 + m^2c^4} - mc^2. \tag{21.175}$$
(relativistic-particle kinetic energy)

Note that expanding to order p^2 gives the usual nonrelativistic result $T = p^2/2m$. This amounts to assuming the Hamiltonian is of the form

$$H(x,p) = T(p) + V(x) + mc^2, \tag{21.176}$$
(relativistic-particle Hamiltonian)

which is a constant of the motion if V does not depend explicitly on time. This then amounts to a definition for T. Also, if we take the Hamiltonian $H(x,p) = E$ to be constant, in the free-particle ($V = 0$) case, Eq. (21.173) becomes

$$E^2 - p^2c^2 = mc^2, \tag{21.177}$$
(mass-shell condition)

which is called the **mass-shell condition**. This appears as a constraint on the momentum, so in the one-dimensional case, the momentum is not an independent variable.

In the relativistic case, note that the Lagrangian (21.174) is *not* of the form $T - V$. However, if we define

$$\tilde{T} = -mc^2\sqrt{1 - \frac{\dot{x}^2}{c^2}} = -\frac{m^2c^4}{\sqrt{p^2c^2 + m^2c^4}}, \tag{21.178}$$

then $L = \tilde{T} - V$, which is as close as we will get to the "standard" nonrelativistic form for the Lagrangian.

21.2.2 Reparameterization Independence

Right now, the action (21.174) is parameterized in terms of the local-time coordinate t. We can also write it in terms of the proper-time parameter as

$$S[x,t] = -mc^2 \int d\tau \left[\sqrt{\left(\frac{dt}{d\tau}\right)^2 - \frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2} - \frac{dt}{d\tau} V(x,t) \right]. \tag{21.179}$$

In this form, we see that the action has a particular property. If we introduce a (bijective) function $\lambda(\tau)$ that acts as a rescaled time (possibly rescaled in a nonlinear way), then using

$$d\lambda = \left(\frac{d\lambda}{d\tau}\right) d\tau, \tag{21.180}$$

the action has the same form:

$$S[x,t] = -mc^2 \int d\lambda \left[\sqrt{\left(\frac{dt}{d\lambda}\right)^2 - \frac{1}{c^2}\left(\frac{dx}{d\lambda}\right)^2} - \frac{dt}{d\lambda} V(x,t) \right]. \tag{21.181}$$

The point is that the action is invariant under arbitrary rescaling of the time parameter, or that it is **reparameterization-independent**.

Reparameterization independence is an important concept in relativity. The idea is that the action S should characterize the **world line** of a particle—the set of all space-time events occupied by the particle. Recall that space-time events are observer-independent, and thus manifestly independent of any coordinate system used to describe them. The world line itself is thus a geometric object, independent of a particular choice of coordinate system. Since the world line fully determines the action (and vice versa), the action should also be described only in terms of the world-line geometry, and it should not include any "artifacts" introduced by a coordinate system. Hence the importance of reparameterization independence.

In this sense, reparameterization we have something like a gauge freedom: choosing different proper-time-like parameters are analogous to different gauge choices that lead to the same physical quantities (i.e., the world line). We will see the significance of this soon when we treat this gauge invariance explicitly in the action.

21.2.3 Quadratic Action

The action (21.174) and Hamiltonian (21.173) are perfectly valid, and we have seen that the action has the advantage of being independent of the world-line parameterization. However, the obvious feature is the square root, which becomes awkward in some calculations, particularly when carrying the Hamiltonian over to quantum mechanics. Now, using a proper-time parameterization, we will seek to write down a variational principle that is more similar to the nonrelativistic case, hiding the awkward square root. We will begin by replacing Eq. (21.164) with the action principle

$$S = \alpha \int d\tau = \alpha \int \frac{dt^2 - dx^2/c^2}{d\tau^2}\, d\tau = \alpha \int \left[\left(\frac{dt}{d\tau}\right)^2 - \frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2 \right] d\tau = \alpha \int \left[\dot{t}^2 - \frac{\dot{x}^2}{c^2} \right] d\tau \tag{21.182}$$

where we are explicitly maintaining the action integral in terms of the proper time τ instead of the local time t, and α is again an undetermined (and potentially a different) constant parameter, with dimensions of energy. Then, while we will refrain for the moment from introducing a background potential $V(x)$, we will introduce an energy offset of $mc^2/2$, with the form of a potential energy, whose purpose will become clear later:

$$S = \int \left[\alpha\left(\dot{t}^2 - \frac{\dot{x}^2}{c^2} \right) - \frac{1}{2}mc^2 \right] d\tau. \tag{21.183}$$

The momentum conjugate to the spatial coordinate x is

$$p_x = \frac{\partial L}{\partial \dot{x}} = -\frac{2\alpha}{c^2}\dot{x} = -\frac{2\alpha}{c^2}\frac{dx}{d\tau}, \tag{21.184}$$

while the momentum conjugate to the local time is

$$p_t = \frac{\partial L}{\partial \dot{t}} = 2\alpha\dot{t} = 2\alpha\frac{dt}{d\tau}. \tag{21.185}$$

For the spatial momentum to coincide with the usual relativistic momentum $p_x = m(dx/d\tau)$, we then require

$$\alpha = -\frac{mc^2}{2}. \tag{21.186}$$

Then the conjugate momenta are

$$p_x = m\frac{dx}{d\tau}, \qquad p_t = -mc^2\frac{dt}{d\tau}. \tag{21.187}$$

The Euler–Lagrange equation for the action (21.183) then gives

$$\frac{d^2t}{d\tau^2} = 0, \qquad \frac{d^2x}{d\tau^2} = 0. \tag{21.188}$$

(Euler–Lagrange equations)

The two equations together imply straight world lines. The first equation in particular relates local time and proper time. In particular, it states that $dt/d\tau = a$ for some constant a. Then inverting this and writing $d\tau = \sqrt{dt^2 - dx^2/c^2}$, we find $1/a = \sqrt{1 - \dot{x}^2/c^2}$, and thus

$$d\tau = dt\sqrt{1 - \frac{\dot{x}^2}{c^2}}. \tag{21.189}$$

This is the usual relation between local and proper time. However, this came from inserting the definition (21.162) of the proper time, not from the dynamical equation.

The Hamiltonian in this case is

$$\begin{aligned} H &= p_x\dot{x} + p_t\dot{t} - L \\ &= p_x\dot{x} + p_t\dot{t} - \frac{m}{2}\dot{x}^2 + \frac{mc^2}{2}\dot{t}^2 + \frac{mc^2}{2}, \end{aligned} \tag{21.190}$$

or in canonical coordinates,

$$H(x, t, p_x, p_t; \tau) = \frac{p_x^2}{2m} - \frac{p_t^2}{2mc^2} + \frac{mc^2}{2}. \tag{21.191}$$

Using the conjugate momenta (21.187) and (21.189), we can see that

$$\frac{p_x^2}{2m} - \frac{p_t^2}{2mc^2} = -\frac{mc^2}{2}, \tag{21.192}$$

and thus

$$H(x, t, p_x, p_t; \tau) = 0. \tag{21.193}$$

In the absence of the external potential, the Hamiltonian always has the value $H = 0$, which is characteristic of the ''extended'' phase space (including the temporal degree of freedom and the proper time as the new time). This is the reason for introducing the $-mc^2/2$ in the Lagrangian; otherwise the Hamiltonian could have some other (arbitrary) constant value.

Then to summarize the action and Lagrangian here (after dropping the external potential), we have

$$\begin{aligned} S[x] &= \int d\tau\, L(x, t, \dot{x}, \dot{t}; \tau) \\ L(x, t, \dot{x}, \dot{t}; \tau) &= \frac{1}{2}m\dot{x}^2 - \frac{1}{2}mc^2\dot{t}^2 - \frac{1}{2}mc^2, \end{aligned}$$

(relativistic-particle Lagrangian and action, proper-time parameterization) (21.194)

with Hamiltonian

$$H(x, t, p_x, p_t; \tau) = \frac{p_x^2}{2m} - \frac{p_t^2}{2mc^2} + \frac{mc^2}{2}.$$

(Hamiltonian, proper-time parameterization) (21.195)

For this action principle, the parameterization can be changed, but an alternate parameter λ will only work as a parameter for the Euler–Lagrange equations, provided $\lambda = a\tau + b$ for some constants a and b. We can see this by writing out

$$\frac{d^2x}{d\lambda^2} = \frac{d}{d\lambda}\frac{dx}{d\lambda} = \frac{d\tau}{d\lambda}\frac{d}{d\tau}\frac{d\tau}{d\lambda}\frac{dx}{d\tau} = \frac{d^2x}{d\tau^2}, \tag{21.196}$$

where the final equality only holds if $d\tau/d\lambda$ is constant. In this case λ is called an **affine parameter**.[7]

Note that had we introduced a background-potential term of the form $\int d\tau\, V(x)$ [or $\int dt\, V(x)$, as in the square-root action (21.174)], in the action (21.183), we would have arrived at the odd-looking result $H = V(x)$ in place of the null-value Hamiltonian (21.193). This says that introducing a potential in this way is somewhat unnatural—remember that if the Hamiltonian contains no explicit reference to a time parameter, then the Hamiltonian is a constant of the motion. We will see shortly how to introduce a potential in a way consistent with the simpler result $H = 0$.

[7] The actions (21.174) and (21.194) impose different requirements on reparameterizations in the free-particle case; in particular the former allows any parameterization, while the latter admits only affine parameters. For more details, see Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler, *Gravitation* (W. H. Freeman, 1973), p. 322, Boxes 13.2 and 13.3.

21.2.3.1 Massless-Particle World Lines

While the action (21.194) works fine for *massive* relativistic particles, it is problematic when it comes to *massless* particles. [Note that the first action (21.174) is similarly problematic for massless particles.] This is due both to the presence of the mass in the kinetic-energy terms, as well as the parameterization by the proper time (the change in proper time is always zero along a light cone). We can cure this by letting $\tau \longrightarrow m\tau/c^2$ in the action integral. This amounts to parameterizing the paths by $\lambda = c^2\tau/m$, which remains well-defined in the limit $m \longrightarrow 0$. In this case, the action becomes

$$S[x] = \int d\lambda\, L_m(x, t, \dot{x}, \dot{t}; \lambda), \tag{21.197}$$

(reparameterized action)

where we are defining the rescaled Lagrangian

$$L_m(x, t, \dot{x}, \dot{t}; \lambda) := \frac{1}{2c^2}\dot{x}^2 - \frac{1}{2}\dot{t}^2 - \frac{m^2c^4}{2}, \tag{21.198}$$

(reparameterized Lagrangian)

and $\dot{x}$ and $\dot{t}$ now refer to derivatives with respect to λ. The rescaled version Hamiltonian (21.191) is then

$$H_m(x, t, p_x, p_t; \lambda) = \frac{p_x^2c^2}{2} - \frac{p_t^2}{2} + \frac{m^2c^4}{2}, \tag{21.199}$$

(reparameterized Hamiltonian)

where $p_x = \dot{x}/2c^2$ and $p_t = -\dot{t}/2$. Everything here is well-defined as $m \longrightarrow 0$. Note, though, that both Hamiltonian and Lagrangian here have dimensions of square energy. Note that from Eq. (21.193), we still have $H_m = 0$, and the Hamilton and Euler-Lagrange equations are equivalent to the ones before, but the Hamilton equations have a slightly different form, owing to different canonical momenta,

$$p_x = \frac{1}{c^2}\frac{dx}{d\tau}, \qquad p_t = \frac{dt}{d\tau}. \tag{21.200}$$

which are equivalent to the former momenta (21.187) under the same rescaling.

21.2.3.2 Variable-Mass Potential

With this rescaled Hamiltonian, we can introduce an external potential by regarding the mass as a space-dependent quantity. If we allow the mass to vary in space, separating constant and variable components via

$$m(x) = m_0 + \delta m(x), \tag{21.201}$$

then we may assign the space-dependent part to a potential via

$$V(x) := \frac{1}{2}\Big[2m_0\delta m(x) + \delta m^2(x)\Big]c^4, \tag{21.202}$$

which again has dimensions of squared energy. Then the Hamiltonian (21.199) becomes

$$H_m(x, t, p_x, p_t; \lambda) = \frac{p_x^2c^2}{2} - \frac{p_t^2}{2} + \frac{m_0^2c^4}{2} + V(x), \tag{21.203}$$

(reparameterized Hamiltonian with variable-mass potential)

and the Lagrangian (21.198) becomes

$$L_m(x, t, \dot{x}, \dot{t}; \lambda) = \frac{1}{2c^2}\dot{x}^2 - \frac{1}{2}\dot{t}^2 - \frac{m_0^2c^4}{2} - V(x). \tag{21.204}$$

(reparameterized Lagrangian)

Now both have a suitable potential that keeps a constant (null) value of the Hamiltonian. Note that in implementing the mass rescaling of the proper time to obtain these functions, in the case of a space-dependent mass, the reparameterization had to be done on a trajectory-dependent basis.

Then the new Hamilton equations are

$$\frac{dp_x}{d\tau} = -\frac{\partial V}{\partial x}, \qquad \frac{dp_t}{d\tau} = 0, \tag{21.205}$$

where the first equation gives the force law, which is slightly different from the earlier law (21.169), due to a difference in factor of $dt/d\tau$, and in the background mass m_0 not entering the force law here.

21.2.4 Constrained Action

Now, informed by our attempts above, let's rederive a quadratic action for the relativistic particle, but now keeping both a sensible $m \longrightarrow 0$ limit, *and* making sure the action is (generally) independent under reparameterizations. Thus, let's first start with a Lagrangian of the form (21.204),

$$L(x, t, \dot{x}, \dot{t}; \tau) = \frac{1}{2c^2}\dot{x}^2 - \frac{1}{2}\dot{t}^2 - \frac{m^2c^4}{2}, \tag{21.206}$$

and corresponding action

$$S[x, t] = \frac{1}{2}\int d\tau \left[\frac{1}{c^2}\left(\frac{dx}{d\tau}\right)^2 - \left(\frac{dt}{d\tau}\right)^2 - m^2c^4\right]. \tag{21.207}$$

For simplicity, we are temporarily ignoring the external potential (i.e., lumping it into the mass m). Again, the problem that we saw is that this action is not generally reparameterization-independent: it is only independent under affine parameter changes. To see this, if we introduce a new time parameter $\lambda = \lambda(\tau)$, such that $d\lambda = (d\lambda/d\tau)\, d\tau =: \dot{\lambda}\, d\tau$, and $\dot{\lambda}(d/d\lambda) = (d/d\tau)$, then we have

$$S[x, t] = \frac{1}{2}\int \frac{d\lambda}{\dot{\lambda}} \left[\frac{\dot{\lambda}^2}{c^2}\left(\frac{dx}{d\lambda}\right)^2 - \dot{\lambda}^2\left(\frac{dt}{d\lambda}\right)^2 - m^2c^4\right] = \frac{1}{2}\int d\lambda \left[\frac{\dot{\lambda}}{c^2}\left(\frac{dx}{d\lambda}\right)^2 - \dot{\lambda}\left(\frac{dt}{d\lambda}\right)^2 - \frac{1}{\dot{\lambda}}m^2c^4\right]. \tag{21.208}$$

Note that this has the same form as before only if $\dot{\lambda} = 1$, which is a stronger requirement than λ being an affine parameter. More generally, to preserve the form of the action, we can introduce an "**einbein**" function $e(\tau)$ in the action, so that it reads

$$S[x, t, e] = \frac{1}{2}\int d\tau \left[\frac{1}{ec^2}\left(\frac{dx}{d\tau}\right)^2 - \frac{1}{e}\left(\frac{dt}{d\tau}\right)^2 - em^2c^4\right].$$

(relativistic-particle action with einbein) (21.209)

Now recalling that a time reparameterization is basically a gauge freedom, the "gauge transformation" here is that we are switching from τ to $\lambda(\tau)$ as the parameter, and to keep the form of the action, we must make the simultaneous parameter and einbein replacements

$$\tau \longrightarrow \lambda(\tau), \qquad e(\tau) \longrightarrow e_\lambda[\lambda(\tau)] := e(\tau)\left(\frac{d\lambda}{d\tau}\right)^{-1}, \tag{21.210}$$

(gauge freedom)

under which the form of the action is explicitly invariant.

As far as the action (21.209) is concerned, now e appears as an extra variable, but note that the action is explicitly independent of $\dot{e}$, which signals the presence of a constraint (i.e., the gauge freedom). The Euler-Lagrange equation for e gives

$$\frac{\delta S}{\delta e} = -\frac{\dot{x}^2}{e^2c^2} + \frac{\dot{t}^2}{e^2} - m^2c^4 = 0, \tag{21.211}$$

which reduces to the constraint equation

$$\frac{\dot{x}^2}{c^2} - \dot{t}^2 + e^2m^2c^4 = 0. \tag{21.212}$$

That is, on any given world line, e is determined by the coordinates along the world line (in particular, the velocities). Solving for e,

$$e = \frac{\sqrt{\dot{t}^2 - \dot{x}^2/c^2}}{mc^2}, \tag{21.213}$$

and putting this into the action (21.209)

$$S[x,t,e] = -mc^2 \int d\tau \sqrt{\dot{t}^2 - \dot{x}^2/c^2}. \tag{21.214}$$

This is identical to the first (square-root) action (21.174), with $V = 0$.

To get a bit more intuition for the constraint (21.212), first note that the conjugate momenta from the action (21.209) are

$$p_x = \frac{\dot{x}}{ec^2}, \qquad p_t = -\frac{\dot{t}}{e}. \tag{21.215}$$

Then we can write Eq. (21.212) as

$$p_t^2 - p_x^2 c^2 = m^2 c^4, \tag{21.216}$$
(mass-shell condition)

which is known as the **mass-shell condition**. From the form of the Hamiltonian (21.173) with $V = 0$, we usually identify $p^2c^2 + m^2c^4$ with the square of the relativistic energy E, as in Eq. (21.177), in which case we have the more familiar form

$$E^2 - p_x^2 c^2 = m^2 c^4 \tag{21.217}$$
(mass-shell condition)

for the mass-shell condition. In any case this acts as a constraint on p_t, showing that it is not independent from p_x.

21.2.4.1 Gauge Fixing

From this discussion, it seems that we may just as well "gauge fix" the einbein at $e = 1$ and ignore it in the development of the action, which is essentially what we did in the first quadratic action that we developed in Section 21.2.3. That is, what do we get by introducing the einbein e? The reason is the necessity of having the constraint (21.216), which comes from the gauge freedom of e. This is analogous to electromagnetism, in which it is fine to fix a particular gauge, but we must also make sure to implement Gauss' law as a constraint—a constraint that arose from the gauge freedom of electromagnetic.

To see that we don't naturally get the constraint (21.216), from the gauge-fixed theory, consider the Euler–Lagrange equations corresponding to the action (21.209). First, the x equation gives

$$\frac{d}{d\tau}\left(\frac{1}{e}\frac{dx}{d\tau}\right) = 0, \tag{21.218}$$

and with the momentum p_x in (21.215),

$$\frac{dp_x}{d\tau} = 0. \tag{21.219}$$

Similarly, for the temporal momentum,

$$\frac{dp_t}{d\tau} = 0. \tag{21.220}$$

This implies that p_x and p_t are constants of the motion, and thus that the combination $p_t^2 - p_x^2c^2$ from Eq. (21.216) is *constant*. However, it does not tell us *what* that constant is; that comes only from an explicit treatment of the gauge freedom. In the gauge-fixed version, this value must be supplied as *extra* information, in the form of an initial value for this constant of the motion. This can also come in the form of assuming a proper-time parameterization, as we did in Section 21.2.3.

21.2.4.2 Terminology

Finally, before continuing, it's worth noting why e is called an "einbein." The reason is that under the reparameterization $\tau \longrightarrow \lambda$, the coordinates transform like scalars:

$$x^\mu(\tau) = \tilde{x}^\mu(\lambda). \tag{21.221}$$

However, from Eq. (21.210), the einbein transforms as

$$e(\tau)\, d\tau = \tilde{e}(\lambda)\, d\lambda, \tag{21.222}$$

where the twiddles indicate the λ parameterization. Then e transforms like a one-form (i.e., it transforms covariantly) in one dimension, hence the name "einbein."

21.2.4.3 Variable-Mass Potential

We can now introduce a variable-mass potential, as in the gauge-fixed treatment of Section 21.2.3.2. Thus again if

$$m(x) = m_0 + \delta m(x), \tag{21.223}$$

and we associate the space-dependent part of the mass with a potential,

$$V(x) := \frac{1}{2}\Big[2m_0\delta m(x) + \delta m^2(x)\Big]c^4, \tag{21.224}$$

then we obtain the action

$$\begin{aligned} S[x,t,e] &= \int d\tau\, L(x,t,e;\tau) \\ L(x,t,e;\tau) &= \frac{1}{2ec^2}\left(\frac{dx}{d\tau}\right)^2 - \frac{1}{2e}\left(\frac{dt}{d\tau}\right)^2 - \frac{em_0^2c^4}{2} - eV(x). \end{aligned}$$

(relativistic-particle action with einbein and variable mass) (21.225)

Note that in doing this, we still obey the mass-shell constraint (21.216), and we do not have any awkward results with a nonconstant Hamiltonian value as we saw in Section section:relativistic-ptte-quadratic-action.

Now let's work out the Euler–Lagrange equations. The equation for the local time is the same as in the constant-mass case, Eq. (21.220). The x equation is more interesting, however. Writing out the result for $\partial L/\partial x - (d/d\tau)\partial L/\partial\dot{x} = 0$,

$$\frac{d}{d\tau}\left(\frac{1}{ec^2}\frac{dx}{d\tau}\right) = -e\frac{dV(x)}{dx}. \tag{21.226}$$

Using the momentum p_x in Eqs. (21.215), we can write this out compactly in the form

$$\frac{dp_x}{d\tau} = -e\frac{dV(x)}{dx}. \tag{21.227}$$

We can also write out the equation of motion in a more explicit, but less compact form, by writing out (but not gauge-fixing) the einbein. Using the value (21.213) for e in the form

$$e = \frac{\sqrt{(dt/d\tau)^2 - (dx/d\tau)^2/c^2}}{mc^2} = \frac{dt}{d\tau}\frac{\sqrt{1-(dx/dt)^2/c^2}}{mc^2}, \tag{21.228}$$

which is still valid provided we interpret m as the space-dependent mass, we find

$$\frac{d}{dt}\left(\frac{m(dx/dt)}{\sqrt{1-(dx/dt)^2/c^2}}\right) = -\frac{\sqrt{1-(dx/dt)^2/c^2}}{mc^2}\frac{dV(x)}{dx}. \tag{21.229}$$

Note that the left-hand side here is the same as in the analogous result for the square-root action with external potential, Eq. (21.169). However, the potential couples in slightly differently. There is a factor of mc^2 on

the right-hand side, which makes the dimensions come out correctly. Also, there is a Lorentz-contraction factor on the right-hand side, which amounts to the same factor $dt/d\tau$ that appears in the square-root action (21.179). Not to mention, the mass here is space-dependent, so a better analogous expression may be to put the momentum on the left-hand side in terms of the mass offset m_0:

$$\frac{d}{dt}\left(\frac{m_0(dx/dt)}{\sqrt{1-(dx/dt)^2/c^2}}\right) = -\frac{\sqrt{1-(dx/dt)^2/c^2}}{[m^2(x)/m_0]c^2}\frac{dV(x)}{dx}. \tag{21.230}$$

This puts all the spatial dependence of the force on the right-hand side. In principle, then, we can choose a mass function $m(x)$ to emulate whatever mass and force law we like in the square-root-action equation of motion, Eq. (21.169).

Note that in this form of the action, *any* external potential we tack on in a parameterization-independent way (i.e., that includes a factor of e) has the form of a space-dependent mass. Again, the nonconstant Hamiltonian value as we saw in Section 21.2.3 shows that introducing an external potential in a form that does *not* modulate the mass introduces problems.

21.2.5 Klein–Gordon Equation

The **Klein–Gordon equation** is the wave equation corresponding to the relativistic particle. The motivation for the Klein–Gordon equation starts[8] with the relativistic energy

$$E = \sqrt{p^2c^2 + m^2c^4}. \tag{21.231}$$

If we take this to be the Hamiltonian, apply the usual Schrödinger equation via $i\hbar\partial_t \equiv H$ with the momentum identification $p \equiv -i\hbar\partial_x$ (in one dimension), then we obtain

$$i\hbar\partial_t\phi = \sqrt{-\hbar^2c^2\partial_x^2 + m^2c^4}\,\phi. \tag{21.232}$$

However, the square root here is awkward, as its Taylor expansion implies the presence of derivatives at all orders, and thus an inconveniently nonlocal wave equation.

The Klein–Gordon alternative is to try the *square* of the identification $i\hbar\partial_t \equiv H$, which gives $-\hbar^2\partial_t^2 \equiv H^2$. The corresponding wave equation is

$$\left(\partial_x^2 - \frac{1}{c^2}\partial_t^2\right)\phi = \left(\frac{mc}{\hbar}\right)^2\phi, \tag{21.233}$$

(Klein–Gordon equation)

after a bit of rearrangement. This is the Klein–Gordon equation. The cost of getting rid of the square root is to introduce negative-energy solutions of the form $E = -\sqrt{p^2c^2 + m^2c^4}$. These have the interpretation of solutions propagating *backwards* in time (or equivalently, *antiparticles*). Again, we can add in a background potential $V(x)$,

$$\left(\partial_x^2 - \frac{1}{c^2}\partial_t^2\right)\phi = \left(\frac{m_0c}{\hbar}\right)^2\phi + \frac{2}{\hbar^2c^2}V(x)\,\phi, \tag{21.234}$$

(Klein–Gordon equation, with background potential)

by regarding the mass $m(x)$ to be space-dependent.

21.2.6 Path Integral

One approach to developing the path integral for the Klein–Gordon equation (21.233) is to start with the action (21.174) corresponding to the relativistic particle, and assume that the nonrelativistic form (20.24)

$$K(x,t;x_0,t_0) = \int Dx\, \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt\, L(x,\dot{x})\right] = \int Dx\, \exp\left[\frac{i}{\hbar}\int_{t_0}^{t} dt\, S[x]\right] \tag{21.235}$$

[8] James D. Bjorken and Sidney D. Drell, *Relativistic Quantum Mechanics* (McGraw–Hill, 1964), Chapter 1.

for the propagator carries over. Unfortunately, the square root in the action here is again problematic, and leads to non-Gaussian integrals. It is possible to pursue this, and obtain an explicit expression for the short-time propagator in terms of a Bessel function.[9] Whether one chooses to go this way or to simply object to the use of the exponentiated action for this propagator,[10] clearly this is not the ''world-line'' path integral that we seek here.

21.2.6.1 Proper-Time Path Integral

An alternate approach to the path-integral quantization of the Klein–Gordon equation is due to Feynman.[11] Suppose that we introduce an auxiliary ''energy'' parameter E in in the Klein–Gordon equation (21.233),

$$\left(-\frac{\hbar^2 c^2}{2}\partial_x^2 + \frac{\hbar^2}{2}\partial_t^2 \right)\phi + \frac{m_0^2 c^4}{2}\phi + V(x)\,\phi = E\phi, \tag{21.236}$$

such that we recover the original equation by setting $E = 0$. (Note that the mass m_0 can also take the role of E, but introducing the extra parameter allows us to leave the effective Hamiltonian for the Klein–Gordon equation intact.) Then defining the ''E-rotating'' wave function by

$$\varphi := \phi\, e^{-iE\mathcal{T}}, \tag{21.237}$$

we can write the Klein–Gordon equation in terms of this variable as

$$i\partial_{\mathcal{T}}\varphi = \left(-\frac{\hbar^2 c^2}{2}\partial_x^2 + \frac{\hbar^2}{2}\partial_t^2 \right)\varphi + \frac{m_0^2 c^4}{2}\varphi + V(x)\,\varphi. \tag{21.238}$$

This has the form of the Schrödinger equation with ''time'' $\mathcal{T}$ (which has dimensions of inverse-squared energy), which acts something like a ''proper time'' with respect to the other space-time coordinates (x, t). In the solution of this form of the equation, $\varphi(x, t; \mathcal{T})$ contains contributions from all values of E. Since the equation is linear in the wave function, we can write the solution as a superposition

$$\varphi(x, t; \mathcal{T}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dE\, \tilde{\varphi}(x, t; E)\, e^{-iE\mathcal{T}}, \tag{21.239}$$

where we are interested only in the $E = 0$ component of solution $\tilde{\varphi}(x, t; E = 0)$, which corresponds to the solution of the original equation (21.233). We can obtain the component for any E by a simple Fourier projection:

$$\phi(x, t) = \int_{-\infty}^{\infty} d\mathcal{T}\, \varphi(x, t; \mathcal{T})\, e^{iE\mathcal{T}}. \tag{21.240}$$

The case $E = 0$ becomes the simple integral

$$\phi(x, t) = \int_{-\infty}^{\infty} d\mathcal{T}\, \varphi(x, t; \mathcal{T}), \tag{21.241}$$

which we can easily verify by

$$\phi(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\mathcal{T} \int_{-\infty}^{\infty} dE\, \tilde{\varphi}(x, t; E)\, e^{-iE\mathcal{T}} = \int_{-\infty}^{\infty} dE\, \tilde{\varphi}(x, t; E)\, \delta(E) = \tilde{\varphi}(x, t; E = 0). \tag{21.242}$$

Thus, by solving the extended equation (21.238), we can in principle obtain the propagator for the Klein–Gordon equation (21.233).

[9] Ian H. Redmount and Wai–Mo Suen, ''Path integration in relativistic quantum mechanics,'' *International Journal of Modern Physics A* **8**, 1629 (1993) (doi: 10.1142/S0217751X93000667), arXiv.org preprint (arXiv: gr-qc/9210019).

[10] for a rather critical account, see Arlen Anderson, ''Use of $\exp(iS[x])$ in the sum over histories,'' *Physical Review D* **49**, 4049 (1994) (doi: 10.1103/PhysRevD.49.4049).

[11] R. P. Feynman, ''Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction,'' *Physical Review* **80**, 440 (1950) (doi: 10.1103/PhysRev.80.440); L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, 1981), Chapter 25.

To do this, we will be careful and use the language of Green functions and resolvent operators from Chapter 15. First, defining the effective Hamiltonian for the Klein–Gordon equation as

$$H = -\frac{\hbar^2 c^2}{2}\partial_x^2 + \frac{\hbar^2}{2}\partial_t^2 + \frac{m_0^2 c^4}{2} + V(x), \tag{21.243}$$

then the extended Klein–Gordon equation (21.236) becomes

$$H\phi = E\phi. \tag{21.244}$$

This has the form of a time-independent Schrödinger equation, which defines a retarded, energy-domain Green operator (resolvent operator) via [see Eq. (15.7)]

$$(E - H + i0^+)\tilde{G}^+_{\mathcal{T}}(E) = 1. \tag{21.245}$$

The tilde here emphasizes that the Green operator is in energy space, and the $\mathcal{T}$ subscript is a reminder that this is a Green operator for the extended Hamiltonian. Similarly, Eq. (21.238) is the time-domain version of the extended Schrödinger equation, which we may rewrite as

$$i\partial_{\mathcal{T}}\varphi = H\varphi. \tag{21.246}$$

This as well defines the retarded Green function (in proper time) via [Eq. (15.14)]

$$(i\partial_{\mathcal{T}} - H)G^+_{\mathcal{T}}(\mathcal{T}, 0) = i\delta(\mathcal{T}). \tag{21.247}$$

These two Green functions are related via the integral [Eq. (15.8)]

$$G^+_{\mathcal{T}}(E) = \frac{1}{i}\int_0^\infty d\mathcal{T}\, e^{i(E+i0^+)\mathcal{T}/\hbar}\, G^+_{\mathcal{T}}(\mathcal{T}, 0), \tag{21.248}$$

which for $E = 0$ becomes

$$\tilde{G}^+_{\mathcal{T}}(E = 0) = \frac{1}{i}\int_0^\infty d\mathcal{T}\, G^+_{\mathcal{T}}(\mathcal{T}, 0). \tag{21.249}$$

Now the propagator for the Klein–Gordon equation should satisfy Eq. (21.234) with a delta-function term added. Specifically, upon rescaling, we can write this as

$$\left(-\frac{\hbar^2}{2}\partial_t^2 + \frac{\hbar^2 c^2}{2}\partial_x^2 - \frac{m_0^2 c^4}{2} - V(x)\right) G^+(t, t_0) = i\delta(t - t_0),$$

(defining relation for Klein–Gordon propagator) (21.250)

in analogy with Eq. (21.247). Since the operator on the left-hand side is $-H$, by comparison with (21.245), we have

$$G^+(t, t_0) = i\langle t|\tilde{G}^+_{\mathcal{T}}(E = 0)|t_0\rangle. \tag{21.251}$$

Combination with Eq. (21.249) and writing out the position matrix elements gives the propagator for $t \geq t_0$:

$$K(x, t; x_0, t_0) = \int_0^\infty d\mathcal{T}\, \langle x, t|G^+_{\mathcal{T}}(\mathcal{T}, 0)|x_0, t_0\rangle. \tag{21.252}$$

The matrix element on the right-hand side is the propagator for the extended Klein–Gordon equation (21.238). Since this is formally a Schrödinger equation, we can write this in terms of the standard non-relativistic propagator. Adapted to this case from Eq. (20.24), the extended propagator for this solution reads

$$K_{\mathcal{T}}(x, t, \mathcal{T}; x_0, t_0, 0) = \int Dx(\tau)\, \exp\left[i\int_0^{\mathcal{T}} d\tau \left(\frac{\dot{x}^2}{2c^2} - \frac{\dot{t}^2}{2} - \frac{m_0^2 c^4}{2} - V(x)\right)\right], \tag{21.253}$$

where the Dx refers to integration over (x,t). Notice that the integrand in the exponential is just the Lagrangian $L_m(x,t,\dot{x},\dot{t};\lambda=\tau)$ from Eq. (21.204). Thus, we finally have

$$\begin{aligned} K(x,t;x_0,t_0) &= \int_0^\infty d\mathcal{T} \int_{(x_0,t_0)}^{(x,t)} Dx(\tau)\, \exp\left[i \int_0^{\mathcal{T}} d\tau \left(\frac{\dot{x}^2}{2c^2} - \frac{\dot{t}^2}{2} - \frac{m_0^2 c^4}{2} - V(x) \right) \right] \\ &= \int_0^\infty d\mathcal{T} \int_{(x_0,t_0)}^{(x,t)} Dx(\tau)\, \exp\left[i \int_0^{\mathcal{T}} d\tau\, L_m(x,t,\dot{x},\dot{t};\tau) \right] \end{aligned}$$

(world-line Klein–Gordon propagator) (21.254)

as the ''world-line'' path integral form for the Klein–Gordon equation.[12]

Essentially, this is a path integral over world lines that travel from (x_0,t_0) to (x,t) in proper time $\mathcal{T}$; then, we integrate over all possible proper times $\mathcal{T}$. In a more ''string-inspired'' approach, the world-line parameter $\mathcal{T}$ plays the role of a constraint, and the integral form here is effectively a particular choice of gauge.[13]

21.2.6.2 Imaginary Time

Imagining that we are to do statistical mechanics for the relativistic particle, by analogy with the nonrelativistic case, we can consider the propagator in imaginary (local) time:

$$K(x,-i\hbar\beta;x,0) = \int_0^\infty d\mathcal{T} \int_{(x_0,0)}^{(x,-i\hbar\beta)} Dx\, \exp\left[i \int_0^{\mathcal{T}} d\tau \left(\frac{(\partial_\tau x)^2}{2c^2} + \frac{(\partial_\tau \tilde{\beta})^2}{2} - \frac{m_0^2 c^4}{2} - V(x) \right) \right]. \tag{21.255}$$

Here, $\tilde{\beta} := \hbar\beta$ as usual. Further, going to imaginary *proper* time via the Wick rotation $\tau \longrightarrow -i\tau$ and $\mathcal{T} \longrightarrow -i\mathcal{T}$,

$$\tilde{K}(x,-i\hbar\beta;x,0) = -i \int_0^\infty d\mathcal{T} \int_{(x_0,0)}^{(x,-i\hbar\beta)} Dx\, \exp\left[- \int_0^{\mathcal{T}} d\tau \left(\frac{(\partial_\tau x)^2}{2c^2} + \frac{(\partial_\tau \tilde{\beta})^2}{2} + \frac{m_0^2 c^4}{2} + V(x) \right) \right]. \tag{21.256}$$

Note, however, that we cannot directly associate this with the *partition function* for the relativistic particle simply by tracing over x, because this identification was based on the analogy of the nonrelativistic evolution operator with the partition function. In particular, in imaginary (local and proper) times $t \longrightarrow -it$ and $\tau \longrightarrow -i\tau$, what appears is the Hamiltonian

$$H_m(x,t,p_x,p_t;\tau) = \frac{p_x^2 c^2}{2} + \frac{p_t^2}{2} + \frac{m_0^2 c^4}{2} + V(x) = \frac{p_x^2 c^2}{2} + \frac{p_t^2}{2} + \frac{m^2 c^4}{2} \tag{21.257}$$

[c.f. Eq. (21.203)]. To evaluate this, we recall that this was a reparameterized form of

$$H(x,t,p_x,p_t;\tau) = \frac{p_x^2}{2m} + \frac{p_t^2}{2mc^2} + \frac{mc^2}{2}, \tag{21.258}$$

where the proper-time parameter was scaled by a factor of m/c^2, with corresponding changes to the momenta (which we must restore to find proper factors of the mass that are otherwise hidden). The partition function corresponding to the imaginary-time propagator is then the trace of $\exp[-\beta H(x,t,p_x,p_t;\tau)]$, which might seem reasonable. However, using the conjugate momenta

$$p_x = \frac{1}{c}\frac{dx}{d\tau}, \qquad p_t = \frac{dt}{d\tau}, \tag{21.259}$$

[12] cf. C. Schubert, ''Perturbative quantum field theory in the string-inspired formalism,'' *Physics Reports* **355**, 73 (2001) (doi: 10.1016/S0370-1573(01)00013-8), arXiv.org preprint (arXiv: hep-th/0101036v2), especially Eq. (3.10), where the same propagator is given after evident Wick rotation and change to imaginary times.

[13] Philip R. Johnson, ''Relativistic Particle Trajectories from Worldline Path Integral Quantization,'' *Proceedings of the 2001 Particle Accelerator Conference*, 1781 (2001) (doi: 10.1109/PAC.2001.987181).

and using the conjugate momenta (21.187) and (21.189), we find

$$\frac{p_x^2}{2m} + \frac{p_t^2}{2mc^2} = \frac{p^2}{m} + \frac{mc^2}{2}, \tag{21.260}$$

with $p = p_x$ the momentum, now referred to the local time. Thus,

$$H(x, t, p_x, p_t; \tau) = \frac{p^2}{m} + mc^2. \tag{21.261}$$

This means that the parameterized form was

$$H_m(x, t, p_x, p_t; \tau) = p^2c^2 + m^2c^4, \tag{21.262}$$

which is the *square* of the relativistic energy. This would lead to a partition function with terms of the form $\exp(-\beta E_n^2)$, which is not directly useful. Nevertheless, this form of the imaginary-time Green function enters the world-line path integral for the scalar-field partition function, albeit with a different weighting of the integral over $\mathcal{T}$.

21.3 Exercises

Problem 21.1

Recalling that the mean energy is given by the partition-function derivative (21.13)

$$E = \langle H \rangle = -\partial_\beta \log Z, \tag{21.263}$$

show that the *second* derivative

$$\sigma_E^2 = \left\langle (H - \langle H \rangle)^2 \right\rangle = \partial_\beta^2 \log Z \tag{21.264}$$

gives the *variance* in the energy.

What does this say about fluctuations in Casimir energies at zero temperature?

Problem 21.2

The inhomogeneous Helmholtz equation

$$\left[\nabla^2 + k^2 - \frac{2}{\hbar^2 c^2} V(\mathbf{r})\right] \psi(\mathbf{r}) = -f(\mathbf{r}), \tag{21.265}$$

where $f(\mathbf{r})$ is an arbitrary source function and $V(\mathbf{r})$ is an added background-potential function, has a Green function (resolvent) defined by [see Eq. (14.57), noting that we are ditching the ϵ_0 but keeping the minus sign]

$$\left[-\nabla^2 - k^2 + \frac{2}{\hbar^2 c^2} V(\mathbf{r})\right] \tilde{G}(\mathbf{r}, \mathbf{r}'; k^2) = \delta^d(\mathbf{r} - \mathbf{r}'). \tag{21.266}$$

(See also Problem 15.2). Show that the *retarded* Green function has the world-line representation[14]

$$\tilde{G}^+(\mathbf{r}, \mathbf{r}'; k) = i \int_0^\infty dT\, e^{ik^2 T} \int_{\mathbf{r}}^{\mathbf{r}'} D\mathbf{x}(\tau) \exp\left[i \int_0^T d\tau \left(\frac{\dot{x}^2}{4} - \frac{2}{\hbar^2 c^2} V(\mathbf{x}) \right) \right], \tag{21.267}$$

or with the Wick rotations $T \longrightarrow -i\mathcal{T}$ and $k \longrightarrow i\kappa$ to the "Euclidean" form,

$$\tilde{G}^+(\mathbf{r}, \mathbf{r}'; -i\kappa) = \frac{1}{2} \int_0^\infty d\mathcal{T}\, e^{-\mathcal{T}\kappa^2/2} \int_{\mathbf{r}}^{\mathbf{r}'} D\mathbf{x}(\tau) \exp\left[- \int_0^{\mathcal{T}} d\tau \left(\frac{\dot{x}^2}{2} + \frac{1}{\hbar^2 c^2} V(\mathbf{x}) \right) \right], \tag{21.268}$$

[14] Marco Schäfer, Idrish Huet, and Holger Gies, "Energy-momentum tensors with worldline numerics," *International Journal of Modern Physics Conference Series* **14**, 511 (2012) (doi: 10.1142/S2010194512007647), arXiv.org preprint (arXiv: quant-ph/0605180v3).

after also rescaling $\mathcal{T}$ by a factor of 2. Both expressions should be interpreted in the renormalized sense of being compared to a "background-potential configuration" (e.g., with $V = 0$). Also, write down a normalized (Monte-Carlo) form of the Euclidean path integral in terms of Brownian bridges.

Problem 21.3

Show that the world-line form of the Casimir–Polder energy (before renormalization)

$$V_{\text{CP}}(\mathbf{r}) = \frac{\hbar c \alpha_0}{2(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar^2 c^2}\int_0^{\mathcal{T}} d\tau\, V[\mathbf{x}(\tau)]\right]\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{21.269}$$

for paths in D spacetime dimensions beginning and ending at $\mathbf{r}$, as is consistent with Eq. (21.118), can be obtained from the expression

$$V_{\text{CP}}(\mathbf{r}) = \frac{\hbar \alpha_0}{2\pi\epsilon_0 c^2} \int_0^\infty ds\, s^2\, \tilde{G}^+(\mathbf{r},\mathbf{r};is), \tag{21.270}$$

where $\tilde{G}^+(\mathbf{r},\mathbf{r};is)$ is the retarded Green function for the Helmholtz equation from Problem 21.2, and s is an imaginary frequency ($\omega = is$).

Problem 21.4

(a) Starting with the worldline path integral (21.109) for the Casimir energy, and using the calculation for the strong-coupling Casimir energy between two planar mirrors in Section 21.1.5.3, show that the following path integral yields the Casimir force per unit area for this configuration:

$$\frac{F}{A} = -\frac{\hbar c}{(2\pi)^{D/2}} \int dx_0 \left\langle\!\!\left\langle \frac{\Theta\{-\min[x(\tau)]\}}{d(x_0)\,\mathcal{T}_{\min}^{D/2}[x(\tau)]}\right\rangle\!\!\right\rangle_{x(\tau)}. \tag{21.271}$$

This expression implicitly assumes two mirrors located at $x = 0$ and $x = L$, and gives the force on the $x = L$ mirror. In the expression, d is the number of spatial dimensions, $D = \mathsf{d} + 1$ is the spacetime dimension, $d(x_0)$ is the distance from the path source point x_0 to the $x = L$ mirror, and $\mathcal{T}_{\min}[x(\tau)]$ is the smallest value of the path running time $\mathcal{T}$ such that the path touches the $x = L$ mirror. Note that the Heaviside function "fires" when the path $x(\tau)$ touches the $x = 0$ mirror.

(b) How does this expression generalize to arbitrary geometries?

Chapter 22

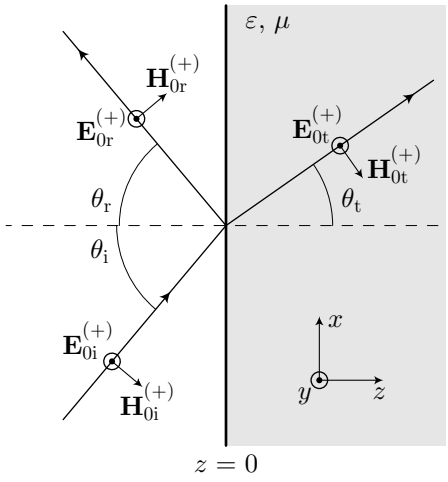
Electromagnetic Casimir Energies as Path Integrals

22.1 Scalar Representation of Electromagnetism

The main deficiency in the worldline method in the previous chapter for computing Casimir and Casimir–Polder energies is that it applies to an artificial field—a massless, scalar field coupled to a background "potential" is *not* the same thing as the electromagnetic field. The next step in improving the method is to develop worldline path integrals for *scalar* representations of elecromagnetism. In propagation problems, scalar representations often make for good approximations to the vector propagation in the paraxial approximation, when the polarization does not vary much as the field propagates (as in the propagation of a well-collimated laser beam). In Casimir-related calculations, we might not expect the scalar approximation to hold, as fields propagating in *all* directions are involved. However, there are some situations in which a scalar treatment is effective: namely when the geometry of space has sufficient symmetry that the two possible polarizations decouple, in which case the two polarizations can act as independent scalar fields. A trivial example is in free space. Less trivial examples include the fields at a planar, dielectric interface, which we treat in more detail below, and a spherical interface (if we take the center of the sphere to define a vector multipole basis).

22.1.1 Planar Interface: TE polarization

Consider the modes associated with a planar interface between vacuum and a (nondispersive) magnetodielectric material (Section 8.9.1), or a more complicated interface with planar symmetry. The case of TE polarization is diagrammed below. The key observation here is that while the magnetic-field direction varies between the incident, reflected, and transmitted components, the electric-field direction is the same among the components.



In particular, we may write the electric field as

$$\mathbf{E} = E_y \hat{y}. \tag{22.1}$$

We will then proceed to work with the *classical* field, where the spatial profile of the classical field carries over to the quantized case. The electromagnetic Hamiltonian (in noncanonical coordinates) is the total field energy (as we will justify in Section 22.2.3)

$$H_{\text{EM}} = \frac{1}{2}\int d^3r \Big[\mathbf{E}\cdot\mathbf{D} + \mathbf{H}\cdot\mathbf{B}\Big] = \frac{1}{2}\int d^3r \Big[\epsilon E^2 + \mu H^2\Big], \tag{22.2}$$

where we are assuming a linear magnetodielectric medium with possibly inhomogeneous permittivity $\epsilon(\mathbf{r})$ and permeability $\mu(\mathbf{r})$.

We will continue by treating the field as if there is a single mode of wave vector $\mathbf{k}$ and frequency $\omega = ck$. In this case we can identify

$$\nabla \equiv i\mathbf{k}, \qquad \partial_t \equiv -i\omega. \tag{22.3}$$

In view of the normal-mode decomposition of the electromagnetic field (Section 8.4), there is an implied sum over all such modes, in which case these identifications are justified in the general case.

Then the Maxwell equation

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B} \tag{22.4}$$

becomes

$$\mathbf{B} = \frac{1}{\omega}\mathbf{k}\times\mathbf{E}, \tag{22.5}$$

and so

$$H^2 = \frac{1}{\mu^2}B^2 = \frac{1}{\mu^2\omega^2}(\mathbf{k}\times\mathbf{E})^2 = \frac{k^2}{\mu^2\omega^2}E_y^2 = -\frac{1}{\mu^2\omega^2}(\nabla E_y)^2, \tag{22.6}$$

where we have used the fact that $\mathbf{k}$ and $\mathbf{E}$ are orthogonal. Then we can write the Hamiltonian (22.2) in terms of only one component E_y of the field as

$$H_{\text{TE}} = \frac{1}{2}\int d^3r \left[\epsilon E_y^2 - \frac{1}{\mu\omega^2}(\nabla E_y)^2\right]. \tag{22.7}$$

We would then like to define a scalar potential that satisfies

$$E_y = i\omega\phi = -\partial_t\phi, \tag{22.8}$$

so that up to an arbitrary constant (corresponding to a choice of "zero" time), we should define

$$\phi(\mathbf{r},t) := -\int_0^t dt'\, E_y(\mathbf{r},t'). \tag{22.9}$$

(TE scalar potential)

In this case, the Hamiltonian finally becomes

$$H_{\text{TE}} = \frac{1}{2}\int d^3r \left[\epsilon(\partial_t\phi)^2 + \frac{1}{\mu}(\nabla\phi)^2\right]. \tag{22.10}$$
(scalar TE Hamiltonian)

This is equivalent to the Hamiltonian (22.2) in the sense of producing the same decomposition in normal modes, restricted to TE-polarized modes. Note that while we have effectively changed a curl to a gradient in the magnetic-field term, but the relevant boundary condition is still the same. In the vector case, the component of $\mathbf{H}$ parallel to the interface is continuous through the interface. In the scalar case, the normal derivative $\partial_z\phi$ is continuous across the interface, which is the same requirement as continuity the normal component of $\nabla\phi$. These are equivalent constraints on the magnetic-field vectors in each case.

Note also that in restricting the electric-field representation to E_y, we have effectively restricted our analysis to a particular plane of incidence. However, our final Hamiltonian (22.10) does not depend on this choice, and thus functions as a sum over all planes of incidence.

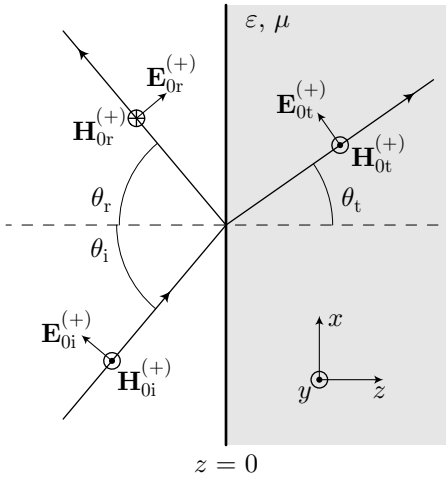
The (noncanonical) Hamiltonian (22.10) has a scalar momentum field $\pi_\phi = -\epsilon\partial_t\phi$ conjugate to ϕ. Then the Hamilton equation $\dot{\pi}_\phi = -\delta H_{\text{TE}}/\delta\phi$ yields

$$\left(\epsilon\partial_t^2 - \nabla\cdot\frac{1}{\mu}\nabla\right)\phi = 0 \tag{22.11}$$
(TE scalar wave equation)

as the appropriate wave equation for ϕ.

22.1.2 Planar Interface: TM polarization

For TM polarization, we have a similar situation, but now $\mathbf{H}$ acts as the scalar field, while $\mathbf{E}$ has a vector character.



Again, we will emphasize this by writing

$$\mathbf{H} = H_y\hat{y}. \tag{22.12}$$

Then the Maxwell equation

$$\nabla\times\mathbf{H} = \partial_t\mathbf{D} \tag{22.13}$$

becomes

$$\mathbf{E} = -\frac{1}{\epsilon\omega}\mathbf{k}\times\mathbf{H}, \tag{22.14}$$

and so

$$E^2 = \frac{1}{\epsilon^2\omega^2}(\mathbf{k}\times\mathbf{H})^2 = \frac{k^2}{\epsilon^2\omega^2}H_y^2 = -\frac{1}{\epsilon^2\omega^2}(\nabla H_y)^2, \tag{22.15}$$

where we have used the fact that $\mathbf{k}$ and $\mathbf{E}$ are orthogonal. Then the Hamiltonian (22.2) becomes

$$H_{\text{TM}} = \frac{1}{2}\int d^3r \left[-\frac{1}{\epsilon\omega^2}(\nabla H_y)^2 + \mu H_y^2\right]. \tag{22.16}$$

Then given the condition

$$H_y = i\omega\varphi = -\partial_t\varphi, \tag{22.17}$$

we define the alternate scalar potential

$$\varphi(\mathbf{r},t) = -\int_0^t dt'\, H_y(\mathbf{r},t'), \tag{22.18}$$
(scalar TM field)

so that the Hamiltonian becomes

$$H_{\text{TM}} = \frac{1}{2}\int d^3r \left[\mu(\partial_t\varphi)^2 + \frac{1}{\epsilon}(\nabla\varphi)^2\right]. \tag{22.19}$$
(scalar TM Hamiltonian)

This is again equivalent to the Hamiltonian (22.2) in the sense of producing the same mode decomposition, restricted to TM modes. Notice that this scalar Hamiltonian has the same form as the TE Hamiltonian (22.10), except that ϵ and μ are switched, while the scalar field now represents $\mathbf{H}$ instead of $\mathbf{E}$. This symmetry is also apparent in the two Maxwell equations we used, if we write them in the form

$$\begin{aligned}\nabla\times\mathbf{E} &= -\mu\partial_t\mathbf{H}\\ \nabla\times\mathbf{H} &= \epsilon\partial_t\mathbf{E}.\end{aligned} \tag{22.20}$$

Here we see that under the combined transformation $\mathbf{E}\longrightarrow\mathbf{H}$, $\mathbf{H}\longrightarrow-\mathbf{E}$, $\epsilon\longrightarrow\mu$, and $\mu\longrightarrow\epsilon$, these two equations transform into each other.

In analogy with the TE case, the (noncanonical) TM Hamiltonian (22.19) has the scalar momentum field $\pi_\varphi = -\mu\partial_t\varphi$ conjugate to φ. Then the Hamilton equation $\dot\pi_\varphi = -\delta H_{\text{TM}}/\delta\varphi$ yields

$$\left(\mu\partial_t^2 - \nabla\cdot\frac{1}{\epsilon}\nabla\right)\varphi = 0 \tag{22.21}$$
(TM scalar wave equation)

as the appropriate wave equation for φ.

22.2 Interaction Energy

In computing Casimir–Polder potentials, we are considering the interactions of point dielectric particles with other material bodies. And generally for Casimir potentials, we are considering the variation in the electromagnetic field energy as a material configuration varies. There are some subtleties here, and so we will expend a fair amount of effort in carefully considering electromagnetic energies. For example: the electromagnetic-field energy in the presence of a magnetodielectric material can be written

$$H_{\text{EM}} = \frac{1}{2}\int d^3r \left(\epsilon E^2 + \mu H^2\right). \tag{22.22}$$

If we consider an atom to be a localized, weak perturbation $(\delta\epsilon,\delta\mu)$ on the medium, a naïve approach would be to make the replacements $\epsilon\longrightarrow\epsilon+\delta\epsilon$ and $\mu\longrightarrow\mu+\delta\mu$ in the Hamiltonian here, in which case we would obtain interaction-potential terms of the form $\delta\epsilon E^2 + \delta\mu H^2$. If we interpret the fields as their unperturbed counterparts, then the first one has the wrong sign for an electric-dipole interaction, while the second one is (accidentally) correct. A more natural choice for the magnetic-field term, as we will see, is to write the energy as B^2/μ, and if we expanded this, we would again obtain the wrong sign. So let's be careful and sort all this out.

22.2.1 Maxwell Equations in Magnetodielectric Materials

We will begin with the fundamentals: a review the construction of the Maxwell equations in magnetodielectric media. Although we are proceeding classically, the same considerations will apply to the quantum-mechanical case. We will start with the vacuum Maxwell equations for the electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$, respectively, including sources,

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{B} &= \mu_0 \epsilon_0 \partial_t \mathbf{E} + \mu_0 \mathbf{j}, \end{aligned} \tag{22.23}$$

(Maxwell's equations with sources)

where the charge density ρ and the current density $\mathbf{j}$ satisfy the continuity equation

$$\partial_t \rho + \nabla \cdot \mathbf{j} = 0. \tag{22.24}$$

(continuity equation)

Now we characterize the response of media via the polarization field $\mathbf{P}$ (electric-dipole-moment density) and the magnetization field $\mathbf{M}$ (magnetic-dipole-moment density). We can then introduce the fields $\mathbf{D}$ and $\mathbf{H}$ via the **constitutive relations**

$$\begin{aligned} \mathbf{D} &= \epsilon_0 \mathbf{E} + \mathbf{P} \\ \mathbf{H} &= \frac{1}{\mu_0} \mathbf{B} - \mathbf{M}. \end{aligned} \tag{22.25}$$

(constitutive relations)

We are already at the crucial point. The fields $\mathbf{D}$ and $\mathbf{H}$ are *introduced* for convenience, since they will simplify the representation of the medium response in Maxwell's equations. The original fields $\mathbf{E}$ and $\mathbf{B}$ are the fundamental fields.[1]

We can then introduce the electric susceptibility χ and the magnetic susceptibility χ_{m} for a medium response,

$$\begin{aligned} \mathbf{P} &= \epsilon_0 \chi \mathbf{E} \\ \mathbf{M} &= \chi_{\mathrm{m}} \mathbf{H}, \end{aligned} \tag{22.26}$$

(linear medium response)

where the susceptibilities are constant for a *linear* medium, but can also represent nonlinear-medium response. Note the asymmetry of the prevailing convention here:[2] The polarization is induced by the electric field $\mathbf{E}$, but by convention the magnetization is ''induced'' by the *introduced* field $\mathbf{H}$—rather than what would seem to be the sensible parallel $\mathbf{M} = (\chi_{\mathrm{m}}/\mu_0)\mathbf{B}$ to the dielectric case. With these susceptibilities, Eqs. (22.25) become

$$\begin{aligned} \mathbf{D} &= \epsilon_0 (1 + \chi) \mathbf{E} \\ \mu_0 (1 + \chi_{\mathrm{m}}) \mathbf{H} &= \mathbf{B}, \end{aligned} \tag{22.27}$$

and thus defining permittivity ϵ and permeability μ for the medium by

$$\begin{aligned} \epsilon &:= \epsilon_0 (1 + \chi) \\ \mu &:= \mu_0 (1 + \chi_{\mathrm{m}}), \end{aligned} \tag{22.28}$$

(permittivity, permeability)

the constitutive relations become

$$\begin{aligned} \mathbf{D} &= \epsilon \mathbf{E} \\ \mathbf{B} &= \mu \mathbf{H}. \end{aligned} \tag{22.29}$$

(linear constitutive relations)

[1] See John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 5.8, especially p. 193.
[2] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice Hall, 2013), Section 6.4.1, p. 284.

Again, notice the asymmetry of the convention: it appears that the electric and magnetic cases are parallel by comparing ϵ and μ, and then by $\mathbf{E}$ vs. $\mathbf{H}$ and $\mathbf{D}$ vs. $\mathbf{B}$. This is, in fact, basically the symmetry that we mentioned before for the Maxwell equations. in Section 22.1.2. However, keep in mind that this is a *formal* symmetry, induced by the definitions in the constitutive relations; the fundamental similarity is between $\mathbf{E}$ vs. $\mathbf{B}$.

Now let us apply the constitutive relations (22.25) to the Maxwell equations (22.23), which are still valid in the presence of magnetodielectric media. We will keep the polarization and magnetization fields explicit this time, rather than appeal to ϵ and μ as we did above to highlight an apparent symmetry in the fields. We will start with the first Maxwell equation,

$$\nabla \cdot \mathbf{E} = \frac{\rho_{\text{total}}}{\epsilon_0}, \tag{22.30}$$

where by the *total charge* label we are emphasizing that the charge density includes any migration of charges due to the polarization response of the medium. Using the first of Eqs. (22.26), we have

$$\nabla \cdot \mathbf{D} = \rho_{\text{total}} + \nabla \cdot \mathbf{P}. \tag{22.31}$$

We can interpret the last term in terms of a **bound charge density**

$$\rho_{\text{P}} := -\nabla \cdot \mathbf{P} \tag{22.32}$$
(bound charge density)

associated with the charge migration due to the dielectric polarization. Then defining the **free charge density** as the part of the total charge *not* associated with the medium response,

$$\rho_{\text{free}} := \rho_{\text{total}} - \rho_{\text{P}}, \tag{22.33}$$
(free charge density)

we have the new Maxwell equation

$$\nabla \cdot \mathbf{D} = \rho_{\text{free}}, \tag{22.34}$$
(first Maxwell equation)

in which the medium response has shifted from the charge density to the $\mathbf{D}$ field.

Similarly, we can take the fourth Maxwell equation in (22.23)

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \partial_t \mathbf{E} + \mu_0 \mathbf{j}_{\text{total}}, \tag{22.35}$$

where again we are using the *total current* label to emphasize that any medium response here is contained in the source term, which is now the current density. Using Eqs. (22.26) to eliminate both $\mathbf{B}$ and $\mathbf{E}$, we have

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{j}_{\text{total}} - \partial_t \mathbf{P} - \nabla \times \mathbf{M}. \tag{22.36}$$

We can again interpret the last two terms as current densities due to the response of media to the fields. The first is a current density associated with charges moving due to a changing polarization,

$$\mathbf{j}_{\text{P}} := \partial_t \mathbf{P}. \tag{22.37}$$
(polarization current density)

Note that this current density is due to motion of bound charges, and as it should, it obeys a continuity equation:

$$\nabla \cdot \mathbf{j}_{\text{P}} = \nabla \cdot \partial_t \mathbf{P} = -\partial_t \rho_{\text{P}}. \tag{22.38}$$

The last term in Eq. (22.36) represents a bound current density due to the magnetization of the medium:

$$\mathbf{j}_{\text{M}} := \nabla \times \mathbf{M}. \tag{22.39}$$
(magnetization current density)

Once again, defining the free current density as the current density not associated with the medium response,

$$\mathbf{j}_{\text{free}} := \mathbf{j}_{\text{total}} - \mathbf{j}_{\text{P}} - \mathbf{j}_{\text{M}}, \tag{22.40}$$
(free current density)

Eq. (22.36) becomes

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D} + \mathbf{j}_{\text{free}}. \tag{22.41}$$
(fourth Maxwell equation)

Now using Eqs. (22.34) and (22.41) in place of the corresponding equations in (22.23), we have the usual Maxwell equations for media:

$$\begin{aligned} \nabla \cdot \mathbf{D} &= \rho \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\partial_t \mathbf{B} \\ \nabla \times \mathbf{H} &= \partial_t \mathbf{D} + \mathbf{j}. \end{aligned} \tag{22.42}$$
(Maxwell's equations in media)

Here, we have dropped the "free" labels, but it is important to remember that the explicit sources here no longer include sources associated with medium response. At the risk of belaboring the point, $\mathbf{E}$ and $\mathbf{B}$ are the original fields, and are the only fields appearing in the two *homogeneous* Maxwell equations. The two *introduced* fields $\mathbf{D}$ and $\mathbf{H}$ serve to reorganize the sources, so we only have to explicitly consider *free* sources. Note that the symmetry $\mathbf{E} \longleftrightarrow -\mathbf{H}$ simultaneously with $\mathbf{B} \longleftrightarrow \mathbf{D}$ (and $\epsilon \longleftrightarrow \mu$, when written explicitly) of the Maxwell equations is apparent here, but only in the *absence* of sources.

22.2.2 Interaction Energies for Magnetodielectric Materials: Static Fields

Having developed the Maxwell equations, we will now apply these in dealing with the energies form materials in *static fields*, which will be sufficient for many purposes (i.e., as long as we are ignoring dispersion), and which will serve as relatively simple studies for developing our intuition of electromagnetic energies and materials.

22.2.2.1 Static-Electric-Field Energies

We begin with the free-space electromagnetic Hamiltonian [Eq. (8.34)]

$$H_{\text{EM}} = \frac{1}{2} \int d^3r \left(\epsilon_0 E^2 + \frac{1}{\mu_0} B^2 \right) \tag{22.43}$$

in terms of the fields, which we will take to be the total field energy. Our goal will be to modify this to encompass the field energy in the presence of magnetodielectric media. First, we will focus on the electric-field energy,

$$H_{\text{E}} = \frac{\epsilon_0}{2} \int d^3r \, E^2, \tag{22.44}$$

and in going to the static, source-free case, since $\nabla \times \mathbf{E} = 0$, we can introduce the scalar potential ϕ such that

$$\mathbf{E} = -\nabla \phi. \tag{22.45}$$

Then the energy becomes

$$H_{\text{E}} = \frac{\epsilon_0}{2} \int d^3r \, |\nabla \phi|^2 = -\frac{\epsilon_0}{2} \int d^3r \, \phi \, \nabla^2 \phi \tag{22.46}$$

after integration by parts. Then using the first Maxwell equation (22.23) in the form of the Poisson equation $\nabla^2 \phi = -\rho/\epsilon_0$, the energy becomes

$$H_{\text{E}} = \frac{1}{2} \int d^3r \, \phi(\mathbf{r}) \, \rho(\mathbf{r}) \tag{22.47}$$
(energy of static electric field)

in terms of the potential and charge density.

It is tempting at this point to try translating this result to a dielectric by separating the free and bound charge densities, $\rho = \rho_{\text{free}} + \rho_{\text{P}}$, and use Eq. (22.32) to introduce the material polarization. However, this does *not* give the correct result, as it only computes the energy associated with assembling the total (free and bound) charge, but not the energy associated with polarizing the medium. Thus, we instead begin by considering the effect δH_{E} on the energy due to a small change $\delta\rho$ in the charge density. Starting with Eq. (22.44), the small variation $\delta\mathbf{E}$ in the electric field due to the change in the charge density

$$\delta H_{\text{E}} = \epsilon_0 \int d^3r\, \mathbf{E} \cdot \delta\mathbf{E}. \tag{22.48}$$

Then using Eq. (22.45) and integrating by parts,

$$\delta H_{\text{E}} = \epsilon_0 \int d^3r\, \nabla\phi \cdot \nabla\delta\phi = -\epsilon_0 \int d^3r\, \phi\, \nabla^2\delta\phi, \tag{22.49}$$

and once again using the Poisson equation, we find

$$\delta H_{\text{E}} = \int d^3r\, \phi(\mathbf{r})\, \delta\rho(\mathbf{r}). \tag{22.50}$$

Note that this differs from what we would have gotten from varying ρ in Eq. (22.47) by a factor of two, because the change in charge modifies both ρ and ϕ. The interpretation here is that by varying ρ against a fixed field ϕ, we are counting the energy due to bringing in charge elements from infinity to assemble $\rho(\mathbf{r})$. In bringing in each of the charge elements, we must do work against the electromagnetic forces against the already-assembled charges, and the total work is the energy of the final configuration.

Now using $\nabla \cdot \mathbf{D} = \rho$, we have

$$\delta H_{\text{E}} = \int d^3r\, \phi\, \nabla \cdot \delta\mathbf{D} = -\int d^3r\, \nabla\phi \cdot \delta\mathbf{D} = \int d^3r\, \mathbf{E} \cdot \delta\mathbf{D}. \tag{22.51}$$

Then if we assume a **linear** dielectric,

$$\mathbf{D} = \epsilon\mathbf{E}, \tag{22.52}$$

with ϵ independent of $\mathbf{E}$ (but may depend on $\mathbf{r}$ or even be generalized to a tensor), then

$$\delta H_{\text{E}} = \int d^3r\, \epsilon\mathbf{E} \cdot \delta\mathbf{E}, \tag{22.53}$$

and upon integration of the field from 0 to $\mathbf{E}$, we find the electric-field energy in the presence of a dielectric:[3]

$$H_{\text{E}} = \frac{1}{2}\int d^3r\, \epsilon E^2 = \frac{1}{2}\int d^3r\, \mathbf{E} \cdot \mathbf{D}.$$

(static-field energy for linear dielectric) (22.54)

The important point here is that this result depends on the assumption of a *linear* dielectric; more generally, the relation can be far more complicated.

22.2.2.2 Static-Magnetic-Field Energies

The magnetic case proceeds in analogy to the electric case. We begin with the magnetic part of the electromagnetic Hamiltonian [Eq. (8.34)],

$$H_{\text{M}} = \frac{1}{2}\int d^3r\, \frac{1}{\mu_0} B^2. \tag{22.55}$$

[3] John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 4.7, especially p. 165; Julius Adams Stratton, *Electromagnetic Theory* (McGraw-Hill, 1941), Sections 2.7-2.8, pp. 104-111.

Since $\nabla \cdot \mathbf{B} = 0$, we can introduce the vector potential $\mathbf{A}$ such that

$$\mathbf{B} = \nabla \times \mathbf{A}, \tag{22.56}$$

with the further condition that $\nabla \cdot \mathbf{A} = 0$, since the longitudinal part of $\mathbf{A}$ can't affect $\mathbf{B}$ anyway. Then

$$H_{\text{M}} = \frac{1}{2}\int d^3r\, \frac{1}{\mu_0}(\nabla \times \mathbf{A})^2 = \frac{1}{2}\int d^3r\, \frac{1}{\mu_0}\mathbf{A} \cdot \nabla \times \nabla \times \mathbf{A} \tag{22.57}$$

where to integrate by parts we used the identity

$$\nabla \cdot \mathbf{A} \times \nabla \times \mathbf{A} = (\nabla \times \mathbf{A}) \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \nabla \times \nabla \times \mathbf{A}. \tag{22.58}$$

Using the last Maxwell equation (22.42) in the static limit and with no magnetic media,

$$\nabla \times \nabla \times \mathbf{A} = \nabla \times \mathbf{B} = \mu_0 \mathbf{j}, \tag{22.59}$$

we then have the source-coupling form of the energy

$$H_{\text{M}} = \frac{1}{2}\int d^3r\, \mathbf{A} \cdot \mathbf{j} \tag{22.60}$$

(energy of static magnetic field)

in terms of the vector potential and the source current density.

Proceeding along the same lines as before, the change in the energy due to a small change in the magnetic field is from Eq. (22.55)

$$\begin{aligned}
\delta H_{\text{M}} &= \frac{1}{\mu_0}\int d^3r\, \mathbf{B} \cdot \delta\mathbf{B} \\
&= \frac{1}{\mu_0}\int d^3r\, (\nabla \times \mathbf{A}) \cdot (\nabla \times \delta\mathbf{A}) \\
&= \frac{1}{\mu_0}\int d^3r\, \delta\mathbf{A} \cdot \nabla \times \nabla \times \mathbf{A} \\
&= \int d^3r\, \delta\mathbf{A} \cdot \mathbf{j}.
\end{aligned} \tag{22.61}$$

Note that we have ended up with a varying field against a fixed current. The interpretation here is as follows: in order to have a static magnetic field in the presence of a source current, the magnetic field must be brought up from zero. In doing so, the induced electromotive forces attempt to modify the flow of charge in the source currents. In order to maintain fixed currents, work must be done against the induced electromotive forces. The total work done is then the total energy of the configuration. Note that we can alternately hold $\mathbf{A}$ fixed while varying $\mathbf{j}$. However, this corresponds to a fixed magnetic field due to the sources, assembling the current density by bringing in current-density elements from infinity. This is not a natural interpretation in the absence of magnetic charge.

Then using the (static) fourth Maxwell equation in the form $\nabla \times \mathbf{H} = \mathbf{j}$ to replace the variation in $\mathbf{j}$,

$$\delta H_{\text{M}} = \int d^3r\, \delta\mathbf{A} \cdot \nabla \times \mathbf{H} = \int d^3r\, (\nabla \times \delta\mathbf{A}) \cdot \mathbf{H} = \int d^3r\, \delta\mathbf{B} \cdot \mathbf{H}, \tag{22.62}$$

where to integrate by parts we used $\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B})$. Again, assuming a **linear** magnetic material (e.g., not a ferromagnetic material), we have

$$\mathbf{B} = \mu\mathbf{H}, \tag{22.63}$$

with μ independent of $\mathbf{H}$, so then

$$\delta H_{\text{M}} = \frac{1}{\mu}\int d^3r\, \mathbf{B} \cdot \delta\mathbf{B}. \tag{22.64}$$

and upon integration of the field from 0 to $\mathbf{B}$, we find the magnetic-field energy in the presence of a magnetic material:[4]

$$H_{\text{M}} = \frac{1}{2}\int d^3r\,\frac{1}{\mu}B^2 = \frac{1}{2}\int d^3r\,\mathbf{B}\cdot\mathbf{H}.$$

(static-field energy for linear magnetic material) (22.65)

To reiterate the main point, this result depends on the assumption of a *linear* magnetic material (in particular, without hysteresis, in which case the energy could depend on the entire history of the magnetic field).

22.2.2.3 Variation of Dielectric Materials

Now a more useful question is: suppose we change the configuration of a dielectric medium by modifying $\epsilon(\mathbf{r})$. What is the difference in energy between the configurations?[5] In answering this question, we should be careful to specify, what happens to the sources? In this section we will explicitly assume *fixed* sources while modifying the dielectric. And we are still working only with static fields and linear dielectrics.

Suppose we have an energy associated with a ''reference'' dielectric configuration $\epsilon_1(\mathbf{r})$,

$$H_{\text{E}}^{(1)} = \frac{1}{2}\int d^3r\,\mathbf{E}_1\cdot\mathbf{D}_1, \tag{22.66}$$

and a second energy associated with a modified dielectric $\epsilon_2(\mathbf{r})$,

$$H_{\text{E}}^{(2)} = \frac{1}{2}\int d^3r\,\mathbf{E}_2\cdot\mathbf{D}_2, \tag{22.67}$$

again with the same sources in both cases. We then define the ''interaction potential'' as the difference of these energies,

$$V_{\text{E}} := H_{\text{E}}^{(2)} - H_{\text{E}}^{(1)}, \tag{22.68}$$

and our goal is to find expressions for V. The concept of an interaction potential is meaningful here in the sense of thinking of the second configuration as introducing a new object into the ''background'' defined by the reference configuration. Then V measures the energy associated with introducing the new object.

First, we can write out the potential as

$$\begin{aligned} V_{\text{E}} &= \frac{1}{2}\int d^3r\,\Big(\mathbf{E}_2\cdot\mathbf{D}_2 - \mathbf{E}_1\cdot\mathbf{D}_1\Big) \\ &= \frac{1}{2}\int d^3r\,\Big(\mathbf{E}_2\cdot\mathbf{D}_1 - \mathbf{E}_1\cdot\mathbf{D}_2\Big) + \frac{1}{2}\int d^3r\,\big(\mathbf{E}_2+\mathbf{E}_1\big)\cdot\big(\mathbf{D}_2-\mathbf{D}_1\big). \end{aligned} \tag{22.69}$$

Then since $\nabla\times(\mathbf{E}_2+\mathbf{E}_1)=0$ for static fields, we can introduce a scalar potential ϕ such that

$$\mathbf{E}_2+\mathbf{E}_1 = -\nabla\phi, \tag{22.70}$$

and so the last integral is proportional to

$$\int d^3r\,(\nabla\phi)\cdot\big(\mathbf{D}_2-\mathbf{D}_1\big) = -\int d^3r\,\phi\,\nabla\cdot\big(\mathbf{D}_2-\mathbf{D}_1\big) = -\int d^3r\,\phi\,\big(\rho_2-\rho_1\big) = 0, \tag{22.71}$$

because we have assumed fixed sources. Thus, the interaction energy is

$$\begin{aligned} V_{\text{E}} &= \frac{1}{2}\int d^3r\,\Big(\mathbf{E}_2\cdot\mathbf{D}_1 - \mathbf{E}_1\cdot\mathbf{D}_2\Big) \\ &= -\frac{1}{2}\int d^3r\,(\epsilon_2-\epsilon_1)\mathbf{E}_2\cdot\mathbf{E}_1. \end{aligned}$$

(energy difference between dielectric configurations) (22.72)

[4] John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 5.16, p. 212; Julius Adams Stratton, *Electromagnetic Theory* (McGraw-Hill, 1941), Section 2.10, p. 112.

[5] John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 4.7, especially p. 167; Julius Adams Stratton, *Electromagnetic Theory* (McGraw-Hill, 1941), Section 2.10, p. 112; L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon, 1960), Section 11, p. 52.

Note that the integrals here in principle extend over all space, but for a bounded object, the integration need only extend over the object. The second form highlights the following observation: *increasing* the dielectric, and in particular introducing a dielectric object into a vacuum background *always decreases the electric-field energy*, provided the sources are fixed.

Suppose that the reference configuration is vacuum, $\epsilon_1 = \epsilon_0$. Then $\epsilon_2 - \epsilon_0 = \epsilon_0\chi$, where $\chi(\mathbf{r})$ is the susceptibility of the object, and using $\mathbf{P} = \epsilon_0\chi\mathbf{E}_2$, the interaction energy becomes

$$V_{\text{E}} = -\frac{1}{2}\int d^3r\, \mathbf{P}(\mathbf{r}) \cdot \mathbf{E}_1(\mathbf{r}). \tag{22.73}$$

This is the usual interaction with a polarizable medium, with a factor of $1/2$ since the polarization is induced by the field. In particular, for a point dipole $\mathbf{d}$ at $\mathbf{r}_0$,

$$\mathbf{P}(\mathbf{r}) = \mathbf{d}\,\delta^3(\mathbf{r} - \mathbf{r}_0), \tag{22.74}$$

so that the interaction becomes

$$V_{\text{E}} = -\frac{1}{2}\mathbf{d} \cdot \mathbf{E}_1(\mathbf{r}_0). \tag{22.75}$$

This is the usual expression for the interaction energy of an induced point dipole in a background field $\mathbf{E}_1$.

22.2.2.4 Variation of Magnetic Materials

The analogous problem of computing the change in the (static) magnetic-field energy when changing the (linear) magnetic material proceeds along similar lines to the electric case.[6] Again, we will assume fixed sources—this time, meaning a fixed current density $\mathbf{j}(\mathbf{r})$.

As before, suppose we have an energy associated with a reference magnetic-material configuration $\mu_1(\mathbf{r})$,

$$H_{\text{M}}^{(1)} = \frac{1}{2}\int d^3r\, \mathbf{B}_1 \cdot \mathbf{H}_1, \tag{22.76}$$

and a second energy associated with a modified configuration $\mu_2(\mathbf{r})$,

$$H_{\text{M}}^{(2)} = \frac{1}{2}\int d^3r\, \mathbf{B}_2 \cdot \mathbf{H}_2, \tag{22.77}$$

with the same source in both cases. The magnetic interaction potential is now the difference of these energies,

$$V_{\text{M}} := H_{\text{M}}^{(2)} - H_{\text{M}}^{(1)}. \tag{22.78}$$

Writing out the potential explicitly,

$$\begin{aligned} V_{\text{M}} &= \frac{1}{2}\int d^3r \left(\mathbf{B}_2 \cdot \mathbf{H}_2 - \mathbf{B}_1 \cdot \mathbf{H}_1\right) \\ &= \frac{1}{2}\int d^3r \left(\mathbf{B}_2 \cdot \mathbf{H}_1 - \mathbf{B}_1 \cdot \mathbf{H}_2\right) + \frac{1}{2}\int d^3r \left(\mathbf{B}_2 + \mathbf{B}_1\right) \cdot \left(\mathbf{H}_2 - \mathbf{H}_1\right). \end{aligned} \tag{22.79}$$

Then since $\nabla \times \mathbf{H} = \mathbf{j}$ for static fields, and thus $\nabla \times (\mathbf{H}_2 - \mathbf{H}_1) = 0$ for a fixed source current, we can introduce a scalar potential φ such that

$$\mathbf{H}_2 - \mathbf{H}_1 = \nabla\varphi, \tag{22.80}$$

and so the last integral is proportional to

$$\int d^3r\, (\nabla\varphi) \cdot \left(\mathbf{B}_2 + \mathbf{B}_1\right) = -\int d^3r\, \varphi\, \nabla \cdot \left(\mathbf{B}_2 + \mathbf{B}_1\right) = 0, \tag{22.81}$$

[6] John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 5.16, especially p. 214; Julius Adams Stratton, *Electromagnetic Theory* (McGraw-Hill, 1941), Section 2.17, p. 126.

because $\nabla \cdot \mathbf{B} = 0$. Thus, the interaction energy is

$$\begin{aligned} V_{\text{M}} &= \frac{1}{2}\int d^3r \left(\mathbf{B}_2 \cdot \mathbf{H}_1 - \mathbf{B}_1 \cdot \mathbf{H}_2\right) \\ &= \frac{1}{2}\int d^3r \,(\mu_2 - \mu_1)\mathbf{H}_2 \cdot \mathbf{H}_1 \\ &= \frac{1}{2}\int d^3r \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)\mathbf{B}_2 \cdot \mathbf{B}_1. \end{aligned}$$

(energy difference between magnetic-material configurations) (22.82)

Again, the integrals here in principle extend over all space, but for a bounded object introduced in the case of μ_2, the integration need only extend over the object. The last two forms highlight the following observation: *increasing* the magnetic permeability, and in particular introducing a paramagnetic object ($\mu_2 > \mu_0$) into a vacuum background material *always increases the magnetic energy*, provided the sources are fixed. This is opposite to the dielectric case. Note that everything up to the first of the above three expressions for V_{M} is the same as the dielectric case with $\mathbf{E} \longrightarrow \mathbf{B}$ and $\mathbf{D} \longrightarrow \mathbf{H}$. However, it is the asymmetry of the $\mathbf{D} = \epsilon\mathbf{E}$ and $\mathbf{B} = \mu\mathbf{H}$ that causes the energy change to move in the opposite direction.

Again, if we take configuration 1 to be free space ($\mu_1 = \mu_0$), then we can write the interaction energy as

$$V_{\text{M}} = \frac{1}{2}\int d^3r \left(\frac{\mu_2}{\mu_0} - 1\right)\mathbf{H}_2 \cdot \mathbf{B}_1. \tag{22.83}$$

But since the magnetization is given by $\mathbf{M} = \chi_{\text{m}}\mathbf{H} = (\mu_2/\mu_0 - 1)\mathbf{H}$, we can write

$$V_{\text{M}} = \frac{1}{2}\int d^3r\, \mathbf{M}(\mathbf{r}) \cdot \mathbf{B}_1(\mathbf{r}). \tag{22.84}$$

This is the usual interaction energy of a magnetized body, but with the *opposite sign* compared to what we would normally expect, and with a factor of 1/2 that reflects the induced nature of the magnetization. As in the electric case, a point magnetic dipole $\mathbf{m}$, represented by $\mathbf{M}(\mathbf{r}) = \mathbf{m}\,\delta(\mathbf{r} - \mathbf{r}_0)$, leads to an interaction energy of $(1/2)\mathbf{m} \cdot \mathbf{B}_1(\mathbf{r}_0)$, which again has the opposite sign to the usual dipole energy $-\mathbf{m} \cdot \mathbf{B}$.

22.2.2.5 Interpretation of Minus Signs

The relative minus sign in the electric vs. magnetic interaction energies that we have developed in the last two sections highlights an important difference between the two cases, and also the importance of being precise about what ''the energy'' means.

To illustrate this, let's return to the dielectric case, but modify the conditions of the sources: rather than fixing the *sources*, we will instead consider fixing the *potentials* on some bounding surfaces,[7] e.g., via conducting electrodes held at fixed voltage by a power supply. In this case, charge is free to migrate in order to maintain the fixed potentials. To compute the energy here, we will begin by considering a small variation in the energy (22.47) due to small changes in the potential and current:

$$\delta H_{\text{E}} = \frac{1}{2}\int d^3r \left(\delta\phi(\mathbf{r})\,\rho(\mathbf{r}) + \phi(\mathbf{r})\,\delta\rho(\mathbf{r})\right). \tag{22.85}$$

Recall that we previously derived a similar expression, Eq. (22.50)

$$\delta H_{\text{E}} = \int d^3r\, \phi(\mathbf{r})\,\delta\rho(\mathbf{r}), \tag{22.86}$$

in the case where the dielectric was fixed. This implies that if the dielectric is held fixed, then the two terms in (22.85) are equal (but not in general equal if the dielectric is changed).

To proceed, we will break the change in the dielectric at fixed potential into two steps: first, disconnect the electrodes from the power supply and change the dielectric at fixed charge; second, reconnect the

[7] See John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 4.7, especially pp. 168-9.

electrodes, and allow charge to flow and restore the original potentials (at least on the electrodes). In step 1, $\delta\rho = 0$, so the energy change is

$$\delta H_{\text{E}}^{(1)} = \frac{1}{2}\int d^3r\, \delta\phi_1(\mathbf{r})\, \rho(\mathbf{r}), \tag{22.87}$$

where $\delta\phi_1$ is the potential change induced by the small dielectric change. We have already computed the energy change V_{E} for a change in dielectric at fixed charge in Eqs. (22.72). In the second step, the batteries must undo the effect of the first-step change $\delta\phi_1$, or written otherwise, $\delta\phi_2 = -\delta\phi_1$. Thus, the energy change in this step is

$$\begin{aligned}\delta H_{\text{E}}^{(2)} &= \frac{1}{2}\int d^3r \left(\delta\phi_2(\mathbf{r})\, \rho(\mathbf{r}) + \phi(\mathbf{r})\, \delta\rho_2(\mathbf{r})\right) \\ &= -\int d^3r\, \delta\phi_1(\mathbf{r})\, \rho(\mathbf{r}) \\ &= -2\delta H_{\text{E}}^{(1)},\end{aligned} \tag{22.88}$$

where we used the observation from above that for a fixed dielectric, the two terms have the same contribution. Thus, the energy change from the second step is opposite to but double the change in the first, so the net energy change is $-V_{\text{E}}$, or the same as the fixed-charge potential (22.72) except for a minus sign, so that in this case, introducing a dielectric object *increases* the total energy. As an example, these observations predict that in introducing a dielectric slap between the plates of a charged capacitor, the dielectric should be sucked into the capacitor if the electrodes are disconnected (maintained at fixed charge), and repelled from the capacitor if the plates are connected to a power supply. In the latter case, the power supply is also doing work on the dielectric, which yields the change in sign.[8]

With the above result, we can now more intuitively interpret the difference in minus sign due to introducing electric vs. magnetic objects, Eqs. (22.72) vs. (22.82).[9] As we saw from the energy variation in Eqs. (22.61),

$$\delta H_{\text{M}} = \int d^3r\, \delta\mathbf{A}\cdot\mathbf{j}. \tag{22.89}$$

the magnetic-field energy (22.65) includes work done by moving charges (in the source current) against electromotive forces as the magnetic field is brought up from zero. This is a closer analogy to the dielectric energy at fixed voltage, which is dominated by the work done by moving charges against the potential. To see this more explicitly, suppose we calculate the work of the current against the electromotive forces. Using $\mathbf{E} = -\partial_t\mathbf{A}$, as potential is brought up from zero, the rate of change of the energy from Eq. (22.89) is

$$\partial_t H_{\text{M}} = \int d^3r\, \partial_t\mathbf{A}\cdot\mathbf{j} = -\int d^3r\, \mathbf{E}\cdot\mathbf{j} \tag{22.90}$$

if the current is fixed. Integrating this over the time over which the potential is ramped to the steady-state value, the change in energy associated with work done against the electromotive forces is

$$\Delta H_{\text{M}} = \int d^3r\, \mathbf{A}\cdot\mathbf{j}, \tag{22.91}$$

where the $\mathbf{A}$ here is the steady-state potential. Note that this is twice the value of the magnetic-field energy (22.60), which led to the magnetic-material energy (22.65). Thus, if we subtract away this contribution to the energy, we are left with the same interaction energy, but with a negative sign, which is in closer analogy to the dielectric energy with fixed charges. Thus, the usual magnetic energy density of the form $-\mathbf{M}\cdot\mathbf{B}$ (or equivalently, the dipole energy $-\mathbf{m}\cdot\mathbf{B}$) counts *just* the energy associated with (permanent) magnetic dipoles being placed in a magnetic field, *after* the dipoles have been created and made permanent elsewhere. The energy we have derived here, $(1/2)\mathbf{M}\cdot\mathbf{B}$, has the opposite sign (and a factor of 1/2, due to the linearity of the magnetic medium), because it *additionally* counts the energy associated with *creating* the magnetic dipoles.

[8] See David J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Prentice Hall, 2013), Section 4.4.4, pp. 202-4.

[9] The interpretations here are straight out of John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), Section 5.16, especially pp. 214-5.

22.2.2.6 Small Material Perturbations

One awkward feature of the interaction energies Eqs. (22.72) and (22.82) for variations in the electric and magnetic materials is that the expressions refer to products of fields in the presence of each of two different configurations. However, there is a useful limit where this complication goes away: that of a *small* change in the material. Returning to the dielectric interaction energy (22.72), suppose we set the background dielectric notation to $\epsilon_1 = \epsilon$, and the perturbed dielectric $\epsilon_2 = \epsilon + \delta\epsilon$, with $\delta\epsilon = \epsilon_0\chi_{\rm obj}$, where $\chi_{\rm obj}$ represents the susceptibility of a weakly polarizable object. Then the interaction potential becomes

$$V_{\rm E} = -\frac{1}{2}\int d^3r\,\delta\epsilon\,\mathbf{E}(\epsilon+\delta\epsilon)\cdot\mathbf{E}(\epsilon). \tag{22.92}$$

To lowest order in $\delta\epsilon$, we can write this solely in terms of the unperturbed field:

$$V_{\rm E} = -\frac{1}{2}\int d^3r\,\delta\epsilon\,E^2(\epsilon) + O(\delta\epsilon^2). \tag{22.93}$$

If we take the small perturbation to be an atom modeled by a point dipole at $\mathbf{r}_0$ with (static) polarizability α_0 (i.e., such that $\mathbf{d} = \alpha_0\mathbf{E}$),

$$\delta\epsilon(\mathbf{r}) = \alpha_0\delta(\mathbf{r}-\mathbf{r}_0), \tag{22.94}$$

then the energy becomes

$$V_{\rm E} = -\frac{1}{2}\alpha_0 E^2(\mathbf{r}_0;\epsilon), \tag{22.95}$$

where $\epsilon(\mathbf{r}_0)$ is the background dielectric (e.g., vacuum or a fluid). This is the usual interaction energy for a linear, induced dipole. It may seem objectionable that we expanded to lowest order in $\delta\epsilon$, and then set it to a delta function, which diverges as a "function." However, it is important to remember that the intent here is to examine the linear response of the atom by expanding to lowest order in the atomic response α_0. The *extent* of the atom, in the point-dipole regime, is smaller than any length scale in the field. Really, we should keep the atomic polarization as a localized but finite distribution, tempered by the small quantity α_0, and then at the end of the day, take the limit as the atom becomes arbitrarily small. Introducing the delta function is a shortcut for this well-defined procedure.

Note also that in computing the atomic response, we have looked at the effect of a localized perturbation to $\epsilon(\mathbf{r})$ at $\mathbf{r}_0$, to linear order in the perturbation, dropped the perturbation, and replaced it by α_0. In the language of functional differentiation, we have thus shown that

$$V_{\rm E}(\mathbf{r}_0) = \alpha_0\,\frac{\delta H_{\rm E}}{\delta\epsilon} \tag{22.96}$$

is the interaction energy for a point-dipole atom at $\mathbf{r}_0$, with static fields and fixed sources.

The same considerations apply to the magnetic-field case. Returning to Eq. (22.82), to lowest order in a perturbation $\delta\mu$ on a background μ,

$$V_{\rm M} = \frac{1}{2}\int d^3r\,\delta\mu\,H^2(\mu) = \frac{1}{2}\int d^3r\,\frac{\delta\mu}{\mu^2}\,B^2(\mu). \tag{22.97}$$

Again, we take the small perturbation to be an atom modeled by a point (magnetic) dipole at $\mathbf{r}_0$ with static, magnetizability β_0. Then since $\mathbf{m} = \beta_0\mu_0\mathbf{H}$,[10] and $\mathbf{M} = \chi_{\rm m}\mathbf{H} = (\delta\mu/\mu_0)\mathbf{H}$, but also $\mathbf{M} = \mathbf{m}\,\delta(\mathbf{r}-\mathbf{r}_0) = \beta_0\mu_0\delta(\mathbf{r}-\mathbf{r}_0)\mathbf{H}$, we have

$$\delta\mu(\mathbf{r}) = \beta_0\mu_0^2\delta(\mathbf{r}-\mathbf{r}_0). \tag{22.98}$$

Note that this differs slightly in form from the electric case (22.94),

$$\delta\epsilon(\mathbf{r}) = \alpha_0\delta(\mathbf{r}-\mathbf{r}_0), \tag{22.99}$$

[10] The convention for the magnetizability induced by $\mathbf{H}$ is used, e.g., by John David Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, 1999), p. 423, and is consistent with the way the magnetization is induced via $\mathbf{M} = \chi_{\rm m}\mathbf{H}$. The factor of μ_0 gives β_0 dimensions of μ_0^{-1} times a volume, in analogy with α_0 having dimensions of ϵ_0 times a volume.

which came from comparing $\mathbf{P} = \epsilon_0\chi\mathbf{E} = \delta\epsilon\,\mathbf{E}$ with $\mathbf{P} = \mathbf{d}\,\delta(\mathbf{r}-\mathbf{r}_0) = \alpha_0\delta(\mathbf{r}-\mathbf{r}_0)\mathbf{E}$.[11] Then the energy becomes

$$V_{\text{M}} = \frac{\mu_0^2}{2\mu^2(\mathbf{r}_0)}\beta_0 B^2(\mathbf{r}_0;\mu) = \frac{\mu_0^2}{2}\beta_0 H^2(\mathbf{r}_0;\mu), \tag{22.100}$$

where $\mu(\mathbf{r}_0)$ is the background magnetic material. In the absence of a background material ($\mu = \mu_0$), this is the usual interaction for an induced dipole with a magnetic field, $-\mathbf{m}\cdot\mathbf{B}/2$, except without the minus sign that we would normally expect due to the reasons we covered in the previous section. Thus, we have shown that for the energy of an atom in a static magnetic field,

$$V_{\text{M}}(\mathbf{r}_0) = -\beta_0\mu_0^2\,\frac{\delta H_{\text{M}}}{\delta\mu} \tag{22.101}$$

in terms of the functional derivative.

22.2.2.7 Application to Casimir–Polder Potentials

Closely related to the above electrostatic energy shifts is the Casimir–Polder effect, which is of course what we're after in all this background material. Note that the Casimir–Polder effect is *not* a static-field phenomenon. However, recall that if the atom is far away from the surface compared to any of its transition wavelengths, the near-dc wavelengths are the only important ones (i.e., we can ignore dispersion), and we can use the dc properties of the atom. Also, the (effectively dc) sources are quantum-vacuum fluctuations in the magnetodielectric material, and will act essentially as fixed sources.

In considering the Casimir–Polder force, we will consider some magnetodielectric body (or configuration of multiple bodies), described by $\epsilon(\mathbf{r})$ and $\mu(\mathbf{r})$. The Casimir–Polder energy is *defined* as the change in the energy when we introduce an atom (precisely, by bringing it in from infinity), which we can model as a point particle with (static) polarizability and magnetizability α_0 and β_0, respectively. Thus, beginning with the electromagnetic Hamiltonian

$$H_{\text{EM}} = \frac{1}{2}\int d^3r\left(\epsilon E^2 + \frac{1}{\mu}B^2\right), \tag{22.102}$$

the idea is to consider small perturbations

$$\epsilon \longrightarrow \epsilon + \delta\epsilon, \qquad \mu \longrightarrow \mu + \delta\mu, \tag{22.103}$$

and consider the first-order expansion of the difference in vacuum expectation values (in the quantum case)

$$V_{\text{CP}} = \Big\langle H_{\text{EM}}(\epsilon+\delta\epsilon,\mu+\delta\mu)\Big\rangle - \Big\langle H_{\text{EM}}(\epsilon,\mu)\Big\rangle = \int d^3r'\left[\frac{\delta\langle H_{\text{EM}}\rangle}{\delta\epsilon}\,\delta\epsilon(\mathbf{r}') + \frac{\delta\langle H_{\text{EM}}\rangle}{\delta\mu}\,\delta\mu(\mathbf{r}')\right], \tag{22.104}$$

in terms of the (partial) functional derivatives of the energy. If we take the perturbations to be due to an atom located at $\mathbf{r}$, where the dimensions of that atom are small compared to other length scales in the calculation, we can formally replace the (finite but localized) perturbations by delta functions:

$$\delta\epsilon(\mathbf{r}') = \alpha_0\delta(\mathbf{r}'-\mathbf{r}), \qquad \delta\mu(\mathbf{r}') = \beta_0\mu_0\delta(\mathbf{r}'-\mathbf{r}). \tag{22.105}$$

This procedure gives a result for the energy that is lowest-order in terms of the atom's size. Then the Casimir–Polder energy becomes

$$V_{\text{CP}}(\mathbf{r}) = \alpha_0\frac{\delta\langle H_{\text{EM}}\rangle}{\delta\epsilon} + \beta_0\mu_0^2\frac{\delta\langle H_{\text{EM}}\rangle}{\delta\mu},$$

(Casimir–Polder potential, no dispersion) (22.106)

where the functional derivatives are to be evaluated at the atomic position ($\mathbf{r}$).

[11] The expressions (22.98) and (22.99) are consistent with, for example, S. Y. Buhmann, H. Safari, Ho Trung Dung, and D.-G. Welsch, "Two-atom van der Waals interaction between polarizable/magnetizable atoms near magnetoelectric bodies," *Optics and Spectroscopy* **103**, 374 (2007) (doi: 10.1134/S0030400X07090068), Eqs. (6) and (12).

However, recall that the Casimir effect is generally by considering atomic interactions with vacuum fields, to leading order via dipole interactions of the form $-\mathbf{d}\cdot\mathbf{E}$ and $-\mathbf{m}\cdot\mathbf{B}$. Our *static* treatment in the past few sections seems to be consistent with the former but opposite to the latter, as we see in Eq. (22.100). Thus, it seems like we have a reversed magnetic-field contribution, and it is tempting to take the Casimir–Polder energy to be

$$V_{\text{CP}}(\mathbf{r}) = \alpha_0 \frac{\delta\langle H_{\text{EM}}\rangle}{\delta\epsilon} - \beta_0\mu_0^2 \frac{\delta\langle H_{\text{EM}}\rangle}{\delta\mu} \qquad \text{(wrong!)} \tag{22.107}$$

So how do we reconcile the Casimir–Polder expression (22.106) with the static-field results? The answer is that the sign in the magnetic-field case is already built into the result (22.106). Specifically, this equation gives the energy change for introducing a particle *without direct reference to the fields*. Thus, the static results give us *intuition* as to what will happen to the field energy as we make small changes to the magnetodielectric material. Specifically, these results suggest that when introducing a purely dielectric vs. purely magnetic atom into the same background configuration, the *sign* of the energy change should be different. (The *magnitude* is not necessarily the same, as each of the two possible ''atoms'' couples to different fields.) We will confirm this intuition in Section 22.11, where we will consider all combinations of purely electric and magnetic particles and planar surfaces. The result: the potential is attractive if both are dielectric or both are magnetic; the potential is repelling whenever the two have opposite characters.

22.2.2.8 Functional Derivatives and Expectation Values: A Second Look

The results in the previous section are somewhat counterintuitive at first glance. To see this, let's set up a heuristic (but wrong!) calculation of the Casimir–Polder potential. Suppose for simplicity that we have an atom located *outside* a dielectric body. A neutral atom interacting with an electric field

$$V_{\text{E}} = -\frac{1}{2}\alpha_0 E^2 \tag{22.108}$$

if we ignore dispersion and model the atom as an electric dipole. This comes from the energy $-\mathbf{d}\cdot\mathbf{E}$, with $\mathbf{d} = \alpha_0\mathbf{E}$ and the factor of $1/2$ arises for a dipole induced by the field. For the magnetic-dipole case we can analogously write

$$V_{\text{M}} = -\frac{\mu_0^2}{2\mu^2}\beta_0 B^2. \tag{22.109}$$

We have seen these energies in the previous section. Quantum-mechanically, this holds as well. Even though the field is in the vacuum state, the field fluctuations are captured by the field variance, and we simply interpret the squared fields as expectation values

$$V_{\text{CP}} = -\frac{1}{2}\alpha_0\langle E^2\rangle - \frac{\mu_0^2}{2\mu^2}\beta_0\langle B^2\rangle\,, \tag{22.110}$$

representing the mean vacuum-induced energy induced by the magnetodielectric body.

Everything we have done is fine so far. What is now tempting is to look at the Hamiltonian (22.102), and functionally differentiate to obtain

$$\frac{\delta H_{\text{EM}}}{\delta\epsilon} = \frac{E^2}{2}, \qquad \frac{\delta H_{\text{EM}}}{\delta\mu} = -\frac{B^2}{2\mu^2} \qquad \text{(wrong!)}. \tag{22.111}$$

Now we could solve these for the squared fields, substitute into Eq. (22.110), and arrive at

$$V_{\text{CP}} = -\alpha_0\left\langle \frac{\delta H_{\text{EM}}}{\delta\epsilon}\right\rangle + \beta_0\mu_0^2\left\langle \frac{\delta H_{\text{EM}}}{\delta\mu}\right\rangle \qquad \text{(wrong!)}. \tag{22.112}$$

However, this is wrong, since there is an overall minus sign on the dielectric term compared to the correct answer (22.106). The problem, to reiterate, is that the *fields themselves* also depend on ϵ and μ, so the variation should also vary the fields. However, their variation is not obvious, and must be established

from the more careful arguments above, which precisely account for the variation of the fields with the magnetodielectric configuration.

To take this a bit further: recall Eq. (22.93):

$$V_{\text{E}} = -\frac{1}{2}\int d^3r\, \delta\epsilon\, E^2(\epsilon) + O(\delta\epsilon^2). \tag{22.113}$$

This is the interaction energy for a small change in the permittivity, to lowest order. The point is, this has the opposite sign, compared to naïvely differentiating an energy involving ϵE^2 with respect to ϵ while holding $\mathbf{E}$ fixed. Evidently, this means that to lowest order, the field variation is

$$\mathbf{E}(\epsilon + \delta\epsilon) = \left(1 - \frac{\delta\epsilon}{\epsilon}\right)\mathbf{E}(\epsilon), \tag{22.114}$$

and this variation in ϵE^2 conspires with the obvious variation of ϵ to yield the overall minus sign in Eq. (22.113). However, note that $\mathbf{D}$ stays *constant* under the same variation, to linear order:

$$\mathbf{D}(\epsilon + \delta\epsilon) = (\epsilon + \delta\epsilon)\mathbf{E}(\epsilon + \delta\epsilon) = (\epsilon + \delta\epsilon)\left(1 - \frac{\delta\epsilon}{\epsilon}\right)\epsilon\mathbf{E}(\epsilon) = \mathbf{D}(\epsilon). \tag{22.115}$$

Thus, we *can* make a naïve version of the argument work out, but only if we rewrite the electromagnetic energy (22.102) in terms of $\mathbf{D}$ and $\mathbf{B}$,

$$H_{\text{EM}} = \frac{1}{2}\int d^3r \left(\frac{D^2}{\epsilon} + \frac{B^2}{\mu}\right), \tag{22.116}$$

and *then* differentiate while holding the fields here fixed. This procedure leads to something more like the previous result (22.106).

22.2.3 Hamiltonian Structure of Electromagnetism and Linear Materials

Why have we been dealing with static fields when Casimir–Polder potentials arise fundamentally from vacuum-field *fluctuations*? Here we will show that the results we have derived for vacuum potentials are still valid if we relax the assumption of static fields, and in doing so we will review some of the symplectic structure of the electromagnetic fields with media.

To see how all this comes out of the Hamiltonian structure of the field, we can write the Lagrangian for the electromagnetic field coupled to matter with (linear) permittivity ϵ and (linear) permeability μ as

$$L_{\text{EM}} = \frac{1}{2}\int d^3r \Big(\mathbf{E}\cdot\mathbf{D} - \mathbf{B}\cdot\mathbf{H}\Big), \tag{22.117}$$

where as usual the fields are related by

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = \epsilon\mathbf{E}, \qquad \mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M}) = \mu\mathbf{H}, \tag{22.118}$$

for (linear) polarization $\mathbf{P}$ and magnetization $\mathbf{M}$. In the Coulomb gauge, $\mathbf{E} = -\partial_t\mathbf{A}$ and $\mathbf{B} = \nabla\times\mathbf{A}$, so the Lagrangian becomes

$$L_{\text{EM}}(\mathbf{A}, \partial_t\mathbf{A}) = \frac{1}{2}\int d^3r \left[\epsilon(\partial_t\mathbf{A})^2 - \frac{1}{\mu}(\nabla\times\mathbf{A})^2\right]$$

(electromagnetic Lagrangian with linear material) (22.119)

in terms of the vector potential, which plays the role of the generalized-coordinate field. We can check that this Lagrangian is sensible by writing out the Euler–Lagrange equation,

$$\frac{\delta L}{\delta\mathbf{A}} - \partial_t\frac{\delta L}{\delta(\partial_t\mathbf{A})} = 0, \tag{22.120}$$

which implies

$$\nabla \times \left(\frac{1}{\mu}\nabla \times \mathbf{A}\right) - \epsilon\partial_t^2 \mathbf{A} = 0, \tag{22.121}$$

yielding the Maxwell equation

$$\nabla \times \mathbf{H} = \partial_t \mathbf{D}, \tag{22.122}$$

assuming a time-independent (and hence nondispersive) permittivity.

Now to derive the Hamiltonian, we will make the usual transformation. That is, the momentum field is

$$\mathbf{\Pi} = \frac{\delta L_{\text{EM}}}{\delta(\partial_t \mathbf{A})} = \epsilon\partial_t \mathbf{A} = -\epsilon\mathbf{E} = -\mathbf{D}, \tag{22.123}$$

in which case we obtain the perturbed Hamiltonian

$$\begin{aligned} H_{\text{EM}}(\mathbf{\Pi}, \mathbf{A}) &= \int d^3r\, \mathbf{\Pi} \cdot \partial_t \mathbf{A} - L \\ &= \frac{1}{2}\int d^3r \left[\epsilon(\partial_t \mathbf{A})^2 + \frac{1}{\mu}(\nabla \times \mathbf{A})^2\right] \\ &= \frac{1}{2}\int d^3r \left[\frac{1}{\epsilon}\mathbf{\Pi}^2 + \frac{1}{\mu}(\nabla \times \mathbf{A})^2\right] \end{aligned} \tag{22.124}$$

in terms of canonical coordinate fields. In terms of the electromagnetic fields, we have

$$H_{\text{EM}} = \frac{1}{2}\int d^3r \left(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}\right), \tag{22.125}$$

(electromagnetic Hamiltonian)

which we can identify as the (time-invariant) total energy of the electromagnetic field.

22.2.3.1 Material Perturbations

Then to find the effect on the *electrodynamic* energy due to a change in material properties ϵ and μ (or equivalently, introducing an extra magnetodielectric object), we proceed as in the static case. Referring to the electromagnetic energy as represented by the Hamiltonian (22.125), we should write down our first electromagnetic energy for the reference magnetodielectric-material configuration, characterized by $\epsilon_1(\mathbf{r})$ and $\mu_1(\mathbf{r})$, as

$$H_{\text{EM}}^{(1)} = \frac{1}{2}\int d^3r \left(\mathbf{E}_1\cdot\mathbf{D}_1 + \mathbf{B}_1\cdot\mathbf{H}_1\right), \tag{22.126}$$

and a second energy associated with a modified configuration characterized by $\epsilon_2(\mathbf{r})$ and $\mu_2(\mathbf{r})$, as

$$H_{\text{EM}}^{(2)} = \frac{1}{2}\int d^3r \left(\mathbf{E}_2\cdot\mathbf{D}_2 + \mathbf{B}_2\cdot\mathbf{H}_2.\right), \tag{22.127}$$

Again, we wish to hold the sources fixed between the two cases. The (dynamical) electromagnetic interaction potential is now the difference of these energies,

$$V_{\text{EM}} := H_{\text{EM}}^{(2)} - H_{\text{EM}}^{(1)}. \tag{22.128}$$

Writing out the potential explicitly,

$$\begin{aligned} V_{\text{EM}} &= \frac{1}{2}\int d^3r \left(\mathbf{E}_2\cdot\mathbf{D}_2 + \mathbf{B}_2\cdot\mathbf{H}_2\right) - \frac{1}{2}\int d^3r \left(\mathbf{E}_1\cdot\mathbf{D}_1 + \mathbf{B}_1\cdot\mathbf{H}_1\right) \\ &= \frac{1}{2}\int d^3r \left(\mathbf{E}_2\cdot\mathbf{D}_2 - \mathbf{E}_1\cdot\mathbf{D}_1\right) + \frac{1}{2}\int d^3r \left(\mathbf{B}_2\cdot\mathbf{H}_2 - \mathbf{B}_1\cdot\mathbf{H}_1\right). \end{aligned} \tag{22.129}$$

Then separating out cross-terms as before,

$$\begin{aligned} V_{\text{EM}} &= \frac{1}{2}\int d^3r \left(\mathbf{E}_2\cdot\mathbf{D}_1 - \mathbf{E}_1\cdot\mathbf{D}_2\right) + \frac{1}{2}\int d^3r \left(\mathbf{B}_2\cdot\mathbf{H}_1 - \mathbf{B}_1\cdot\mathbf{H}_2\right) \\ &\quad + \frac{1}{2}\int d^3r \left[\left(\mathbf{E}_2+\mathbf{E}_1\right)\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right) + \left(\mathbf{B}_2+\mathbf{B}_1\right)\cdot\left(\mathbf{H}_2-\mathbf{H}_1\right)\right]. \end{aligned} \tag{22.130}$$

The first term here has the form of the static-electric interaction energy (22.72), while the second term has the form of the static-magnetic interaction energy (22.82). The last integral contains two terms that we showed vanished separately, under the assumption of static fields. We can't make the same assumption, but we still wish to show that the last integral vanishes, and we can do this as follows. Noting that we can write the electric sum field in terms of potentials in the most general way as

$$\mathbf{E}_2 + \mathbf{E}_1 = -\nabla\phi - \partial_t\mathbf{A}, \tag{22.131}$$

the first term in the integral has the form

$$\begin{aligned} I_{ED} &= \frac{1}{2}\int d^3r \left(\mathbf{E}_2+\mathbf{E}_1\right)\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right) \\ &= -\frac{1}{2}\int d^3r \left(\nabla\phi\right)\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right) - \frac{1}{2}\int d^3r \left(\partial_t\mathbf{A}\right)\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right). \end{aligned} \tag{22.132}$$

The first integral vanishes after integrating by parts, using $\nabla\cdot\mathbf{D} = \rho$, and assuming fixed sources. Thus, we are left with

$$I_{ED} = -\frac{1}{2}\int d^3r \left(\partial_t\mathbf{A}\right)\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right). \tag{22.133}$$

The second part of the last integral in (22.130) has the form

$$\begin{aligned} I_{BH} &= \frac{1}{2}\int d^3r \left(\mathbf{B}_2+\mathbf{B}_1\right)\cdot\left(\mathbf{H}_2-\mathbf{H}_1\right) \\ &= \frac{1}{2}\int d^3r \left(\nabla\times\mathbf{A}\right)\cdot\left(\mathbf{H}_2-\mathbf{H}_1\right) \\ &= \frac{1}{2}\int d^3r\, \mathbf{A}\cdot\nabla\times\left(\mathbf{H}_2-\mathbf{H}_1\right) \\ &= \frac{1}{2}\int d^3r\, \mathbf{A}\cdot\partial_t\left(\mathbf{D}_2-\mathbf{D}_1\right), \end{aligned} \tag{22.134}$$

where we wrote the magnetic sum field in terms of the same vector potential as the electric sum field, integrated by parts using $\nabla\cdot(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot(\nabla\times\mathbf{A}) - \mathbf{A}\cdot(\nabla\times\mathbf{B})$, and then used the Maxwell equation $\nabla\times\mathbf{H} = \partial_t\mathbf{D} + \mathbf{j}$, where the current density drops out because we assume it to be the same in both configurations. Now the critical point is to assume that a decomposition into normal modes exists; this is true because we can still assume monochromatic solutions of frequency ω (e.g., right in the Maxwell equations), which separates the time and space variables of the problem, just as in the vacuum case. In this case, we can identify $\partial_t \equiv -i\omega$, and

$$I_{ED} = \frac{i\omega}{2}\int d^3r\, \mathbf{A}\cdot\left(\mathbf{D}_2-\mathbf{D}_1\right) = -I_{BH}. \tag{22.135}$$

Thus, the last integral in Eq. (22.130) vanishes for each mode, and we are left with the energy

$$V_{\text{EM}} = \frac{1}{2}\int d^3r \left(\mathbf{E}_2\cdot\mathbf{D}_1 - \mathbf{E}_1\cdot\mathbf{D}_2\right) + \frac{1}{2}\int d^3r \left(\mathbf{B}_2\cdot\mathbf{H}_1 - \mathbf{B}_1\cdot\mathbf{H}_2\right), \tag{22.136}$$

(electromagnetic interaction energy between material configurations)

which is the same as the electrostatic and magnetostatic results added together. Thus, the intuition from the static case carries over to full electrodynamics, concerning the opposite signs for a dielectric vs. magnetic particle interacting with the same magnetodielectric body.

22.3 Development of the Path Integrals

As usual in the development of path integrals, we will need to start with the Lagrangian. In scalar electromagnetism, we have two Lagrangians, one for each of the two (uncoupled) polarizations. For TE polarization, we have the Hamiltonian (22.10), which arises from the Lagrangian

$$L_{\text{TE}} = \frac{1}{2}\int d^3r \left[\epsilon(\partial_t\phi)^2 - \frac{1}{\mu}(\nabla\phi)^2\right]. \tag{22.137}$$
(scalar TE Lagrangian)

For the TM polarization, we have from the Hamiltonian (22.19) the Lagrangian

$$L_{\text{TM}} = \frac{1}{2}\int d^3r \left[\mu(\partial_t\varphi)^2 - \frac{1}{\epsilon}(\nabla\varphi)^2\right]. \tag{22.138}$$
(scalar TM Lagrangian)

Note in this latter case that this is a Lagrangian that is related to the Hamiltonian solely in terms of the field φ. In terms of the original electromagnetic Hamiltonian, recall that the kinetic and potential terms swapped roles in switching to the field φ.

22.3.1 Gaussian Field Integral

Now we will proceed as in the case of the scalar field coupled to a potential, and we will first do the TE field. The Wick-rotated partition function is

$$Z = \int D\phi \, \exp\left[\frac{i}{\hbar}\int dt\, L(\phi, \partial_t\phi)\right]_{t\longrightarrow -i\tilde{\beta}/c}, \tag{22.139}$$

where the temperature "length" is $\tilde{\beta} = \hbar c\beta$, which becomes

$$\begin{aligned} Z_{\text{TE}} &= \int D\phi \, \exp\left[-\frac{1}{2\hbar c}\int d\tilde{\beta}\int d^3r \, \left(c^2\epsilon(\mathbf{r})(\partial_{\tilde{\beta}}\phi)^2 + \frac{1}{\mu(\mathbf{r})}(\nabla\phi)^2\right)\right] \\ &= \int D\phi \, \exp\left[-\frac{\epsilon_0 c}{2\hbar}\int d\tilde{\beta}\, d^3r \, \phi(\mathbf{r},\tilde{\beta})\left(-\frac{\epsilon(\mathbf{r})}{\epsilon_0}\partial_{\tilde{\beta}}^2 - \nabla\cdot\frac{\mu_0}{\mu(\mathbf{r})}\nabla\right)\phi(\mathbf{r},\tilde{\beta})\right] \\ &\propto \sqrt{\det\left(-\epsilon_{\text{r}}\partial_{\tilde{\beta}}^2 - \nabla\cdot\frac{1}{\mu_{\text{r}}}\nabla\right)^{-1}} \end{aligned} \tag{22.140}$$

with the Lagrangian (22.137). Here we have introduced the relative permittivity and permeability

$$\epsilon_{\text{r}}(\mathbf{r}) := \frac{\epsilon(\mathbf{r})}{\epsilon_0}, \qquad \mu_{\text{r}}(\mathbf{r}) := \frac{\mu(\mathbf{r})}{\mu_0}, \tag{22.141}$$
(relative permittivity/permeability)

and we have used $\epsilon_0\mu_0 c^2 = 1$. Then

$$\begin{aligned} \log Z_{\text{TE}} &= -\frac{1}{2}\log\det\left(-\epsilon_{\text{r}}\partial_{\tilde{\beta}}^2 - \nabla\cdot\frac{1}{\mu_{\text{r}}}\nabla\right) \\ &= -\frac{1}{2}\text{Tr}\log\left(-\epsilon_{\text{r}}\partial_{\tilde{\beta}}^2 - \nabla\cdot\frac{1}{\mu_{\text{r}}}\nabla\right), \end{aligned} \tag{22.142}$$
(trace-log form of partition function, TE)

up to an additive constant that disappears in renormalization. The TM case is the same as the TE case, but with $\epsilon_{\text{r}} \longleftrightarrow \mu_{\text{r}}$.

22.3.2 Imaginary-Time Green Operator

As an alternate derivation of Eq. (22.142), for the TE polarization, we can start with the wave equation (22.11),

$$\left(\epsilon\partial_t^2 - \nabla\cdot\frac{1}{\mu}\nabla\right)\phi = 0, \tag{22.143}$$

which defines a Green function via

$$\mu_0\left(\epsilon\partial_t^2 - \nabla\cdot\frac{1}{\mu}\nabla\right)G_{\text{TE}}(\mathbf{r},t;\mathbf{r}',t') = \delta^3(\mathbf{r}-\mathbf{r}')\delta(t-t'), \tag{22.144}$$

where we have introduced an (arbitrary) overall factor of μ_0 to define the Green function. This in turn implies the Green operator

$$G_{\text{TE}} = \left(\frac{\epsilon_{\text{r}}}{c^2}\partial_t^2 - \nabla\cdot\frac{1}{\mu_{\text{r}}}\nabla\right)^{-1}. \tag{22.145}$$

The imaginary-time Green operator is

$$\tilde{G}_{\text{TE}} = G_{\text{TE}}|_{t\longrightarrow i\tilde{\beta}/c} = \left(-\epsilon_{\text{r}}\partial_{\tilde{\beta}}^2 - \nabla\cdot\frac{1}{\mu_{\text{r}}}\nabla\right)^{-1}$$

(imaginary-time Green operator) (22.146)

The short-cut partition-function rule is

$$\log Z_{\text{TE}} = \frac{1}{2}\text{Tr}\log\tilde{G}_{\text{TE}},$$

(Green-operator form of partition function) (22.147)

which is consistent with the result (22.142). The TM scalar wave equation (22.21) is the same as the TE wave function but with $\epsilon \longleftrightarrow \mu$, so (with the appropriate normalization) the expression for $\log Z_{\text{TM}}$ follows in the same way, but with $\epsilon_{\text{r}} \longleftrightarrow \mu_{\text{r}}$ in the end.[12]

22.3.3 Rescaling the Green Operator: Coupling to the Potential

Now consider again the TE Green operator (22.146), in particular noting the possibility of space dependence of $\epsilon_{\text{r}}(\mathbf{r})$ and $\mu_{\text{r}}(\mathbf{r})$. When we change this partition function into a worldline path integral in the following section, recall that the Green-operator derivatives will become kinetic-energy terms. The factor $\mu_{\text{r}}(\mathbf{r})$ that accompanies ∇^2 is particularly problematic, because for the worldline particle, the permeability will act as a space-dependent mass, effectively curving space-time. This drastically complicates the path integral. The factor $\epsilon_{\text{r}}(\mathbf{r})$ also introduces a space-dependent mass for the time direction of the worldline particle. While this technically also induces a nontrivial space-time, it turns out that this will be much easier to deal with.

To deal with this curvature in advance, we can work with *rescaled* Green operators. Since we will be computing energies from the log partition function (22.147), it is important to recall that only *differences* in energies are physically meaningful (or even finite, in Casimir calculation). This means that we can rescale $\tilde{G}$ by an arbitrary constant factor, and as long as we are computing energy differences between different configurations (or *ratios* of partition functions), the scaling factor will cancel out. Similarly, we can also rescale $\tilde{G}$ by an arbitrary *function*. In this case, the function splits off in the determinant, and the functional determinants cancel in the energy difference (partition-function ratio). What is somewhat *less* obvious is that we can rescale $\tilde{G}$ by functions of ϵ_{r} and μ_{r}. These are *not* necessarily the same between the two different configurations, and so this rescaling requires more careful interpretation. Suppose we are considering the Casimir effect between two bodies. A sensible comparison would contrast two different configurations of the *same bodies*, but translated or rotated slightly (or even possibly deformed). An infinitesimal change would be the first step in the calculation of the Casimir force, for example. As long as there is the same amount

[12]Similar expressions for scalar Green operators were used by Julian Schwinger, "Casimir energy for dielectrics," *Proceedings of the National Academy of Sciences* **89**, 4091 (1992) (doi: 10.1073/pnas.89.9.4091).

of *stuff* in the two configurations, the functional determinants of $\epsilon_{\rm r}$ and $\mu_{\rm r}$ are equivalent in both cases, because they are effectively infinite products of these functions at every position in space, but the *order* in the products is not important.

In particular, we would like to define ''flat-space,'' imaginary-time Green operators by rescaling via

$$\begin{aligned} \bar{G}^{-1}_{\rm TE} &:= \sqrt{\mu_{\rm r}}\,\tilde{G}^{-1}_{\rm TE}\sqrt{\mu_{\rm r}} \\ \bar{G}^{-1}_{\rm TM} &:= \sqrt{\epsilon_{\rm r}}\,\tilde{G}^{-1}_{\rm TM}\sqrt{\epsilon_{\rm r}}. \end{aligned} \qquad \text{(flat-space Green operators)} \tag{22.148}$$

Writing these out explicitly,

$$\begin{aligned} \bar{G}^{-1}_{\rm TE} &= -\epsilon_{\rm r}\mu_{\rm r}\partial^2_{\tilde{\beta}} - \sqrt{\mu_{\rm r}}\,\nabla\cdot\frac{1}{\mu_{\rm r}}\nabla\sqrt{\mu_{\rm r}} \\ \bar{G}^{-1}_{\rm TM} &= -\epsilon_{\rm r}\mu_{\rm r}\partial^2_{\tilde{\beta}} - \sqrt{\epsilon_{\rm r}}\,\nabla\cdot\frac{1}{\epsilon_{\rm r}}\nabla\sqrt{\epsilon_{\rm r}}. \end{aligned} \tag{22.149}$$

We still have functions of position on the gradient term, but that term is overall ''neutral'' with respect to position-dependent functions.[13] Using the operator-commutation relation

$$\partial_x g(x) = g(x)\,\partial_x + g'(x), \tag{22.150}$$

we can then write

$$g\partial_x\frac{1}{g^2}\partial_x g = (\partial_x g - g')\frac{1}{g^2}(g\partial_x + g') = \partial_x^2 - \frac{g'^2}{g^2} + \left[\partial_x, \frac{g'}{g}\right] = \partial_x^2 - 2\frac{g'^2}{g^2} + \frac{g''}{g} \tag{22.151}$$

Putting $g = \sqrt{h}$, so that $g' = h'/2\sqrt{h}$ and $g'' = h''/2\sqrt{h} - h'^2/4h^{3/2}$,

$$\sqrt{h}\,\partial_x\frac{1}{h}\partial_x\sqrt{h} = \partial_x^2 + \frac{h''}{2h} - \frac{3h'^2}{4h^2}. \tag{22.152}$$

Note that if we put $h = f'$, then the last two terms here have the form of a Schwarzian derivative (with a factor of 1/2). Also noting that $\partial_x \log h = h'/h$ and $\partial_x^2 \log h = h''/h - h'^2/h^2$, we can alternately write

$$\sqrt{h}\,\partial_x\frac{1}{h}\partial_x\sqrt{h} = \partial_x^2 + \frac{1}{2}\partial_x^2\log h - \frac{1}{4}\big(\partial_x\log h\big)^2 = \partial_x^2 + \partial_x^2\log\sqrt{h} - \big(\partial_x\log\sqrt{h}\big)^2. \tag{22.153}$$

Thus, we may rewrite the Green operators (22.149) as

$$\begin{aligned} \bar{G}^{-1}_{\rm TE} &= -\epsilon_{\rm r}\mu_{\rm r}\partial^2_{\tilde{\beta}} - \nabla^2 + 2V_{\rm TE} \\ \bar{G}^{-1}_{\rm TM} &= -\epsilon_{\rm r}\mu_{\rm r}\partial^2_{\tilde{\beta}} - \nabla^2 + 2V_{\rm TM}, \end{aligned} \qquad \text{(flat-space Green operators)} \tag{22.154}$$

where we have introduced the material-induced ''potentials'' (which have dimensions of inverse area)

$$\begin{aligned} V_{\rm TE}(\mathbf{r}) &:= \frac{1}{2}\left[\big(\nabla\log\sqrt{\mu_{\rm r}}\big)^2 - \nabla^2\log\sqrt{\mu_{\rm r}}\right] \\ V_{\rm TM}(\mathbf{r}) &:= \frac{1}{2}\left[\big(\nabla\log\sqrt{\epsilon_{\rm r}}\big)^2 - \nabla^2\log\sqrt{\epsilon_{\rm r}}\right]. \end{aligned} \qquad \text{(matter-induced potentials)} \tag{22.155}$$

We will now proceed by using these Green operators, which have no space-dependent masses on the ∇^2 terms (but replaced by effective potentials), in the trace-log expression (22.147).

[13] Similar rescalings in path integrals were noted by S. Pasquali, F. Nitti, and A. C. Maggs, ''Numerical methods for fluctuation-driven interactions between dielectrics,'' *Physical Review E* **77**, 016705 (2008) (doi: 10.1103/PhysRevE.77.016705).

22.3.4 Worldline Form of the Path Integral

We can then formally change to an integral representation of the logarithm,

$$\log Z = \frac{1}{2}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\,\mathrm{Tr}\left[\exp\left(-\frac{\mathcal{T}}{\tilde{G}}\right)\right], \tag{22.156}$$

with the understanding that this expression is divergent unless renormalized. Also, we are dropping the polarization label, since this procedure will work for either polarization, given the appropriate choice of Green operator (22.154). Defining the momentum operators

$$p_{\tilde{\beta}} := -i\hbar\partial_{\tilde{\beta}}, \qquad \mathbf{p} := -i\hbar\nabla, \tag{22.157}$$

the partition density with (22.146) becomes

$$\log Z = \frac{1}{2}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\,\mathrm{Tr}\left\{\exp\left[-\mathcal{T}\left(\epsilon_\mathrm{r}\mu_\mathrm{r}\frac{p_{\tilde{\beta}}^2}{2\hbar^2}+\frac{p^2}{2\hbar^2}+V\right)\right]\right\}, \tag{22.158}$$

where V is the potential (22.155) appropriate to the polarization. Notice that we have inserted a factor of $1/2$ in the inverse Green operator here. We can do this because we are free to rescale $\mathcal{T}$ in Eq. (22.156) by an arbitrary factor.

The trace of the exponential factor here has the form of a partition function

$$Z_\mathrm{eff} := \mathrm{Tr}\left[e^{-\mathcal{T}H_\mathrm{eff}}\right], \tag{22.159}$$

(effective partition function, scalar EM)

where $\mathcal{T}$ plays the role of the temperature parameter β, and

$$H_\mathrm{eff}(p_{\tilde{\beta}},\mathbf{p},r_{\tilde{\beta}},\mathbf{r}) := \epsilon_\mathrm{r}(\mathbf{r})\mu_\mathrm{r}(\mathbf{r})\,\frac{p_{\tilde{\beta}}^2}{2\hbar^2}+\frac{p^2}{2\hbar^2}+V(\mathbf{r}) \tag{22.160}$$

(effective Hamiltonian, scalar EM)

is the effective particle Hamiltonian for the scalar electromagnetic field coupled to the medium. The total (log) partition function is then

$$\log Z = \frac{1}{2}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}}\,Z_\mathrm{eff}, \tag{22.161}$$

(log partition function, in terms of effective Z)

in terms of the effective partition function. Note that there is no ordering ambiguity in this effective Hamiltonian.

22.3.4.1 Recap: Normalization of the Particle Path Integral

At this point, our task will be to write the effective partition function (22.159) as a Monte-Carlo-type path integral. To do this, we will recall the corresponding path integral for a standard particle in one dimension of Hamiltonian $H(x,p) = p^2/2m + V(x)$, and then transform our results to the case of the effective Hamiltonian (22.160). Reviewing our results before in Section 20.3.1, the partition function for the standard particle is

$$Z = \int \tilde{D}x\,\delta[x(\hbar\beta)-x_0]\,\exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta}d\tilde{\beta}\left(\frac{m}{2}(\partial_{\tilde{\beta}}x)^2+V(x)\right)\right] \tag{22.162}$$

where the path-integration measure is defined as

$$\tilde{D}x := \left(\frac{m}{2\pi\hbar^2\delta\beta}\right)^{N/2}\prod_{j=0}^{N}dx_j. \tag{22.163}$$

Taking the kinetic-energy part of the Lagrangian to define the probability measure, we can write the partition function as the ensemble average

$$Z = \int dx_0 \left\langle\!\!\left\langle \delta[x(\hbar\beta) - x_0] \, \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right]\right\rangle\!\!\right\rangle, \tag{22.164}$$

where the average is taken with respect to paths of the form

$$x(\tilde{\beta}) = x_0 + \sqrt{\frac{\hbar}{m}}\, W(\tilde{\beta}), \tag{22.165}$$

where $W(t)$ is a standard Wiener path. The delta function selects only the paths (of zero measure) that return to x_0 at "time" $\hbar\beta$, so it is convenient to switch to more appropriate paths. We will thus use the normalization

$$\left\langle\!\!\left\langle \delta[W(T)]\, F[W(t)]\right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi T}} \left\langle\!\!\left\langle F[B_T(t)]\right\rangle\!\!\right\rangle, \tag{22.166}$$

where $B_T(t)$ is a Brownian bridge [i.e., equivalent to a standard Wiener path, but conditioned on the endpoint $B_T(T) = 0$]. Thus, the path integral (22.164) becomes

$$Z = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar}\int_0^{\hbar\beta} d\tilde{\beta}\, V(x)\right]\right\rangle\!\!\right\rangle_{x(\tilde{\beta}) = x_0 + \sqrt{\hbar/m}\, B_{\hbar\beta}(\tilde{\beta})}, \tag{22.167}$$

without the delta function but with Brownian "loops" $B_T(t)$ in imaginary time.

22.3.4.2 Normalization of the Worldline Path Integral

Transforming the partition function (22.159) into the form of Eq. (22.162) is a fairly straightforward transcription, with $\hbar\beta \longrightarrow \mathcal{T}$ and setting $m = \hbar^2$ and $V \longrightarrow \hbar V$ in the exponential. We also lose the distinction between β and $\tilde{\beta}$, which also absorbs the $1/\hbar$ in the exponential. The result of all this is

$$Z_{\text{eff}} = \int D\mathbf{x}\, Dx_{\tilde{\beta}}\, \delta^{D-1}[\mathbf{x}(\mathcal{T}) - x_0]\, \delta[x_{\tilde{\beta}}(\mathcal{T}) - x_{\tilde{\beta}0}] \, \exp\left[-\int_0^{\mathcal{T}} d\tau \left(\frac{(\partial_\tau x_{\tilde{\beta}})^2}{2\epsilon_{\text{r}}(\mathbf{x})\mu_{\text{r}}(\mathbf{x})} + \frac{(\partial_\tau \mathbf{x})^2}{2} + V(\mathbf{x})\right)\right], \tag{22.168}$$

and we have the integration measures

$$\tilde{D}x_{\tilde{\beta}} := \left(\frac{\hbar}{2\pi\delta\mathcal{T}}\right)^{N/2} \left[\prod_{j=0}^{N} dx_{\tilde{\beta}j}\right] \left[\prod_{j=1}^{N} \left[\epsilon_{\text{r}}(\mathbf{x}_j)\mu_{\text{r}}(\mathbf{x}_j)\right]^{-1/2}\right], \;\; \tilde{D}\mathbf{x} := \left(\frac{\hbar}{2\pi\delta\mathcal{T}}\right)^{(D-1)N/2} \prod_{j=0}^{N} d^{D-1}x. \tag{22.169}$$

Note that we are writing the expression for D dimensions ($D-1$ space dimensions and 1 imaginary-time dimension). Of course, we only obtain physical results in the case $D = 4$, but keeping explicit track of D as a variable allows us to track the origin of various factors, as well as to examine other D values as test cases (and even use noninteger values for dimensional regularization). Note also that $\tilde{\beta}$ here ranges from 0 to $\hbar c\beta$, including an extra factor of c compared to the standard particle case to measure the time-like coordinate in spatial units.

The result here bears some explanation. Note that we had no ordering issues to contend with, as $p_{\tilde{\beta}}$ commutes with $\epsilon(\mathbf{x})$. The main difference with the standard result (22.162) is the dependence of the $p_{\tilde{\beta}}$ term in the Hamiltonian on $\epsilon_{\text{r}}(\mathbf{r})$ and $\mu_{\text{r}}(\mathbf{r})$. In the derivation of (22.162), recall that we had to perform integrations over momenta. Each one of these integrations generated a factor of $\sqrt{m}$. Because of the spatial dependence of $\epsilon_{\text{r}}(\mathbf{r})$ and $\mu_{\text{r}}(\mathbf{r})$, this is a bit more complicated. The idea is to then think of fixing the $\mathbf{x}$ coordinates in the path integral, while performing the integrals over the $x_{\tilde{\beta}}$ coordinates. Then for these integrals, $\epsilon_{\text{r}}(\mathbf{x})$ and $\mu_{\text{r}}(\mathbf{x})$ appear as different (fixed) masses for each of the $x_{\tilde{\beta}}$ integrals, and thus they contribute the extra factors in the measure (22.169), and they still appear in the time-like kinetic-energy

term in (22.168). The path integral then proceeds just as in the standard-particle case for the space-like dimensions.

Turning this path integral into ensemble-average form,

$$Z_{\text{eff}} = \int d^{D-1}x_0 \int_0^{\hbar c\beta} d\tilde{\beta} \left\langle\!\!\left\langle \delta^{D-1}[\mathbf{x}(\mathcal{T}) - x_0]\, \delta[x_{\tilde{\beta}}(\mathcal{T}) - x_{\tilde{\beta}0}]\, \exp\left[-\int_0^{\mathcal{T}} d\tau\, V(\mathbf{x})\right] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.170}$$

where the (four-dimensional) paths satisfy

$$\begin{aligned} \mathbf{x}(0) &= \mathbf{x}_0 \\ dx_{\tilde{\beta}}(\tau) &= \sqrt{\epsilon_{\text{r}}[\mathbf{x}(\tau)]\, \mu_{\text{r}}[\mathbf{x}(\tau)]}\, dW_{\tilde{\beta}}(\tau) \\ dx_\alpha(\tau) &= dW_\alpha(\tau), \end{aligned} \tag{22.171}$$

where α here refers to the spatial dimensions, and $x_{\tilde{\beta}}$ is the time-like coordinate of the path. Note that the choice of stochastic calculus is not important here, as the noise is additive. Carrying out the derivative and accounting for the δ functions,

$$Z_{\text{eff}} = \frac{\hbar c\beta}{(2\pi\mathcal{T})^{D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}}\mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.172}$$
(effective Z, scalar EM)

The path average here is defined as

$$\langle g \rangle_{\mathbf{x}(\tau)} := \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} d\tau\, g[\mathbf{x}(\tau)] \tag{22.173}$$
(path average)

for a function g of the path, where the time coordinate τ of the path runs from 0 to $\mathcal{T}$. The normalization factors came from the prefactor on the right-hand side of Eq. (22.166), which was the probability density for a Wiener path $W(t)$ to end at $W(T) = 0$. Generalizing this for each the space dimensions gives a factor of $(2\pi\mathcal{T})^{-1/2}$, since the space-like part of the paths (22.171) are standard Wiener paths. The generalization in the $x_{\tilde{\beta}}$ direction is additionally weighted by the root-mean-square average of the spatial factor in Eq. (22.166), resulting in the extra path-average factor. We also carried out the $\tilde{\beta}$ integral, which simply gives an extra factor of $\hbar c\beta$. In this case, the path in the $x_{\tilde{\beta}}$ direction is irrelevant, and so we have

$$\mathbf{x}(\tau) = \mathbf{x}_0 + \mathbf{B}_{\mathcal{T}}(\tau) \tag{22.174}$$
(stochastic path, scalar EM)

as the $(D-1)$-dimensional path, without any variable velocities, and we have switched to Brownian bridges to enforce the path condition on the paths that replaced the delta functions. Note that while $B_{\mathcal{T}}(\tau)$ is a standard Brownian bridge in the sense of being a Wiener process pinned to $B_{\mathcal{T}}(\tau = \mathcal{T}) = 0$, by contrast $\mathbf{B}_{\mathcal{T}}(\tau)$ is a D-dimensional *vector* Brownian bridge, or D independent bridges bundled together as a single vector. Thus, the relevant paths are just Brownian bridges displaced to $\mathbf{x}_0$.

Together with the expression (22.161) for the (field) log partition function, Eq. (22.172) gives the log partition function as

$$\log Z = \frac{\hbar c\beta}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}}\mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.175}$$

or using the energy expression

$$E = -\partial_\beta \log Z, \tag{22.176}$$

we have the energy $E_{\text{EM}} = \langle H_{\text{EM}} \rangle$:

$$E_{\text{EM}} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} .$$

(Casimir energy, scalar EM) (22.177)

This is the (unrenormalized) ground-state Casimir energy in scalar electromagnetism in D space-time dimensions, valid for either TE or TM polarization if we choose the potential V appropriately according to Eqs. (22.155). Again, because the physical case corresponds to $D = 4$, the units here only come out correctly for an energy in $D = 4$. Since $\mathcal{T}$ has the dimensions of squared length, this expression will have dimensions of energy-length and energy-area in $D = 3$ and $D = 2$, respectively.

22.3.4.3 Alternate Derivation of the Normalized Path Integral: Space-Dependent Mass

As an alternative to working with the rescaled Green operators (22.148), we can also work directly with the "raw" TE Green operator (22.146) and its TM counterpart, repeated here:

$$\begin{aligned} \tilde{G}_{\text{TE}} &= \left(-\epsilon_{\text{r}} \partial_{\tilde{\beta}}^2 - \nabla \cdot \frac{1}{\mu_{\text{r}}} \nabla \right)^{-1} \\ \tilde{G}_{\text{TM}} &= \left(-\mu_{\text{r}} \partial_{\tilde{\beta}}^2 - \nabla \cdot \frac{1}{\epsilon_{\text{r}}} \nabla \right)^{-1} . \end{aligned} \tag{22.178}$$

As we mentioned at the beginning of Section (22.3.3), the factor of $\mu_{\text{r}}(\mathbf{r})$ in the TE operator and the corresponding factor of $\epsilon_{\text{r}}(\mathbf{r})$ in the TM operator are problematic, as they represent space-dependent masses, which require more careful path-integration techniques. Fortunately, we have already done the hard work for this in Section 20.4.4.1.

Recall that the one-dimensional Hamiltonian (20.370)

$$H_{\partial g \partial}(x, p) = -\hbar^2 \partial_x \frac{1}{2mg(x)} \partial_x + V(x), \tag{22.179}$$

which has the same form as the space components of the Green operators (22.178), leads to a partition-function path integral of the form (20.375):

$$Z_{\partial g \partial} = \sqrt{\frac{m}{2\pi\hbar^2\beta}} \int dx_0 \left\langle\!\!\left\langle \exp\left[-\frac{1}{\hbar} \int_0^{\hbar\beta} d\tilde{\beta} \left(g(x)\, V(x) - \frac{\hbar^2 (2gg'' - 3g'^2)}{8mg^2} \right) \right] \right\rangle\!\!\right\rangle_{x(\tilde{\beta}) = x_0 + \sqrt{\hbar/m}\, B_{\hbar\beta}(\tilde{\beta})} . \tag{22.180}$$

This followed after converting to the proper curved-space momentum operators and developing the propagator path integral, rescaling time to remove the space dependence from the path-integral measure, and reorganizing the ordering potentials, which lead to the g-derivative potential terms here. Recall that due to the temporal scaling, the stochastic paths here are scaled Brownian bridges (i.e., flat-space paths). Now, noting that the effective ordering potential can be rewritten as

$$V_{\text{eff}}(x) := -\frac{(2gg'' - 3g'^2)}{8g^2} = \frac{1}{2} \left[\left(\partial_x \log \sqrt{g(x)} \right)^2 - \partial_x^2 \log \sqrt{g(x)} \right], \tag{22.181}$$

we will now identify g with μ_{r} or ϵ_{r} for the TE and TM cases, respectively. We will also set $V(x) = 0$ here. Then we can proceed by making the same identifications as we did before in Section 22.3.4.2. The spatial dimensions here give the same contributions to the worldline path integral as before, but the potentials $V_{\text{TE}}(\mathbf{r})$ and $V_{\text{TM}}(\mathbf{r})$ of Eqs. (22.155) come from the ordering potential [instead of from the rescaling step in Eqs. (22.148)].

Then the normalization of the imaginary-time dimension also goes through as before, but with one difference. It would seem from Eqs. (22.178) that the local path variance is scaled by $\epsilon_{\text{r}}[\mathbf{x}(t)]$ or $\mu_{\text{r}}[\mathbf{x}(t)]$ for the TE and TM cases, respectively. [See, e.g., Eqs. (22.171).] However, because we rescaled time by $g[\bar{x}(t)]$ in Eq. (20.365), the local path variance ends up being scaled by $\epsilon_{\text{r}}[\mathbf{x}(t)]\, \mu_{\text{r}}[\mathbf{x}(t)]$, as before. The end result is the same path integral (22.172), but with more technical complication in the derivation.

22.3.5 Worldline Path Integral for Casimir–Polder Potentials

Now if we are interested in the interaction of an atom with a magnetodielectric body (or bodies), we can adapt the classical result Eq. (22.106) to the present case:,

$$V_{\rm CP}(\mathbf{r}) = \frac{\alpha_0}{\epsilon_0}\left(\frac{\delta E_{\rm EM}}{\delta \epsilon_{\rm r}}\right) + \beta_0\mu_0\left(\frac{\delta E_{\rm EM}}{\delta \mu_{\rm r}}\right), \qquad \begin{matrix}(22.182)\\ \text{(Casimir–Polder potential)}\end{matrix}$$

where we are ignoring dispersion (of the atom or the bodies), α_0 is the (static) polarizability of the atom, and β_0 is the (static) magnetizability. Note that this expression is invariant under the duality transformation $\epsilon_{\rm r} \longleftrightarrow \mu_{\rm r}$, $\alpha_0/\epsilon_0 \longleftrightarrow \beta_0\mu_0$.[14]

To evaluate the functional derivatives here, we will first need the partial (functional) derivatives corresponding to the first factor in Eq. (22.177),

$$\begin{aligned}\frac{\delta}{\delta\epsilon_{\rm r}}\int d^{D-1}x_0\,\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2} &= -\frac{\mu_{\rm r}}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}\\ \frac{\delta}{\delta\mu_{\rm r}}\int d^{D-1}x_0\,\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2} &= -\frac{\epsilon_{\rm r}}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}.\end{aligned} \qquad (22.183)$$

These derivatives require some explanation. Here we are regarding $E_{\rm EM}[\epsilon_{\rm r}(\mathbf{r}), \mu_{\rm r}(\mathbf{r})]$ as a functional of the relative permittivity and permeability, which are themselves functions over all space. The integration over all space, as in the Hamiltonian (22.125), reduces these functions to a scalar, and thus define the scalar nature of the functional. But when we vary a path integral of the form

$$\delta\int d^{D-1}x_0\,\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2} = \int d^{D-1}x_0\left(-\frac{\langle\mu_{\rm r}\,\delta\epsilon_{\rm r}\rangle}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2} - \frac{\langle\epsilon_{\rm r}\,\delta\mu_{\rm r}\rangle}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}\right), \qquad (22.184)$$

we must identify the derivatives via the inner products of the form

$$\delta E[\epsilon_{\rm r},\mu_{\rm r}] = \left\langle\frac{\delta E}{\delta\epsilon_{\rm r}},\,\delta\epsilon_{\rm r}\right\rangle + \left\langle\frac{\delta E}{\delta\mu_{\rm r}},\,\delta\mu_{\rm r}\right\rangle. \qquad (22.185)$$

The integration over $\mathbf{x}_0$ is an obvious candidate for the inner product, but in the path integral, it is not so simple, because $\epsilon_{\rm r}$ and $\mu_{\rm r}$ are now functions of the coordinates $x(\tau)$ along the entire path. Thus, the path average (or, equivalently, an integral along the path time τ) is *also* necessary to define a scalar result, and thus a functional, which is the reduction to a scalar by integration over all path averages, where the paths are rigid translations of a single path. The ensemble average over paths is not necessary here to define the inner product, as it implements the integration over the intermediate coordinates $\mathbf{x}_1,\ldots,\mathbf{x}_{N-1}$, whose dependence on $\mathbf{x}_0$ has already been removed by considering the rigid translation of paths. Thus, Eq. (22.184) becomes

$$\delta\int d^{D-1}x_0\,\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2} = \int d^{D-1}x_0\left\langle-\frac{\mu_{\rm r}\,\delta\epsilon_{\rm r}}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2} - \frac{\epsilon_{\rm r}\,\delta\mu_{\rm r}}{2}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}\right\rangle, \qquad (22.186)$$

from which the derivatives (22.183) follow. A check on the the choice of inner product for the functional differentiation is that we should be able to set $\delta\epsilon_{\rm r}(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}_0)$ and $\delta\mu = 0$ to project the functional

[14] Stefan Yoshi Buhmann and Stefan Scheel, "Macroscopic Quantum Electrodynamics and Duality," *Physical Review Letters* **102**, 140404 (2009) (doi: 10.1103/PhysRevLett.102.140404); Hassan Safari, Dirk-Gunnar Welsch, Stefan Yoshi Buhmann, and Stefan Scheel, "van der Waals potentials of paramagnetic atoms," *Physical Review A* **78**, 062901 (2008) (doi: 10.1103/PhysRevA.78.062901).

derivative $\delta E/\delta\epsilon_{\rm r}$ at $\mathbf{r}_0$ out of the inner product (22.185). This leads us to evaluate an integral of the form

$$\begin{aligned}\int d^{D-1}x_0 \left\langle \mu_{\rm r}(\mathbf{x})\,\delta(\mathbf{x}-\mathbf{r}_0)\right\rangle_{\mathbf{x}(t)} &= \frac{1}{\mathcal{T}}\int d^{D-1}x_0 \int_0^{\mathcal{T}} d\tau\,\mu_{\rm r}[\mathbf{x}(\tau)]\,\delta[\mathbf{x}(\tau)-\mathbf{r}_0]\\ &= \frac{1}{N}\int d^{D-1}x_0 \sum_{j=0}^{N-1}\mu_{\rm r}(\mathbf{x}_j)\,\delta(\mathbf{x}_j-\mathbf{r}_0)\\ &= \frac{1}{N}\int d^{D-1}x_0 \sum_{j=0}^{N-1}\mu_{\rm r}[\mathbf{x}_0+(\mathbf{x}_j-\mathbf{x}_0)]\,\delta[\mathbf{x}_0+(\mathbf{x}_j-\mathbf{x}_0)-\mathbf{r}_0]\\ &= \frac{1}{N}\sum_{j=0}^{N-1}\mu_{\rm r}(\mathbf{r}_0)\\ &= \mu_{\rm r}(\mathbf{r}_0),\end{aligned} \tag{22.187}$$

where we switched to the discrete representation of the path average and used the fact that $(\mathbf{x}_j-\mathbf{x}_0)$ is constant for any rigid translation of a particular path. This procedure also leads to the derivatives (22.183).

We will also need functional derivatives corresponding to the exponent (i.e., the potential) of the path integral. Writing the TM potential from Eqs. (22.155) in the form

$$\begin{aligned}V_{\rm TM} &= \frac{1}{2}\left[\left(\nabla\log\sqrt{\epsilon_{\rm r}}\right)^2-\nabla^2\log\sqrt{\epsilon_{\rm r}}\right]\\ &= \frac{1}{8}\left(\nabla\log\epsilon_{\rm r}\right)^2-\frac{1}{4}\nabla^2\log\epsilon_{\rm r}\\ &= \frac{3\left(\nabla\epsilon_{\rm r}\right)^2}{8\epsilon_{\rm r}^2}-\frac{\nabla^2\epsilon_{\rm r}}{4\epsilon_{\rm r}},\end{aligned} \tag{22.188}$$

we can write the partial derivatives for the TM potential as

$$\begin{aligned}\frac{\partial}{\partial\epsilon_{\rm r}}V_{\rm TM} &= -\frac{3\left(\nabla\epsilon_{\rm r}\right)^2}{4\epsilon_{\rm r}^3}+\frac{\nabla^2\epsilon_{\rm r}}{4\epsilon_{\rm r}^2} = -\frac{1}{4}\left(\frac{\left(\nabla\epsilon_{\rm r}\right)^2}{\epsilon_{\rm r}^3}+\nabla^2{\epsilon_{\rm r}}^{-1}\right) = \frac{1}{4\epsilon_{\rm r}}\left(\nabla^2\log\epsilon_{\rm r}-2\left(\nabla\log\epsilon_{\rm r}\right)^2\right)\\ \frac{\partial}{\partial\nabla\epsilon_{\rm r}}V_{\rm TM} &= \frac{3\nabla\epsilon_{\rm r}}{4\epsilon_{\rm r}^2} = -\frac{3}{4}\nabla{\epsilon_{\rm r}}^{-1} = \frac{3}{4\epsilon_{\rm r}}\nabla\log\epsilon_{\rm r}\\ \frac{\partial}{\partial\nabla^2\epsilon_{\rm r}}V_{\rm TM} &= -\frac{1}{4\epsilon_{\rm r}},\end{aligned} \tag{22.189}$$

where we used $\partial_x h^{-1} = -h'/h^2$, $\partial_x^2 h^{-1} = -h''/h^2+2h'^2/h^3$, $\partial_x\log h = -h'/h$, and $\partial_x^2\log h = h''/h-h'^2/h^2$, with similar expressions for the corresponding $\mu_{\rm r}$ derivatives of the TE potential. Then, recalling that for a functional of the form

$$F[x,x_t,x_{tt},\ldots;t] = \int_{t_1}^{t_2} f(t,x,x_t,x_{tt})\,dt, \tag{22.190}$$

where $x_t\equiv\partial x/\partial t$, the functional derivative is [see Eq. (8.12)]

$$\frac{\delta F}{\delta x} = \frac{\partial f}{\partial x}-\frac{d}{dt}\frac{\partial f}{\partial x_t}+\frac{d^2}{dt^2}\frac{\partial f}{\partial x_{tt}}. \tag{22.191}$$

Adapting the functional derivative to the present case, we will need the following derivatives of the partial

derivatives (22.189),

$$\begin{aligned}
\left(\frac{\partial}{\partial\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) &= \frac{1}{4\epsilon_{\rm r}}\left(\nabla^2\log\epsilon_{\rm r} - 2\big(\nabla\log\epsilon_{\rm r}\big)^2\right)P(\mathbf{r}) \\
-\nabla\cdot\left(\frac{\partial}{\partial\nabla\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) &= -\frac{3}{4\epsilon_{\rm r}}\left(\nabla^2\log\epsilon_{\rm r} - \big(\nabla\log\epsilon_{\rm r}\big)^2\right)P(\mathbf{r}) - \frac{3}{4\epsilon_{\rm r}}\left(\nabla\log\epsilon_{\rm r}\right)\cdot\nabla P(\mathbf{r}) \\
\nabla^2\left(\frac{\partial}{\partial\nabla^2\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) &= \frac{1}{4\epsilon_{\rm r}}\left(\nabla^2\log\epsilon_{\rm r} - \big(\nabla\log\epsilon_{\rm r}\big)^2\right)P(\mathbf{r}) + \frac{1}{2\epsilon_{\rm r}}\left(\nabla\log\epsilon_{\rm r}\right)\cdot\nabla P(\mathbf{r}) - \frac{1}{4\epsilon_{\rm r}}\nabla^2 P(\mathbf{r}),
\end{aligned} \tag{22.192}$$

where $P(\mathbf{r})$ is an arbitrary function that represents the rest of the integrand. Combining these terms gives

$$\begin{aligned}
&\left(\frac{\partial}{\partial\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) - \nabla\cdot\left(\frac{\partial}{\partial\nabla\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) + \nabla^2\left(\frac{\partial}{\partial\nabla^2\epsilon_{\rm r}}V_{\rm TM}\right)P(\mathbf{r}) \\
&\qquad = -\frac{1}{4\epsilon_{\rm r}}\left(\nabla^2\log\epsilon_{\rm r}\right)P(\mathbf{r}) - \frac{1}{4\epsilon_{\rm r}}\left(\nabla\log\epsilon_{\rm r}\right)\cdot\nabla P(\mathbf{r}) - \frac{1}{4\epsilon_{\rm r}}\nabla^2 P(\mathbf{r}) \\
&\qquad = -\frac{1}{4\epsilon_{\rm r}}\Big[\left(\nabla^2\log\epsilon_{\rm r}\right) + \left(\nabla\log\epsilon_{\rm r}\right)\cdot\nabla + \nabla^2\Big]P(\mathbf{r}).
\end{aligned} \tag{22.193}$$

Finally, assembling all the parts from Eqs. (22.183) and (22.193) the functional derivatives of Eq. (22.177) become

$$\begin{aligned}
\frac{\delta E^{(\rm TE)}_{\rm EM}}{\delta\epsilon_{\rm r}} &= \frac{\hbar c}{4(2\pi)^{D/2}}\int_0^\infty\frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \mu_{\rm r}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TE}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\
\frac{\delta E^{(\rm TE)}_{\rm EM}}{\delta\mu_{\rm r}} &= \frac{\hbar c}{4(2\pi)^{D/2}}\int_0^\infty\frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \\
&\quad\times\left\langle\!\!\left\langle \epsilon_{\rm r}\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TE}\rangle_{\mathbf{x}(\tau)}} - \frac{\mathcal{T}}{2\mu_{\rm r}}\Big[\left(\nabla^2\log\mu_{\rm r}\right) + \left(\nabla\log\mu_{\rm r}\right)\cdot\nabla + \nabla^2\Big]\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TE}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)},
\end{aligned} \tag{22.194}$$

with analogous expressions for the TM case. Putting these functional derivatives into Eq. (22.182) leads to the somewhat cumbersome expression

$$\begin{aligned}
V^{(\rm TE)}_{\rm CP}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^{D/2}}\int_0^\infty\frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\Bigg\langle\!\!\Bigg\langle &\left(\frac{\alpha_0\mu_{\rm r}}{\epsilon_0} + \beta_0\mu_0\epsilon_{\rm r}\right)\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TE}\rangle_{\mathbf{x}(\tau)}} \\
&- \frac{\beta_0\mu_0\mathcal{T}}{2\mu_{\rm r}}\Big[\left(\nabla^2\log\mu_{\rm r}\right) + \left(\nabla\log\mu_{\rm r}\right)\cdot\nabla + \nabla^2\Big]\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TE}\rangle_{\mathbf{x}(\tau)}}\Bigg\rangle\!\!\Bigg\rangle_{\mathbf{x}(\tau)}.
\end{aligned}$$

(unrenormalized Casimir–Polder potential, TE polarization) (22.195)

For TM polarization, the expression is essentially the same, but with $\epsilon_{\rm r}$ and $\mu_{\rm r}$ reversed, and correspondingly with α_0/ϵ_0 and $\beta_0\mu_0$ reversed. Thus, the resulting TM expression is

$$\begin{aligned}
V^{(\rm TM)}_{\rm CP}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^{D/2}}\int_0^\infty\frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\Bigg\langle\!\!\Bigg\langle &\left(\beta_0\mu_0\epsilon_{\rm r} + \frac{\alpha_0\mu_{\rm r}}{\epsilon_0}\right)\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-3/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TM}\rangle_{\mathbf{x}(\tau)}} \\
&- \frac{\alpha_0\mathcal{T}}{2\epsilon_0\epsilon_{\rm r}}\Big[\left(\nabla^2\log\epsilon_{\rm r}\right) + \left(\nabla\log\epsilon_{\rm r}\right)\cdot\nabla + \nabla^2\Big]\langle\epsilon_{\rm r}\mu_{\rm r}\rangle^{-1/2}_{\mathbf{x}(\tau)}\,e^{-\mathcal{T}\langle V_{\rm TM}\rangle_{\mathbf{x}(\tau)}}\Bigg\rangle\!\!\Bigg\rangle_{\mathbf{x}(\tau)}.
\end{aligned}$$

(unrenormalized Casimir–Polder potential, TM polarization) (22.196)

In the case of a nonmagnetic atom ($\beta_0 = 0$) and a nonmagnetic material ($\mu_{\rm r} = 1$ and $V_{\rm TE} = 0$), the TE

expression simplifies drastically:

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}$$

(unrenormalized, nonmagnetic Casimir–Polder potential, TE) (22.197)

If we make the same assumptions in the TM case and further assume $\nabla\epsilon_{\text{r}}(\mathbf{r}) = 0$, so that the atom is in a region of constant ϵ_{r} (i.e., if the atom is in vacuum outside a dielectric body), then the TM expression also simplifes, though not quite as much:

$$\begin{aligned} V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}} \nabla^2 \left[\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} \right] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}} \nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \end{aligned}$$

(unrenormalized, nonmagnetic Casimir–Polder potential, TM) (22.198)

Recall that the magnetic path-integral potential is defined in (22.155) as

$$V_{\text{TM}}(\mathbf{r}) := \frac{1}{2}\left[\left(\nabla \log\sqrt{\epsilon_{\text{r}}}\right)^2 - \nabla^2 \log\sqrt{\epsilon_{\text{r}}} \right]. \qquad (22.199)$$

(matter-induced potentials)

Also, note that in all these expressions, the paths (22.174) are defined after the functional derivatives such that

$$\mathbf{x}(\tau) = \mathbf{r} + \mathbf{B}_{\mathcal{T}}(\tau). \qquad (22.200)$$

(stochastic path, scalar EM)

That is, the atomic location $\mathbf{r}$ in the interaction potential is the beginning and end-point of the paths (Brownian bridges); the integral over this terminus point was lost in the functional differentiation. Note that, had we derived analogous expressions to (22.197) and (22.198) for a *magnetic* atom interacting with a magnetic surface, we would obtain the same results, but with the following changes: α_0 is replaced by $\beta_0\mu^2$, ϵ_{r} is replaced by μ_{r}, ϵ_0 is replaced by μ_0, TE and TM polarizations are interchanged, and there is an extra overall minus sign for both polarizations. That is, the geometry dependence would be the same, but the *total force* would change sign.

22.3.5.1 Interaction with a Point Particle: Path-Integral View (TE Polarization)

Thus far, we have relied on a rather formal expression (22.182) for the Casimir–Polder potential in terms of functional derivatives of the energy in path-integral form. To gain a more intuitive view of this interaction, we can also exhibit this interaction *directly* in the path integral. To do this, let's return to the path integral (22.177), and specialize to the TE, nonmagnetic ($\mu_{\text{r}} = 1$) energy

$$E_{\text{EM}}^{(\text{TE})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \qquad (22.201)$$

Now let's consider the interaction of a point-dipole atom with a dielectric body (or bodies), which amounts to separating out the atomic contribution from the permittivity as

$$\epsilon_{\text{r}}(\mathbf{x}) \longrightarrow \epsilon_{\text{r}}(\mathbf{x}) + \frac{\alpha_0}{\epsilon_0}\delta^{D-1}(\mathbf{x}-\mathbf{r}), \qquad (22.202)$$

where the atom (of static polarizability α_0) is at position $\mathbf{r}$. Making this replacement in the path integral to obtain the Casimir–Polder potential,

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \left\langle \epsilon_{\text{r}}(\mathbf{x}) + (\alpha_0/\epsilon_0)\,\delta^{D-1}(\mathbf{x}-\mathbf{r}) \right\rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \left[\langle \epsilon_{\text{r}}(\mathbf{x}) \rangle + \frac{\alpha_0}{\epsilon_0 \mathcal{T}} \wp[\mathbf{x}(\tau);\mathbf{r}] \right]_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} , \end{aligned} \qquad (22.203)$$

where we have defined the point-occupation time in $D-1$ dimensions

$$\wp[\mathbf{y}(t');\mathbf{r}] := \int_0^t dt'\, \delta^{D-1}[\mathbf{y}(t') - \mathbf{r}] \tag{22.204}$$

for $\mathbf{y}(t)$ at $\mathbf{r}$, assuming that the process $\mathbf{y}$ runs from 0 to t. Note that in one dimension, this reduces to the local time of $\mathbf{y}$, whereas in higher dimensions this statistic has the typical value of zero unless $\mathbf{r}$ is the starting point of the process.

Expanding to lowest order in α_0, using $(a+b)^{-1/2} = a^{-1/2} - b/2a^{3/2} + O(b^2)$, we then have

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}}(\mathbf{x})\rangle_{\mathbf{x}(\tau)}^{-1/2} - \frac{\alpha_0 \wp[\mathbf{x}(\tau);\mathbf{r}]}{2\epsilon_0 \mathcal{T} \langle \epsilon_{\text{r}}(\mathbf{x})\rangle_{\mathbf{x}(\tau)}^{3/2}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.205}$$

The first term disappears when we renormalize against the atom and ϵ_{r} body being separated by an arbitrarily large distance. The integral renormalized thusly is

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \frac{\alpha_0 \wp[\mathbf{x}(\tau);\mathbf{r}]}{\mathcal{T} \langle \epsilon_{\text{r}}(\mathbf{x})\rangle_{\mathbf{x}(\tau)}^{3/2}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.206}$$

Due to the presence of $\wp$, only paths $\mathbf{x}(\tau)$ that intersect $\mathbf{r}$ will actually contribute to the path average due to the presence of $\wp$ in the integrand. However, any path that passes through $\mathbf{r}$ is equivalent to a path that *begins* at $\mathbf{r}$, and the path itself defines an equivalence class of paths that are cyclic permutations of the path in time, starting at different points along the path. For the purposes of the path average over the permittivity, we will replace all such paths by the equivalent path at $\mathbf{r}$, which is generated with the same probability as each possible path by virtue of being a path. Thus, we replace the contribution of the path $\mathbf{x}(\tau)$ by the contribution of the path $\mathbf{x}_{\mathbf{r}}(\tau)$, which is equivalent except that $\mathbf{x}_{\mathbf{r}}(0) = \mathbf{x}(\tau) = \mathbf{r}$,

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \frac{\alpha_0 \wp[\mathbf{x}(\tau);\mathbf{r}]}{\mathcal{T} \langle \epsilon_{\text{r}}(\mathbf{x})\rangle_{\mathbf{x}_{\mathbf{r}}(\tau)}^{3/2}} \right\rangle\!\!\right\rangle_{\mathbf{x}_{\mathbf{r}}(\tau)}, \tag{22.207}$$

and by changing the ensemble average to reference $\mathbf{x}_{\mathbf{r}}$ instead of $\mathbf{x}$, we have effectively dropped the null contribution of any path that doesn't pass through $\mathbf{r}$. Then the only part of the integrand that depends on $\mathbf{x}_0$ is the $\wp$ factor, since this refers to the original path $\mathbf{x}(\tau)$. By shifting the path, we can carry out this integral:

$$\int d^{D-1}x_0\, \wp[\mathbf{x}(\tau);\mathbf{r}] = \int d^{D-1}x_0\, \wp[\mathbf{x}_{\mathbf{r}}(\tau);\mathbf{x}_0] = \mathcal{T}. \tag{22.208}$$

The last equality here follows from integrating the definition (22.204) over all $\mathbf{r}$. [This is essentially the same argument we used in Eq. (22.187), but here in continuous notation.] Thus, we finally have

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}}\rangle_{\mathbf{x}_{\mathbf{r}}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}_{\mathbf{r}}(\tau)}. \tag{22.209}$$

This is equivalent to the result (22.197) that we derived from functional differentiation.

Note that in this approach, we have perturbed the dielectric permittivity to introduce an atom without regard for what fields, potentials, or sources to hold fixed, as we discussed at length in Section 22.2. Here, we have little choice: the fields and sources (the sources being effective sources due to vacuum fluctuations) are buried in the path integral (22.201), and so we *assume* that the whatever needs to be fixed is so. This is justified from our procedure in Section 22.3.5, which was based on the more careful analysis of electromagnetism.

22.3.5.2 Interaction with a Point Particle: Path-Integral View (TM Polarization)

In the TM case, we can return to Eq. (22.177) for a nonmagnetic interaction,

$$E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.210}$$

where we now must also deal with the potential defined in (22.155):

$$V_{\text{TM}}(\mathbf{r}) := \frac{1}{8}\big(\nabla \log \epsilon_{\text{r}}\big)^2 - \frac{1}{4}\nabla^2 \log \epsilon_{\text{r}}. \tag{22.211}$$

Again letting

$$\epsilon_{\text{r}}(\mathbf{x}) \longrightarrow \epsilon_{\text{r}}(\mathbf{x}) + \frac{\alpha_0}{\epsilon_0}\delta^{D-1}(\mathbf{x}-\mathbf{r}), \tag{22.212}$$

we will expand to lowest order in the atomic polarizability. Note that we will simply expand in spite of the delta function, which only makes sense when the delta function is regularized by a localized function of finite height. This expansion is more involved than in the TE case, because of the nonlinear functions of ϵ_{r} in the potential that appear inside the path average. Expanding the potential first, using $\log(a+b) + b/a + O(b^2)$, the potential with the atom separated out is

$$V_{\text{TM}}(\mathbf{x}) \longrightarrow \frac{1}{8}\left(\nabla \log \epsilon_{\text{r}} + \frac{\alpha_0}{\epsilon_0}\nabla \frac{1}{\epsilon_{\text{r}}}\delta^{D-1}(\mathbf{x}-\mathbf{r})\right)^2 - \frac{1}{4}\left(\nabla^2 \log \epsilon_{\text{r}} + \frac{\alpha_0}{\epsilon_0}\nabla^2 \frac{1}{\epsilon_{\text{r}}}\delta^{D-1}(\mathbf{x}-\mathbf{r})\right). \tag{22.213}$$

Expanding to lowest order in α_0,

$$V_{\text{TM}}(\mathbf{x}) \longrightarrow \frac{1}{8}\big(\nabla \log \epsilon_{\text{r}}\big)^2 - \frac{1}{4}\nabla^2 \log \epsilon_{\text{r}} - \frac{\alpha_0}{4\epsilon_0}\Big[(\nabla \log \epsilon_{\text{r}})\cdot\nabla - \nabla^2\Big]\frac{1}{\epsilon_{\text{r}}}\delta^{D-1}(\mathbf{x}-\mathbf{r}), \tag{22.214}$$

gives the original potential plus an atomic interaction term. Thus, the replacement (22.212) in the path integral (22.210) gives

$$\begin{aligned} E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \\ \times \left\langle\!\!\left\langle \left\langle \epsilon_{\text{r}} + \frac{\alpha_0}{\epsilon_0}\delta^{D-1}(\mathbf{x}-\mathbf{r}) \right\rangle_{\mathbf{x}(\tau)}^{-1/2} e^{-\mathcal{T}\left\langle V_{\text{TM}} - (\alpha_0/4\epsilon_0)[(\nabla \log \epsilon_{\text{r}})\cdot\nabla - \nabla^2]\epsilon_{\text{r}}^{-1}\delta^{D-1}(\mathbf{x}-\mathbf{r})\right\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \end{aligned} \tag{22.215}$$

Now expanding to first order in α_0, and dropping the zeroth-order component (which disappears in renormalization against the configuration sans atom), we obtain

$$\begin{aligned} E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \\ \times \left\langle\!\!\left\langle -\frac{\alpha_0 \big\langle \delta^{D-1}(\mathbf{x}-\mathbf{r})\big\rangle_{\mathbf{x}(\tau)}}{2\epsilon_0 \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{3/2}} e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right. \\ \left. - \frac{\alpha_0 \mathcal{T}}{4\epsilon_0 \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{1/2}} \left\langle \Big[(\nabla \log \epsilon_{\text{r}})\cdot\nabla - \nabla^2\Big]\frac{1}{\epsilon_{\text{r}}}\delta^{D-1}(\mathbf{x}-\mathbf{r}) \right\rangle_{\mathbf{x}(\tau)} e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \end{aligned} \tag{22.216}$$

Now we can integrate by parts twice on the second term, and note that the path average of the delta function takes care of the $\mathbf{x}_0$-integration as in the last section. The result is

$$\begin{aligned} E_{\text{EM}}^{(\text{TM})} = \frac{\hbar c \alpha_0}{2(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \\ \times \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}} \Big[(\nabla^2 \log \epsilon_{\text{r}}) + (\nabla \log \epsilon_{\text{r}})\cdot\nabla + \nabla^2\Big] \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \end{aligned} \tag{22.217}$$

which is equivalent to Eq. (22.196) for $\mu_\mathrm{r} = 1$ and $\beta_0 = 0$, which we derived by functional differentiation, and also to Eq. (22.198) if the gradient of ϵ_r vanishes at the atomic location.

22.4 Casimir–Polder Potential Near a Dielectric Half-Space

To illustrate the Casimir path integrals we have developed so far, we will evaluate some examples, in particular for the Casimir–Polder path integrals (22.197) and (22.198) for an atom near a (planar) dielectric half-space.

22.4.1 TE Polarization: Strong Coupling

Let's begin with a test-drive of the TE path integral (22.197) in a simple regime: a perfectly conducting, planar surface:

$$V_\mathrm{CP}^{(\mathrm{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\epsilon_\mathrm{r}\rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.218}$$

Recall that we must renormalize this by subtracting the same path integral with the atom and surface separated by an arbitrarily large distance. In this case it suffices to subtract the same integral with $\epsilon_\mathrm{r} = 1$:

$$V_\mathrm{CP}^{(\mathrm{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\epsilon_\mathrm{r}\rangle_{\mathbf{x}(\tau)}^{-3/2} - 1 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.219}$$

For a perfect conductor, the susceptibility of the surface is very large. If any part of the path touches the surface, that part will dominate the path average, causing it to diverge, and thus the inverted average $\langle\epsilon_\mathrm{r}\rangle_{\mathbf{x}(\tau)}^{-3/2}$ will vanish for any path that touches the surface. With the renormalization term, the quantity in the ensemble average is 0 for paths *not* touching the surface, and -1 for paths touching the surface. *With* the path average, the ensemble average is -1 times the probability that the paths touch the surface. The relevant path statistic is the boundary-crossing probability, which for a standard Brownian bridge is [Eq. (17.380)]

$$P_\mathrm{cross}(d) = e^{-2d^2}, \tag{22.220}$$

where d is the distance of the boundary from the initial and final point of the bridge. The paths here, defined in Eq. (22.200), go to time $\mathcal{T}$, so they are effectively "larger" by a factor $\sqrt{\mathcal{T}}$. We can adapt the statistic by setting $d = z/\sqrt{\mathcal{T}}$, where z is the distance to the surface. Note that this applies to only one of the spatial dimensions; the others are irrelevant to the problem except for normalization. Then the path integral becomes

$$\begin{aligned} V_\mathrm{CP}^{(\mathrm{TE})}(\mathbf{r}) &= -\frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} P_\mathrm{cross}\left(\frac{z}{\sqrt{\mathcal{T}}}\right) \\ &= -\frac{\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} P_\mathrm{cross}\left(\frac{1}{\sqrt{2\mathcal{T}}}\right) \\ &= -\frac{\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} e^{-1/\mathcal{T}}, \end{aligned} \tag{22.221}$$

where we rescaled $\mathcal{T}$ in the second step to remove the z dependence from the crossing probability and to introduce a factor of 2, and then we inserted the explicit crossing probability. The result is

$$V_\mathrm{CP}^{(\mathrm{TE})}(\mathbf{r}) = -\frac{\Gamma(D/2)\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D}.$$

(Casimir–Polder potential, TE strong coupling) (22.222)

Setting $D = 4$ we obtain

$$V_\mathrm{CP}^{(\mathrm{TE})}(\mathbf{r}) = -\frac{\hbar c \alpha_0}{64\pi^2\epsilon_0 z^4} = -\frac{3\hbar c \alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{1}{6}\right),$$

(Casimir–Polder potential, TE strong coupling) (22.223)

which is the standard result for the TE part of the total potential [see Eq. (14.214) and the following discussion]. We can also write out the result for one-dimensional electromagnetism ($D = 2$),

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2}, \tag{22.224}$$

which is not a standard result, but easily verified by mode-summation (Problem 22.9).

22.4.2 TE Polarization: Weak Coupling

Now, in the limit of a rarefied dielectric surface, we can write

$$\epsilon_{\text{r}}(\mathbf{r}) = 1 + \chi(\mathbf{r}), \tag{22.225}$$

and calculate to first order in the susceptibility χ. In this limit, the TE path integral (22.197) becomes

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0}{8(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \chi \rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.226}$$

where we have already dropped the χ-independent (zeroth-order) term in renormalization. For a dielectric half-space, we have

$$\chi(z) = \chi\Theta(-z), \tag{22.227}$$

where the dielectric occupies $z < 0$ and we will compute the potential for $z > 0$. The path integral then becomes

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0 \chi}{8(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \Theta(-z) \rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.228}$$

The relevant statistic here is the **sojourn time**, measuring the time a process $y(t)$ spends beyond a barrier at d [Eq. (17.499):

$$T_{\text{s}}[y(t); d] := \int_0^t dt' \, \Theta[y(t') - d]. \tag{22.229}$$

Thus, the potential is

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0 \chi}{8(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{2+D/2}} \left\langle\!\!\left\langle T_{\text{s}}[x(\tau); z] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.230}$$

where we have shifted the coordinates such that $x(0) = x(\tau) = 0$, and we have reflected the coordinates in view of the martingale nature of $x(\tau)$. We will again want to transform the sojourn time of $x(\tau)$ into the sojourn time of a standard Brownian bridge, again with $\tau = t\mathcal{T}$:

$$\begin{aligned} T_{\text{s}}[x(\tau); z] &= \int_0^{\mathcal{T}} d\tau \, \Theta[x(\tau) - z] \\ &= \mathcal{T} \int_0^1 dt \, \Theta[x(t\mathcal{T}) - z] \\ &= \mathcal{T} \int_0^1 dt \, \Theta[\sqrt{\mathcal{T}} B(t) - z] \\ &= \mathcal{T} \int_0^1 dt \, \Theta\left[B(t) - \frac{z}{\sqrt{\mathcal{T}}}\right] \\ &= \mathcal{T} \, T_{\text{s}}\left[B(t); \frac{z}{\sqrt{\mathcal{T}}}\right]. \end{aligned} \tag{22.231}$$

We then have the potential

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= -\frac{3\hbar c\alpha_0\chi}{8(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle T_{\text{s}}\left[B(t);\frac{z}{\sqrt{\mathcal{T}}}\right]\right\rangle\!\!\right\rangle_{B(t)} \\ &= -\frac{3\hbar c\alpha_0\chi}{8(4\pi)^{D/2}\epsilon_0 z^D}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle T_{\text{s}}\left[B(t);\frac{1}{\sqrt{2\mathcal{T}}}\right]\right\rangle\!\!\right\rangle_{B(t)}, \end{aligned} \tag{22.232}$$

where we have scaled away z in the integral by letting $\mathcal{T} \longrightarrow 2z^2\mathcal{T}$. The mean sojourn time for a standard Brownian bridge is given by [Eq. (17.588)]

$$\left\langle\!\!\left\langle T_{\text{s}}[B(t);d]\right\rangle\!\!\right\rangle = \frac{e^{-2d^2}}{2} - \sqrt{\frac{\pi}{2}}\,d\,\text{erfc}\left[\sqrt{2}\,d\right], \tag{22.233}$$

and so

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c\alpha_0\chi}{16(4\pi)^{D/2}\epsilon_0 z^D}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left[e^{-1/\mathcal{T}} - \sqrt{\frac{\pi}{\mathcal{T}}}\,\text{erfc}\left(\frac{1}{\sqrt{\mathcal{T}}}\right)\right]. \tag{22.234}$$

Carrying out the integral, the result is

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\Gamma(D/2)\hbar c\alpha_0\chi}{16(D+1)(4\pi)^{D/2}\epsilon_0 z^D}.$$

(Casimir–Polder potential, TE weak coupling) (22.235)

Putting $D = 4$, we find the result

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{\chi}{40}\right).$$

(Casimir–Polder potential, TE weak coupling) (22.236)

This is the standard result for the TE part of the total potential in this regime [see Eq. (14.214) and the following discussion]. Again, the alternate one-dimensional ($D = 2$) case is

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{\hbar c\alpha_0\chi}{64\pi\epsilon_0 z^2}, \tag{22.237}$$

as can be verified by mode-summation (Problem 22.8).

22.4.3 TE Polarization: General Coupling

Using basically the same methods, in the TE case at least, we can derive the potential for an atom interacting with a dielectric half-space of arbitrary (dispersionless) susceptibility χ. Returning to the TE path integral (22.197) after renormalization,

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2} - 1\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.238}$$

In this case, we can write the path average for the permittivity as

$$\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)} = \langle 1+\chi\rangle_{\mathbf{x}(\tau)} = 1 + \chi T_{\text{s}}\left[B(\tau);\frac{z}{\sqrt{\mathcal{T}}}\right], \tag{22.239}$$

in terms of the sojourn time for the path, which represents the fraction of the path inside the dielectric, which is a distance z from the atom, and we have rescaled the paths into standard Brownian bridges. Then

the potential in terms of the sojourn time is

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(1 + \chi T_{\text{s}}\left[B(\tau); \frac{z}{\sqrt{\mathcal{T}}}\right]\right)^{-3/2} - 1 \right\rangle\!\!\right\rangle_{B(\tau)} \\ &= -\frac{\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle 1 - \left(1 + \chi T_{\text{s}}\left[B(\tau); \frac{1}{\sqrt{2\mathcal{T}}}\right]\right)^{-3/2} \right\rangle\!\!\right\rangle_{B(\tau)}, \end{aligned} \tag{22.240}$$

To compute the path average here, we will write

$$\begin{aligned} &\left\langle\!\!\left\langle 1 - (1+\chi T_{\text{s}})^\alpha \right\rangle\!\!\right\rangle \\ &\quad = \int_0^1 dx \, [1 - (1+\chi x)^\alpha] \, f_{T_{\text{s}}}(x) \\ &\quad = \int_0^1 dx \, [1 - (1+\chi x)^\alpha] \left[\sqrt{\frac{4(1-x)}{\pi x \mathcal{T}}} \, e^{-1/(1-x)\mathcal{T}} + \left(1 - \frac{2}{\mathcal{T}}\right) e^{-1/\mathcal{T}} \operatorname{erfc}\left(\sqrt{\frac{x}{(1-x)\mathcal{T}}}\right)\right], \end{aligned} \tag{22.241}$$

where we are integrating against the probability density for the standard Brownian bridge from Eq. (17.584),

$$f_{T_{\text{s}}}(x) = \left[1 - e^{-2d^2}\right] \delta(x - 0^+) + \sqrt{\frac{8d^2(1-x)}{\pi x}} \, e^{-2d^2/(1-x)} + (1 - 4d^2) \, e^{-2d^2} \operatorname{erfc}\left(\sqrt{\frac{2d^2 x}{1-x}}\right), \tag{22.242}$$

with $d = 1/\sqrt{2\mathcal{T}}$. Then computing the $\mathcal{T}$ integral for $D = 4$,

$$\begin{aligned} &\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle 1 - (1+\chi T_{\text{s}})^\alpha \right\rangle\!\!\right\rangle \\ &\quad = \int_0^1 dx \, [1 - (1+\chi x)^\alpha] \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left[\sqrt{\frac{4(1-x)}{\pi x \mathcal{T}}} \, e^{-1/(1-x)\mathcal{T}} + \left(1 - \frac{2}{\mathcal{T}}\right) e^{-1/\mathcal{T}} \operatorname{erfc}\left(\sqrt{\frac{x}{(1-x)\mathcal{T}}}\right)\right] \\ &\quad = \int_0^1 dx \, [1 - (1+\chi x)^\alpha] \left[\frac{3(1-\sqrt{x})^2}{2\sqrt{x}}\right] \end{aligned} \tag{22.243}$$

which for $\alpha = -3/2$ becomes

$$\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle 1 - (1+\chi T_{\text{s}})^{-3/2} \right\rangle\!\!\right\rangle = 1 + \frac{6}{\chi} - \frac{3\sqrt{1+\chi}}{\chi} - \frac{3 \sinh^{-1}\sqrt{\chi}}{\chi^{3/2}}. \tag{22.244}$$

Putting this into Eq. (22.240) with $D = 4$,

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0}{32\pi^2 \epsilon_0 z^4} \left(\frac{1}{6} + \frac{1}{\chi} - \frac{\sqrt{1+\chi}}{2\chi} - \frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\right). \tag{22.245}$$

(Casimir–Polder potential, TE)

The parenthetic quantity is η_{TE} from the Green-tensor treatment in the full electromagnetic case [Eq. (14.214)], and the rest of the result is the strong-coupling result from full electromagnetism. This factor is $\chi/40$ and $1/6$ for small and large χ, respectively, demonstrating the consistency with the weak-coupling calculation (22.236) and the strong-coupling version (22.223).

22.4.4 TM Polarization: Strong Coupling

To begin, a disclaimer: in this section we will do something *wrong*, for illustrative purposes. The idea is to now switch to the TM path integral, which is substantially more complicated due to the presence of a divergent potential. It's important to be careful with this potential in order to avoid trouble.

For the TM path integral, evaluated near a planar dielectric interface, we return to Eq. (22.198),

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}} \nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \tag{22.246}$$

We will assume the atom at position $\mathbf{r}$ to be outside the dielectric body, in which case we can set $\epsilon_{\text{r}}(\mathbf{r}) = 1$:

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2} \nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \tag{22.247}$$

There are two main, extra complications here that we didn't have in the TE case: the potential and the derivative. The derivative will be easy to handle, but let's first take a closer look at the potential.

From Eq. (22.155) the potential is

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{2} \left[\left(\nabla \log \sqrt{\epsilon_{\text{r}}} \right)^2 - \nabla^2 \log \sqrt{\epsilon_{\text{r}}} \right] . \tag{22.248}$$

If the dielectric has susceptibility χ, then we can write the permittivity as

$$\epsilon_{\text{r}}(z) = 1 + \chi \Theta(z) \tag{22.249}$$

for a vacuum–dielectric interface, if the dielectric occupies $z > 0$ (we choose this convention to keep the signs simple). Then putting this into the potential, we have a rather divergent mess:

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{2} \left[\left(\log \sqrt{1+\chi} \right)^2 \delta^2(z) - \left(\log \sqrt{1+\chi} \right) \delta'(z) \right] . \tag{22.250}$$

The first term is divergent in the path average of any nonvanishing path (with unit probability). This is not terribly problematic, because it means that any path touching the surface, the (unrenormalized) contribution of that path vanishes, which is precisely what *should* happen in the strong-coupling limit. However, this *shouldn't* happen for arbitrary χ, which is where we are at the moment, so this term is problematic. The second term in the potential is also problematic: the derivative of the delta function picks out the derivative of the path in the path average, which is divergent (either positive or negative) [see the local-time derivative in Eq. (17.660) and the following discussion]. Here, if the overall term swings large and positive, the result is the same as for the second term. If it swings large and negative, then problems can happen, as now the exponential factor can become arbitrarily large.

But let's ignore these problems for the moment and try to work with the potential in the strong-coupling limit. To do this, we will assume that, for sufficiently large χ, that the second term in the potential dominates the first, because asymptotically $\log^2 \chi \gg \log \chi$. This only really makes sense if we regularize the delta functions, replacing them with corresponding narrow and sharply (but finitely) peaked functions. To proceed, we need to evaluate (renormalized) path integrals of the form

$$I(D, \alpha; \mathbf{r}) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-\alpha} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} - 1 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \tag{22.251}$$

The first term in the path average vanishes whenever the path touches the surface, which is a distance z away. Thus, we are back to the crossing probability as in Section 22.4.1. Using the Brownian-bridge crossing probability of e^{-2d^2}, and then setting $d = z/\sqrt{\mathcal{T}}$ as before, we have

$$\begin{aligned} I(D, \alpha; \mathbf{r}) &= -\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \, P_{\text{cross}} \left(\frac{z}{\sqrt{\mathcal{T}}} \right) \\ &= -2^{-D/2} z^{-D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \, P_{\text{cross}} \left(\frac{1}{\sqrt{2\mathcal{T}}} \right) \\ &= -2^{-D/2} z^{-D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \, e^{-1/\mathcal{T}} \\ &= -2^{-D/2} z^{-D} \, \Gamma(D/2). \end{aligned} \tag{22.252}$$

Then writing Eq. (22.247) in terms of this integral as

$$\begin{aligned} V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0}\left[I(D,3/2;z) - \frac{1}{2}\partial_z^2 I(D-2,1/2;z)\right] \\ &= -\frac{\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0}\left[z^{-D}\,\Gamma(D/2) - \partial_z^2 z^{-(D-2)}\,\Gamma[(D-2)/2]\right] \\ &= -\frac{\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D}\left[\Gamma(D/2) - (D-2)(D-1)\Gamma[(D-2)/2]\right]. \end{aligned} \tag{22.253}$$

Simplifying the last factor, we find

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = -\frac{(3-2D)\Gamma(D/2)\hbar c \alpha_0}{4(4\pi)^{D/2}\epsilon_0 z^D}.$$

(Casimir–Polder potential, TM strong coupling) (22.254)

Note that we were a bit slippery here with the derivative. The derivative nominally operates on the atomic position, but this is equivalent to letting it operate on the atom-surface distance. Equivalently, we could also fix the atom and let the derivative operator work on the surface position.

In the case $D = 4$, we have

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{5\hbar c \alpha_0}{64\pi^2\epsilon_0 z^4} = \frac{3\hbar c \alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{5}{6}\right) \qquad \text{(wrong!)}, \tag{22.255}$$

which is the standard result for the TM part of the total potential [see Eq. (14.214) and the following discussion], *except* for an overall sign, as this potential should be negative (attractive). Note that in 1D electromagnetism ($D = 2$), Eq. (22.254) becomes

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} \qquad \text{(wrong!)}, \tag{22.256}$$

which *should be* identical to the TE result (22.224): in 1D, all waves are normally incident, so the distinction between TE and TM waves is irrelevant. However, we are again off by an overall sign.

What happened? The flaw is in dropping the $\log\chi$ term in the potential compared to the $\log^2\chi$ term, which is invalid even in the limit $\chi \longrightarrow \infty$. It is the *balance* of the two terms in the potential that is critical in obtaining a sensible answer (though we will see in the next section that one of these terms can be dropped for *small* χ). We will see this minus sign showing up later in Section 22.8.6.1 when we treat the TM potential properly in the limit of large χ.

22.4.4.1 Fluctuations and the Interface Potential

Now that we see how the strong-coupling limit works here (or *doesn't* work, as it were), we can wax a bit philosophical. It was the first (δ^2) term in the potential (22.250) that produced the strong-coupling limit (albeit with the wrong sign), and evidently the second (δ') term is the one that produces deviations from it. We will see in the next calculation that for *small* χ, the first (δ^2) term is ignorable, and the second (δ') term yields the correct result. In intermediate regimes, it is the *interplay* of these two terms that conspire to give the correct result. As we have seen, it is even true in the strong-coupling regime that this interplay must hold, since the two terms must conspire to produce an overall minus sign, compared to what we would expect for only the (δ^2) term.

In terms of this interplay, one can intuitively think of a path touching an interface with some intermediate χ. The second term will always attempt to make the potential large and positive, trying to bring about the strong-coupling result. The first term can swing either way as the path crosses the surface, so most of the time the strong-coupling result will indeed happen. Recall that in the unrenormalized path integral, the strong-coupling result is simply a null integrand. However, it may happen that as the path crosses the surface, the total potential will swing large and negative. In this case, rather than having a null contribution, the integrand can be huge. The relative rarity of these events tempers their large contribution, and the

resulting average, while incorporating large fluctuations, converges to the correct intermediate result. Note that the $\mathcal{T}$ integral rescales the paths, so even for a particular path, this outer integral already averages over all these possibilities as different parts of the path are in contact with the interface.

22.4.5 TM Polarization: Weak Coupling

In the case of weak coupling, we can linearizing and then regularize the path integral (although technically the regularization should come *before* the linearization). In this case, we again have the potential

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{2}\left[\left(\nabla \log\sqrt{\epsilon_{\text{r}}}\right)^2 - \nabla^2 \log\sqrt{\epsilon_{\text{r}}}\right]. \tag{22.257}$$

with permittivity

$$\epsilon_{\text{r}}(z) = 1 + \chi\Theta(z), \tag{22.258}$$

so that

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{2}\left[\left(\log\sqrt{1+\chi}\right)^2 \delta^2(z) - \left(\log\sqrt{1+\chi}\right)\delta'(z)\right], \tag{22.259}$$

but now in the limit of small χ. Expanding to first order in χ, we can discard the first term and obtain

$$V_{\text{TM}}(\mathbf{r}) = -\frac{\chi}{4}\delta'(z) = -\frac{\chi}{4}\partial_z^2\Theta(z). \tag{22.260}$$

Note that by rewriting the derivative of the delta function in terms of the Heaviside function, we are already implementing our regularization strategy for the divergent potential (which is the only way in which an expansion like this could *possibly* make sense). Now in the path integral (22.247),

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2}\nabla^2\right)\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.261}$$

we can insert the potential and permittivity, and expand to first order in χ. The result is

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c\alpha_0\chi}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(-\frac{3}{2} + \frac{\mathcal{T}}{2}\partial_z^2 - \frac{\mathcal{T}^2}{8}\partial_z^4\right)\langle\Theta(z)\rangle_{\mathbf{x}(\tau)}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.262}$$

after renormalization against vacuum. Identifying the path average as the sojourn time for $\mathbf{x}(\tau)$ starting a distance z away from the surface, or equivalently a standard Brownian bridge sojourning a distance $z/\sqrt{\mathcal{T}}$ from the surface [see Eqs. (22.231) and the associated discussion], we have

$$\begin{aligned} V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= -\frac{\hbar c\alpha_0\chi}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\frac{3}{2} - \frac{\mathcal{T}}{2}\partial_z^2 + \frac{\mathcal{T}^2}{8}\partial_z^4\right) T_{\text{s}}\left[B(t); \frac{z}{\sqrt{\mathcal{T}}}\right]\right\rangle\!\!\right\rangle_{B(t)} \\ &= -\frac{\hbar c\alpha_0\chi}{4(4\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\frac{3}{2} - \mathcal{T}\partial_z^2 + \frac{\mathcal{T}^2}{2}\partial_z^4\right) T_{\text{s}}\left[B(t); \frac{z}{\sqrt{2\mathcal{T}}}\right]\right\rangle\!\!\right\rangle_{B(t)}. \end{aligned} \tag{22.263}$$

Here, we have integrals to evaluate of the form

$$I(D,\alpha;z) = \partial_z^\alpha \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle T_{\text{s}}\left[B(t); \frac{z}{\sqrt{2\mathcal{T}}}\right]\right\rangle\!\!\right\rangle_{B(t)}. \tag{22.264}$$

We have already computed the integral here, back in the solution to Eq. (22.234), so

$$I(D,\alpha;z) = \partial_z^\alpha \frac{\Gamma(D/2)}{(D+1)\, z^D} = (-1)^\alpha \frac{\Gamma(D/2)\Gamma(D+\alpha)}{2(D+1)\Gamma(D)\, z^{D+\alpha}}. \tag{22.265}$$

Note here that one of the cases we need is $I(0,4,z) = 12/z^2$, which must be done carefully by dimensional regularization, as we will see. For example, if the differentiation is performed naïvely here, by setting $D = 0$ too soon, for example, a spurious null result would obtain. The correct (dimensionally regularized) procedure is to do the differentiation *before* fixing D, and then taking the limit $D \longrightarrow 0$.

Now writing Eq. (22.262) in terms of this integral as

$$\begin{aligned} V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= -\frac{\hbar c \alpha_0 \chi}{4(4\pi)^{D/2}\epsilon_0}\left(\frac{3}{2}I(D,0;z) - I(D-2,2;z) + \frac{1}{2}I(D-4,4;z)\right) \\ &= -\frac{\hbar c \alpha_0 \chi}{4(4\pi)^{D/2}\epsilon_0 z^D}\left(\frac{3\Gamma(D/2)}{4(D+1)} - \frac{\Gamma[(D-2)/2]\,\Gamma(D)}{2(D-1)\Gamma(D-2)} + \frac{\Gamma[(D-4)/2]\,\Gamma(D)}{4(D-3)\Gamma(D-4)}\right). \end{aligned} \tag{22.266}$$

Simplifying the last factor, the result is

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = -\frac{\hbar c \alpha_0 \chi}{4(4\pi)^{D/2}\epsilon_0 z^D}\left(\frac{[4D(D-1)-5]\Gamma(D/2)}{4(D+1)}\right).$$

(Casimir–Polder potential, TM weak coupling) (22.267)

For $D = 4$, the last factor is $43/20$, and we have

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0 \chi}{32\pi^2\epsilon_0 z^4}\left(\frac{43}{120}\right).$$

(Casimir–Polder potential, TM weak coupling) (22.268)

This is the standard result in the small-χ regime for the TM part of the total potential [see Eq. (14.214) and the following discussion]. For $D = 2$, the last factor is $1/4$, and so

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = -\frac{\hbar c \alpha_0 \chi}{64\pi\epsilon_0 z^2}. \tag{22.269}$$

This is equivalent to the TE result (22.237), as it should be in 1D electromagnetism, where the distinction between the polarizations is lost.

Evidently, in obtaining results that are not quite correct, we have played fast and loose with the divergent potential. Although it should in principle be possible to regularize the potential, in the limit where a sharp interface is a good idealization, the behavior of the potential path average is a violent one, swinging between zero and a large positive value. What we are seeing here is that in a perturbative treatment that misses this behavior, the predictions aren't correct.

22.5 Fresnel Reflection Coefficients

In applying the method that follows, we will repeatedly encounter expressions involving the Fresnel reflection coefficients at a dielectric interface, but in a somewhat disguised form. Thus, we will have a short digression and develop some expressions that will greatly simplify algebra and interpretation in later calculations.

22.5.1 TE Reflection Coefficient: Vacuum–Dielectric Interface

The Fresnel reflection coefficient for TE polarization is given by[15]

$$r_{\text{TE}}(\theta_1) = \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2}, \tag{22.270}$$

where the incident angle is θ_1 from the medium of index n_1 into the medium of index n_2, and we have Snell's law,

$$n_1\sin\theta_1 = n_2\sin\theta_2, \tag{22.271}$$

which determines the angle θ_2 in the transmitting medium in terms of θ_1.

[15] Daniel A. Steck, *Classical and Modern Optics* (2006), Chapter 9, available online at `http://steck.us/teaching/`. Note that $r_{\text{TE}} \equiv r_{\text{S}}$.

22.5.1.1 Vacuum-Side Reflection (TE)

For a vacuum–dielectric interface (incident from the vacuum side), we can put $n_1 = 1$ and $n_2^2 = 1+\chi$, where χ is the susceptibility of the medium, to obtain

$$\begin{aligned} r_{\text{TE}}(\theta_1) &= \frac{\cos\theta_1 - \sqrt{1+\chi}\cos\theta_2}{\cos\theta_1 + \sqrt{1+\chi}\cos\theta_2} \\ &= \frac{\cos\theta_1 - \sqrt{(1+\chi)(1-\sin^2\theta_2)}}{\cos\theta_1 + \sqrt{(1+\chi)(1-\sin^2\theta_2)}} \\ &= \frac{\cos\theta_1 - \sqrt{1+\chi-\sin^2\theta_1}}{\cos\theta_1 + \sqrt{1+\chi-\sin^2\theta_1}} \\ &= \frac{\cos\theta_1 - \sqrt{\chi+\cos^2\theta_1}}{\cos\theta_1 + \sqrt{\chi+\cos^2\theta_1}}. \end{aligned} \tag{22.272}$$

Now suppose introduce the shorthand notation

$$\lambda = \cos^2\theta_1. \tag{22.273}$$

Then the reflection coefficient becomes

$$r_{\text{TE}}(\lambda;\chi) = \frac{\sqrt{\lambda} - \sqrt{\lambda+\chi}}{\sqrt{\lambda} + \sqrt{\lambda+\chi}}.$$

(TE reflection coefficient, vacuum side) (22.274)

Note that we are explicitly indicating the dependence on the medium parameter χ and the angle parameter λ.

22.5.1.2 Dielectric-Side Reflection (TE)

If a wave is incident from the *dielectric* side, the reflection coefficient takes on a somewhat different form. The general form (22.270) still applies, but we now take $n_2 = 1$ and $n_1^2 = 1+\chi$:

$$\begin{aligned} r_{\text{TE}}(\theta_1) &= \frac{\sqrt{1+\chi}\cos\theta_1 - \cos\theta_2}{\sqrt{1+\chi}\cos\theta_1 + \cos\theta_2} \\ &= \frac{\sqrt{1+\chi}\cos\theta_1 - \sqrt{1-\sin^2\theta_2}}{\sqrt{1+\chi}\cos\theta_1 + \sqrt{1-\sin^2\theta_2}} \\ &= \frac{\sqrt{1+\chi}\cos\theta_1 - \sqrt{1-(1+\chi)\sin^2\theta_1}}{\sqrt{1+\chi}\cos\theta_1 + \sqrt{1-(1+\chi)\sin^2\theta_1}} \\ &= \frac{\sqrt{1+\chi}\cos\theta_1 - \sqrt{(1+\chi)\cos^2\theta_1 - \chi}}{\sqrt{1+\chi}\cos\theta_1 + \sqrt{(1+\chi)\cos^2\theta_1 - \chi}}. \end{aligned} \tag{22.275}$$

Again identifying λ with $\cos^2\theta_1$, we can then write

$$r'_{\text{TE}}(\lambda;\chi) = \frac{\sqrt{\lambda(1+\chi)} - \sqrt{\lambda(1+\chi)-\chi}}{\sqrt{\lambda(1+\chi)} + \sqrt{\lambda(1+\chi)-\chi}},$$

(TE reflection coefficient, dielectric side) (22.276)

where the prime indicates that the sense of the reflection is from the dielectric side, compared to the vacuum-side coefficient in Eq. (22.274). Note that for the same incidence angles on each side, these two expressions should be equivalent up to a minus sign. However, the forms here are different because λ refers to *different* angles in each case.

22.5.2 TM Reflection Coefficient: Vacuum–Dielectric Interface

The reflection coefficient for TM polarization takes only a slightly different form: [16]

$$r_{\text{TM}}(\theta_1) = \frac{n_1 \cos\theta_2 - n_2 \cos\theta_1}{n_1 \cos\theta_2 + n_2 \cos\theta_1}. \tag{22.277}$$

Again, we will work out what this looks like if we replace $\cos^2\theta_1$ with λ, for a wave incident on either side.

22.5.2.1 Vacuum-Side Reflection (TM)

For incidence from the vacuum side we again put we can put $n_1 = 1$ and $n_2^2 = 1 + \chi$ to obtain

$$\begin{aligned} r_{\text{TM}}(\theta_1) &= \frac{\cos\theta_2 - \sqrt{1+\chi}\cos\theta_1}{\cos\theta_2 + \sqrt{1+\chi}\cos\theta_1} \\ &= \frac{\sqrt{1-\sin^2\theta_1} - \sqrt{1+\chi}\cos\theta_1}{\sqrt{1-\sin^2\theta_2} + \sqrt{1+\chi}\cos\theta_1} \\ &= \frac{\sqrt{1+\chi-\sin^2\theta_1} - (1+\chi)\cos\theta_1}{\sqrt{1+\chi-\sin^2\theta_1} + (1+\chi)\cos\theta_1} \\ &= \frac{\sqrt{\chi+\cos^2\theta_1} - (1+\chi)\cos\theta_1}{\sqrt{\chi+\cos^2\theta_1} + (1+\chi)\cos\theta_1}. \end{aligned} \tag{22.278}$$

With $\lambda = \cos^2\theta_1$, we can write

$$r_{\text{TM}}(\lambda;\chi) = \frac{\sqrt{\lambda+\chi} - \sqrt{\lambda}(1+\chi)}{\sqrt{\lambda+\chi} + \sqrt{\lambda}(1+\chi)},$$

(TM reflection coefficient, vacuum side) (22.279)

for a the reflection coefficient of a TM vacuum-incident wave.

22.5.2.2 Dielectric-Side Reflection (TM)

From the dielectric side, we take $n_2 = 1$ and $n_1^2 = 1 + \chi$, and obtain

$$\begin{aligned} r_{\text{TM}}(\theta_1) &= \frac{\sqrt{1+\chi}\cos\theta_2 - \cos\theta_1}{\sqrt{1+\chi}\cos\theta_2 + \cos\theta_1} \\ &= \frac{\sqrt{(1+\chi)(1-\sin^2\theta_2)} - \cos\theta_1}{\sqrt{(1+\chi)(1-\sin^2\theta_2)} + \cos\theta_1} \\ &= \frac{\sqrt{(1+\chi)(1-\sin^2\theta_2)} - \cos\theta_1}{\sqrt{(1+\chi)(1-\sin^2\theta_2)} + \cos\theta_1} \\ &= \frac{\sqrt{(1+\chi)[1-(1+\chi)\sin^2\theta_1]} - \cos\theta_1}{\sqrt{(1+\chi)[1-(1+\chi)\sin^2\theta_1]} + \cos\theta_1} \\ &= \frac{\sqrt{(1+\chi)[1-(1+\chi)(1-\cos^2\theta_1)]} - \cos\theta_1}{\sqrt{(1+\chi)[1-(1+\chi)(1-\cos^2\theta_1)]} + \cos\theta_1} \\ &= \frac{\sqrt{(1+\chi)[(1+\chi)\cos^2\theta_1 - \chi]} - \cos\theta_1}{\sqrt{(1+\chi)[(1+\chi)\cos^2\theta_1 - \chi]} + \cos\theta_1}. \end{aligned} \tag{22.280}$$

[16] Daniel A. Steck, *Classical and Modern Optics* (2006), Chapter 9, available online at `http://steck.us/teaching/`. Note that $r_{\text{TM}} \equiv r_{\text{P}}$ in the notation there. Note also that the convention here is that $r_{\text{TM}} = +1$ if the reflected wave is in phase with the incident wave, while the opposite-sign convention is common. In particular, this means that $r_{\text{TE}} = r_{\text{TM}}$ at normal incidence onto the same interface.

Thus, we can write

$$r'_{\text{TM}}(\lambda;\chi) = \frac{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}-\sqrt{\lambda}}{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}+\sqrt{\lambda}},$$

(TM reflection coefficient, vacuum side) (22.281)

where again the prime denotes incidence from the dielectric side.

22.6 Laplace–Mellin Method for Evaluating Path Integrals

In the solutions we have gotten so far, we have required probability densities for path statistics to evaluate the path integrals. In the TM case, the path integral is complicated by the presence of multiple statistics (path average of the potential *and* a function of the sojourn time), so in these cases we would need the *joint* density of path statistics, which are hard to come by in closed form. In the strong-coupling TM limit, we were able to reduce the problem to something manageable, so that we didn't need a complicated statistic. In the weak-coupling TM limit we were also able to simplify the path integral, but not without running into problems. All this motivates a more general method for evaluating path integrals that does *not* rely on specific probability densities for path statistics. We will begin by introducing some useful transforms and general integral relations.

22.6.1 Laplace and Mellin Transforms

For a function $f(t)$, recall that we can define the Laplace transform by

$$\mathscr{L}[f](s) \equiv F(s) := \int_0^\infty dt\, e^{-st} f(t). \qquad (22.282)$$

(Laplace transform)

We can also define the **Mellin transform** via

$$\mathscr{M}[f](z) \equiv \mathfrak{F}(z) := \int_0^\infty dt\, t^{z-1} f(t) = \int_0^\infty \frac{dt}{t^{1-z}}\, f(t). \qquad (22.283)$$

(Mellin transform)

These transforms obviously are only defined whenever the integrals converge. However, when they are sensible, the Mellin transform is invertible in the same sense as the Laplace transform, although it is somewhat more obscure.

Suppose that you have a Laplace transform of a function, but you want the Mellin transform. Obviously you could invert the Laplace transform and then stick the result into the Mellin integral. But what if the original function is not readily available? Here we will show that you can obtain the Mellin transform *directly* from the Laplace transform by computing a Mellin transform *of the Laplace transform itself*:[17]

$$\mathscr{M}[f](z) = \frac{1}{\Gamma(1-z)} \mathscr{M}\big[\mathscr{L}[f]\big](1-z).$$

(Laplace–Mellin conversion formula) (22.284)

In more direct notation, we can also write this as

$$\mathfrak{F}(z) = \frac{1}{\Gamma(1-z)} \int_0^\infty ds\, s^{-z} F(s). \qquad (22.285)$$

(Laplace–Mellin conversion formula)

Here,

$$\Gamma(z) = \int_0^\infty dt\, t^{z-1} e^{-t} = \mathscr{M}\big[e^{-t}\big](z) \qquad (22.286)$$

[17] J. S. Lew, "On Some Relations between the Laplace and Mellin transforms," *IBM Journal of Research and Development* **19**, 582 (1975) (doi: 10.1147/rd.196.0582).

is the usual gamma function.

To show this, we begin by writing out the Mellin transform of the Laplace transform $F(s)$ explicitly:

$$\mathscr{M}\big[\mathscr{L}[f]\big](z) = \mathscr{M}\big[F\big](z) = \int_0^\infty ds\, s^{z-1} F(s) = \int_0^\infty ds\, s^{z-1} \int_0^\infty dt\, e^{-st} f(t). \tag{22.287}$$

Then changing z to $1-z$,

$$\mathscr{M}\big[\mathscr{L}[f]\big](1-z) = \int_0^\infty ds\, s^{-z} \int_0^\infty dt\, e^{-st} f(t). \tag{22.288}$$

Changing the order of integration gives

$$\mathscr{M}\big[\mathscr{L}[f]\big](1-z) = \int_0^\infty dt\, f(t) \int_0^\infty ds\, s^{-z}\, e^{-st}, \tag{22.289}$$

and scaling $s \longrightarrow s/t$,

$$\mathscr{M}\big[\mathscr{L}[f]\big](1-z) = \int_0^\infty dt\, t^{z-1} f(t) \int_0^\infty ds\, s^{-z}\, e^{-s}. \tag{22.290}$$

Using

$$\Gamma(1-z) = \int_0^\infty ds\, s^{-z} e^{-s}, \tag{22.291}$$

and recognizing the remaining t integral as $\mathfrak{F}(z)$, we have

$$\mathscr{M}\big[\mathscr{L}[f]\big](1-z) = \mathscr{M}[f](z)\Gamma(1-z), \tag{22.292}$$

which is the result we wanted.

Note that we relied on a coordinate transformation that is ill-defined at the limits of integration. The integrations should really be taken between finite limits a and b, where $0 < a < b$, and then the limits $a \longrightarrow 0$ and $b \longrightarrow \infty$ should be taken at the end.

22.6.2 Laplace Transform and Inverse Moments

Recall that the Laplace transform of a probability density $f(x)$

$$\Big\langle\!\Big\langle e^{-sx} \Big\rangle\!\Big\rangle = \int_0^\infty ds\, e^{-sx} f(x) = \mathscr{L}[f](s) \tag{22.293}$$

acts as a moment-generating function, where the moments arise via differentiation:

$$\Big\langle\!\Big\langle x^n \Big\rangle\!\Big\rangle = (-1)^n\, \partial_s^n \Big\langle\!\Big\langle e^{-sx} \Big\rangle\!\Big\rangle\Big|_{s=0}. \tag{22.294}$$

It is also possible (and useful) to use the moment-generating function for **shifted, inverse moments**,[18]

$$\frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} e^{-s\beta} \Big\langle\!\Big\langle e^{-sx} \Big\rangle\!\Big\rangle = \Big\langle\!\Big\langle \frac{1}{(x+\beta)^\alpha} \Big\rangle\!\Big\rangle, \tag{22.295}$$

(shifted-inverse moment formula)

which holds for any α and β where the integrals make sense (α is not necessarily an integer).

To show this, first we start with Eq. (22.284), setting $z = 1-\alpha$:

$$\mathscr{M}[f](1-\alpha) = \frac{1}{\Gamma(\alpha)} \mathscr{M}\big[\mathscr{L}[f]\big](\alpha). \tag{22.296}$$

[18] For this and other related formulae, see Edward B. Rockower, "Integral Identities for Random Variables," *The American Statistician* **42**, 68 (1988) (doi: 10.1080/00031305.1988.10475526).

Writing out the Mellin-transform integrals from the definition (22.283),

$$\int_0^\infty dx\, x^{-\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} \mathscr{L}[f](s). \tag{22.297}$$

Then writing out the Laplace transform as an expectation value,

$$\int_0^\infty dx\, x^{-\alpha} f(x) = \frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} \left\langle\!\!\left\langle e^{-sx} \right\rangle\!\!\right\rangle. \tag{22.298}$$

Finally, changing variables via $x \longrightarrow x + \beta$ and identifying the left-hand side as an expectation value gives the desired result.

22.6.3 Feynman–Kac Formula: Review

The third ingredient we will need in this method is the Feynman–Kac formula (Section 17.11). Recall that the Feynman–Kac formula states that the distribution $f(x,t)$ of solutions to stochastic differential equations can be written both as the solution of a diffusion-type PDE and as an ensemble average over stochastic trajectories. In particular, we will be interested in the *steady-state* solution $f(x) \equiv f(x, t \longrightarrow \infty)$ for time-independent damping and driving functions, in which case the Feynman–Kac formula for exponentially stopped paths with stopping rate λ [from Eqs. (17.438) and (17.440)] is

$$f(x) = \int_0^\infty dt \left\langle\!\!\left\langle\, g[x + W(t)] \, \exp\left(-\lambda t - s \int_0^t dt'\, V[x + W(t')] \right) \right\rangle\!\!\right\rangle, \tag{22.299}$$

(Feynman–Kac formula)

where the function $f(x)$ also satisfies the ODE

$$f''(x) = 2\big[\lambda + sV(x)\big] f(x) - 2g(x). \tag{22.300}$$

(PDE for Feynman–Kac formula)

By choosing the functions $V(x)$ and $g(x)$ appropriately, we can derive various statistics about Wiener paths and Brownian bridges (see Section 17.12.1 and following sections). The basic idea is to obtain an analytic solution to the ODE (22.300). Then observing that the right-hand side of Eq. (22.299) has the form of a double Laplace transform in the variables λ and s, the idea is to invert the Laplace transforms, thereby obtaining the desired probability density for a statistic (the density here is hidden in the ensemble average). Choosing $V(x) = \Theta(x)$, for example, gives the sojourn-time statistic, and $V(x) = \delta(x)$ gives the local time. Choosing $g(x) = 1$ gives statistics for Wiener paths, while choosing $g(x) = \Theta(x)$ or e^{ikx} give access to statistics for Brownian bridges. Many variations on this theme are possible.[19]

22.6.4 Application to Worldline Path Integrals

The point of all this is to apply these pieces to evaluating worldline path integrals, schematically of the form

$$I = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle\, \langle\gamma\rangle^{-\alpha}_{\mathbf{B}_\mathcal{T}(\tau)}\, e^{-\mathcal{T}\langle \mathsf{V}\rangle_{\mathbf{B}_\mathcal{T}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{B}_\mathcal{T}(\tau)}, \tag{22.301}$$

where we have shifted away the center $\mathbf{x}_0$ of the paths, and $\gamma(\mathbf{x})$ and $\mathsf{V}(\mathbf{x})$ are some scalar functions of position. Then the general strategy is:

1. Design a related path integral of the form (22.299), where $\mathbf{V}$ must reflect both $s\gamma$ (where s is an auxiliary parameter) and V. Choose $g(x)$ to pin the path $W(t)$ to 0 at $\mathcal{T}$.

[19] A. N. Borodin and I. A. Ibragimov, *Limit Theorems for Functionals of Random Walks* (Proceedings of the Steklov Institute of Mathematics, vol. 195) (American Mathematical Society, 1995).

2. Solve the ODE (22.300) to obtain a solution to the path integral (22.299).
3. Use the Mellin-transform formula (22.284) to change the λ Laplace transform into the $\mathcal{T}$ integral.
4. Use the inverse-moment formula (22.295) (with $\beta = 0$) to convert the s Laplace transform to install the $\langle\gamma\rangle_{\mathbf{B}_{\mathcal{T}}(\tau)}^{-\alpha}$ factor.

The details vary with the details of the path integrals, so we will move on to solving some path integrals.

22.7 TE Casimir–Polder Path Integral: Dielectric Interface

As a first example, we will apply this method to the TE path integral for the Casimir–Polder potential in Eq. (22.238),

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2} - 1 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.302}$$

for a vacuum–dielectric interface.

22.7.1 Feynman–Kac Formula

Since the path integral involves the path average of ϵ_{r} in one dimension, the relevant statistic is the sojourn time at distance d:

$$T_{\text{s}}[y(t); d] := \int_0^t dt'\, \Theta[y(t') - d]. \tag{22.303}$$

We have already set up the Feynman–Kac formula to obtain the relevant path integral with exponential stopping. The form that we desire is Eq. (17.563)

$$\begin{aligned} L(\lambda, s; d) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\left(-sT_{\text{s}}[B_t; d]\right) \right\rangle\!\!\right\rangle &= \sqrt{\frac{\pi}{\lambda}} \left[1 - e^{-2\sqrt{2\lambda}\, d} \left(\frac{\sqrt{\lambda+s} - \sqrt{\lambda}}{\sqrt{\lambda+s} + \sqrt{\lambda}}\right)\right] \\ &= \sqrt{\frac{\pi}{\lambda}} \left[1 + r_{\text{TE}}(\lambda; s)\, e^{-2\sqrt{2\lambda}\, d}\right], \end{aligned}$$

(sojourn-time solution to Feynman–Kac formula) (22.304)

where we are using the TE reflection coefficient (22.274):

$$r_{\text{TE}}(\lambda; \chi) = \frac{\sqrt{\lambda} - \sqrt{\lambda+\chi}}{\sqrt{\lambda} + \sqrt{\lambda+\chi}}. \tag{22.305}$$

Again, Eq. (22.304) is equivalently a formula for the iterated Laplace transform of the sojourn time for Brownian bridges pinned to 0 at time t.

As a complete illustration of the general method for solving worldline path integrals, we will briefly review the derivation of this result. Beginning with Eq. (22.299), we will take $V(x) = s\Theta(x-d)$ to put the sojourn time in the exponential, and we will take $g(x) = e^{ikx}$ to (later) pin the Wiener path $W(t) = 0$ to select only Brownian bridges. Further, we will only write down the path integral at $x = 0$, which is the only point we will need:

$$\begin{aligned} f(0) &= \int_0^\infty dt \left\langle\!\!\left\langle\, e^{ikW(t)} \exp\left(-\lambda t - s\int_0^t dt'\, \Theta[W(t') - d]\right) \right\rangle\!\!\right\rangle \\ &= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle\, e^{ikW(t)} \exp\left(-sT_{\text{s}}[W(t); d]\right) \right\rangle\!\!\right\rangle. \end{aligned} \tag{22.306}$$

From Eq. (22.300), we can obtain an expression for $f(0)$ by solving the ODE

$$f''(x) = 2\big[\lambda + s\Theta(x-d)\big] f(x) - 2e^{ikx}. \tag{22.307}$$

The equation is easy to solve piecewise, and the result is [Eq. (17.556)]

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{2e^{ikx}}{2(\lambda+s)+k^2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{2e^{ikx}}{2\lambda+k^2} & (x < d). \end{cases} \tag{22.308}$$

Requiring continuity of the solution and its derivative yield the coefficients [Eq. (17.557)]

$$A = \frac{2\sqrt{2}\,s\left(\sqrt{2\lambda}-ik\right)e^{\sqrt{2(\lambda+s)}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}, \qquad B = -\frac{2\sqrt{2}\,s\left(\sqrt{2(\lambda+s)}+ik\right)e^{-\sqrt{2\lambda}\,d+ikd}}{(2\lambda+k^2)[2(\lambda+s)+k^2]\left(\sqrt{\lambda}+\sqrt{\lambda+s}\right)}, \tag{22.309}$$

completing the Feynman–Kac-formula solution.

Now to complete the pinning of the solution $W(t)$. Integrating the solution (22.306) over k introduces a delta function of the path W at the end-time t:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, f(0) = \int_0^\infty dt\, e^{-\lambda t}\left\langle\!\!\left\langle\, \delta[W(t)]\, \exp\left(-sT_{\rm s}[W(t);d]\right)\right\rangle\!\!\right\rangle. \tag{22.310}$$

Then using the path-integral relation [Eq. (17.561)]

$$\left\langle\!\!\left\langle \delta[W(t)]\, F[W(t')]\right\rangle\!\!\right\rangle = \frac{1}{\sqrt{2\pi t}}\left\langle\!\!\left\langle F[B_t(t')]\right\rangle\!\!\right\rangle, \tag{22.311}$$

where F is a functional and $B_t(t')$ is a Brownian bridge pinned to 0 at $t' = t$, we then have

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, f(0) = \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t}\left\langle\!\!\left\langle\, \exp\left(-sT_{\rm s}[B_t;d]\right)\right\rangle\!\!\right\rangle, \tag{22.312}$$

which is the integral form we desire in Eq. (22.302).

What remains is to evaluate the integral of the solution $f(0)$ to the diffusion ODE. If we assume $d > 0$, corresponding to the case of a Casimir–Polder potential for an atom *outside* the dielectric, then Eq. (22.308) becomes

$$f(0) = B + \frac{2}{2\lambda+k^2}, \tag{22.313}$$

and thus, with B as in Eq. (22.309), we obtain the solution [see Eq. (17.563)]

$$\begin{aligned} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, f(0) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\left(B + \frac{2}{2\lambda+k^2}\right) \\ &= \sqrt{\frac{\pi}{\lambda}}\left[1 - e^{-2\sqrt{2\lambda}\,d}\left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}}\right)\right]. \end{aligned} \tag{22.314}$$

This is the right-hand side of our desired result (22.302).

22.7.2 Mellin Transform

Now we can use the Mellin-Laplace-transform conversion formula (22.284) to compute the integral

$$I_D(s;d) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle e^{-sT_{\rm s}[B_{\mathcal{T}};d]} - 1\right\rangle\!\!\right\rangle, \tag{22.315}$$

where $B_{\mathcal{T}}(t)$ is a Brownian bridge pinned to 0 at time $\mathcal{T}$. First, we rewrite the integral in Eq. (22.304) as

$$L(\lambda,s;d) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1/2}}\, e^{-\lambda\mathcal{T}}\left\langle\!\!\left\langle \exp\left\{-sT_{\rm s}[B_{\mathcal{T}}]\right\}\right\rangle\!\!\right\rangle = \sqrt{\frac{\pi}{\lambda}}\left[1 + r_{\rm TE}(\lambda;s)\, e^{-2\sqrt{2\lambda}\,d}\right]. \tag{22.316}$$

We then adapt Eq. (22.284) with $1 - z = (1 + D/2) - 1/2 = (D + 1)/2$, or $z = (1 - D)/2$, to read

$$\begin{aligned} I_D(s;d) &= \frac{1}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{(D-1)/2} \Big[L(\lambda,s;d) - L(\lambda,0;d) \Big] \\ &= \frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{D/2-1} r_{\text{TE}}(\lambda;s)\, e^{-2\sqrt{2\lambda}\,d}. \end{aligned} \tag{22.317}$$

Note that we can also regard this Laplace–Mellin transformation as being an example of the integral formula

$$\int_0^\infty d\lambda\, \lambda^{(D-1)/2}\, e^{-\lambda\mathcal{T}} = \frac{\Gamma[(D+1)/2]}{\mathcal{T}^{(D+1)/2}}. \tag{22.318}$$

This transformation formally takes care of the first Laplace transform (in λ). Note that the path integral (22.315) that we computed here has the form of a potential-coupled scalar field, where the potential has the form of a step function in one direction.

However, this isn't *quite* the integral that we want for the dielectric interface. The reason is that the path integral (22.302) contains the path average of ϵ_{r}. If we take the the path terminus $\mathbf{r}$ to be the origin and the interface to be at distance z, then, considering 1D paths, the path average is

$$\langle \epsilon_{\text{r}} \rangle_{B_\mathcal{T}} = 1 + \chi \frac{T_{\text{s}}[B_\mathcal{T};z]}{\mathcal{T}} = \mathcal{T}^{-1} \Big(\mathcal{T} + \chi T_{\text{s}}[B_\mathcal{T};z] \Big) \tag{22.319}$$

in terms of the sojourn time $T_{\text{s}}[B_\mathcal{T};d]$ of Brownian bridges $B_\mathcal{T}$. Thus, we need the potential in our path integral here to involve this functional. Note that we can adapt Eq. (22.316) for this purpose by letting $s \longrightarrow s\chi$ and then $\lambda \longrightarrow \lambda + s$, so that

$$\begin{aligned} L(\lambda + s, s\chi; z) &= \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1/2}}\, e^{-\lambda\mathcal{T}} \left\langle\!\!\left\langle e^{-s(\mathcal{T} + \chi T_{\text{s}}[B_\mathcal{T}])} \right\rangle\!\!\right\rangle \\ &= \sqrt{\frac{\pi}{\lambda+s}} \left[1 + r_{\text{TE}}(\lambda + s; s\chi)\, e^{-2\sqrt{2(\lambda+s)}\,z} \right]. \end{aligned} \tag{22.320}$$

This allows us to compute the alternate integral, analogous to (22.315),

$$\tilde{I}_D(s,\chi;z) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle e^{-s(\mathcal{T} + \chi T_{\text{s}}[B_\mathcal{T};z])} - e^{-s\mathcal{T}} \right\rangle\!\!\right\rangle, \tag{22.321}$$

using the same procedure as for (22.317):

$$\begin{aligned} \tilde{I}_D(s,\chi;z) &= \frac{1}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{(D-1)/2} \Big[L(\lambda+s,s\chi;z) - L(\lambda+s,0;z) \Big] \\ &= \frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \frac{\lambda^{(D-1)/2}}{\sqrt{\lambda+s}}\, r_{\text{TE}}(\lambda+s;s\chi)\, e^{-2\sqrt{2(\lambda+s)}\,z}. \end{aligned} \tag{22.322}$$

The motivation for this form of the integral isn't completely clear at this point, but the main idea is that it contains the relevant functional for the problem.

22.7.3 Inverse Moments

Now going back to the original path integral (22.302), we see that for a dielectric interface we need to evaluate a path integral of the form

$$\mathcal{I}_{D,\alpha}(z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \frac{1}{(1 + \chi T_{\text{s}}[B_\mathcal{T};z]/\mathcal{T})^\alpha} - 1 \right\rangle\!\!\right\rangle, \tag{22.323}$$

where $\chi \geq 0$ is the dielectric susceptibility, z is the distance from the vacuum–dielectric interface to the atom (which is on the vacuum side).

The slightly more useful form of Eq. (22.323) is

$$\mathcal{I}_{D,\alpha}(z) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2-\alpha}} \left\langle\!\!\left\langle \frac{1}{(\mathcal{T} + \chi T_{\rm s}[B_{\mathcal{T}}; z])^\alpha} - \mathcal{T}^{-\alpha} \right\rangle\!\!\right\rangle, \tag{22.324}$$

which will allow us to more easily use the results of the previous section. Now using Eq. (22.295) with $\beta = 0$,

$$\frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} \left\langle\!\!\left\langle e^{-sx} \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle \frac{1}{x^\alpha} \right\rangle\!\!\right\rangle, \tag{22.325}$$

and comparing this to Eq. (22.321), we see that

$$\mathcal{I}_{D,\alpha}(z) = \frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} \tilde{I}_{D-2\alpha}(z). \tag{22.326}$$

Putting in the result (22.322) for $\tilde{I}_D(z)$,

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}}{\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]} \int_0^\infty ds\, s^{\alpha-1} \int_0^\infty d\lambda\, \frac{\lambda^{(D-1)/2-\alpha}}{\sqrt{\lambda+s}}\, r_{\rm TE}(\lambda+s; s\chi)\, e^{-2\sqrt{2(\lambda+s)}\, z}. \tag{22.327}$$

Now we just need to simplify this. Scaling out the distance dependence by letting $\lambda \longrightarrow \lambda/8z^2$ and $s \longrightarrow s/8z^2$,

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big] z^D} \int_0^\infty ds\, s^{\alpha-1} \int_0^\infty d\lambda\, \lambda^{(D-1)/2-\alpha} \frac{e^{-\sqrt{\lambda+s}}}{\sqrt{\lambda+s}}\, r_{\rm TE}(\lambda+s; s\chi), \tag{22.328}$$

where we have used the property of the reflection coefficient (22.274) that it is invariant if both parameters are scaled in the same way. Then letting $\lambda \longrightarrow \lambda - s$,

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big] z^D} \int_0^\infty ds\, s^{\alpha-1} \int_s^\infty d\lambda\, (\lambda-s)^{(D-1)/2-\alpha} \frac{e^{-\sqrt{\lambda}}}{\sqrt{\lambda}}\, r_{\rm TE}(\lambda; s\chi), \tag{22.329}$$

and then letting $\lambda \longrightarrow \lambda s$,

$$\begin{aligned} \mathcal{I}_{D,\alpha}(z) &= \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big] z^D} \int_0^\infty ds\, s^{D/2-1} \int_1^\infty d\lambda\, (\lambda-1)^{(D-1)/2-\alpha} \frac{e^{-\sqrt{\lambda s}}}{\sqrt{\lambda}}\, r_{\rm TE}(\lambda; \chi) \\ &= \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big] z^D} \int_1^\infty d\lambda\, \frac{(\lambda-1)^{(D-1)/2-\alpha}\, r_{\rm TE}(\lambda; \chi)}{\sqrt{\lambda}} \int_0^\infty ds\, s^{D/2-1}\, e^{-\sqrt{\lambda s}}, \end{aligned} \tag{22.330}$$

where we again used $r_{\rm TE}(s\lambda; s\chi) = r_{\rm TE}(\lambda; \chi)$. The integral over s has the value $2\Gamma(D)\lambda^{-D/2}$, so

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big] z^D} \int_1^\infty d\lambda\, (\lambda-1)^{(D-1)/2-\alpha} \lambda^{-(D+1)/2}\, r_{\rm TE}(\lambda; \chi). \tag{22.331}$$

This is as far as we can go without being more explicit about α and D.

22.7.4 Result: 3D Electromagnetism

Now comparing the original path integral (22.302) with the integral (22.323), we see that for the vacuum–dielectric interface we may write

$$V_{\rm CP}^{\rm (TE)}(z) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \mathcal{I}_{D,3/2}(z). \tag{22.332}$$

With the result (22.331), this becomes

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\Gamma(D)\hbar c\alpha_0}{2^{3D/2}(2\pi)^{D/2}\Gamma\big[D/2-1\big]\epsilon_0 z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r_{\text{TE}}(\lambda;\chi). \tag{22.333}$$

Now using the duplication formula for the gamma function[20]

$$\Gamma(2z) = (2\pi)^{-1/2}2^{2z-1/2}\Gamma(z)\Gamma(z+1/2), \tag{22.334}$$

and setting $z = D/2$, we have the useful result

$$\frac{\sqrt{\pi}\Gamma(D)}{\Gamma(D/2)} = 2^{D-1}\Gamma[(D+1)/2]. \tag{22.335}$$

Along with $z\Gamma(z) = \Gamma(z+1)$, we then have the marginally simpler form

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0 z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r_{\text{TE}}(\lambda;\chi),$$

(TE Casimir–Polder potential) (22.336)

where again

$$r_{\text{TE}}(\lambda;\chi) = \frac{\sqrt{\lambda}-\sqrt{\lambda+\chi}}{\sqrt{\lambda}+\sqrt{\lambda+\chi}}. \tag{22.337}$$

For $D = 4$, we have

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{3\hbar c\alpha_0}{128\pi^2\epsilon_0 z^4}\int_1^\infty d\lambda\,\lambda^{-5/2}\left(\frac{\sqrt{\lambda+\chi}-\sqrt{\lambda}}{\sqrt{\lambda+\chi}+\sqrt{\lambda}}\right), \tag{22.338}$$

and evaluating the remaining integral, we find

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{1}{6}+\frac{1}{\chi}-\frac{\sqrt{1+\chi}}{2\chi}-\frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\right).$$

(TE Casimir–Polder potential) (22.339)

This matches the earlier result (22.245) that we obtained using the explicit probability density for the sojourn time.

22.7.5 Result: 1D Electromagnetism

In the case of 1D electromagnetism ($D = 2$), the result (22.336) is a bit tricky, as the integrand has a pole at $\lambda = 1$, and the factor $\Gamma(D-2)$ vanishes at $D = 2$. Consequently, we will need to be careful, and we will handle the removable singularity here via dimensional regularization, letting $D \longrightarrow 2^+$ only after we remove the divergences. We begin by putting in $D = 2$ wherever it is not objectionable to do so:

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{(D-2)\hbar c\alpha_0}{32\pi\epsilon_0 z^2}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-3/2}\,r_{\text{TE}}(\lambda;\chi). \tag{22.340}$$

Integrating by parts, we have

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(z) = {}& \frac{(D-2)\hbar c\alpha_0}{\big(D/2-1\big)32\pi\epsilon_0 z^2}\Big[(\lambda-1)^{D/2-1}\lambda^{-3/2}\,r_{\text{TE}}(\lambda;\chi)\Big]_1^\infty \\ & - \frac{(D-2)\hbar c\alpha_0}{\big(D/2-1\big)32\pi\epsilon_0 z^2}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-1}\partial_\lambda\left[\lambda^{-3/2}\,r_{\text{TE}}(\lambda;\chi)\right]. \end{aligned} \tag{22.341}$$

[20] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 256, Eq. (6.1.18).

Since $D > 2$, the boundary terms vanish at both the lower limit (due to the $\lambda - 1$ factor) and upper limits (since the integrand, except for the D-dependent factor, decays asymptotically more quickly than $\lambda^{-3/2}$, and we are assuming D is close to 2). Also, canceling the $(D-2)$ terms,

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{\hbar c \alpha_0}{16\pi\Gamma\epsilon_0 z^2} \int_1^\infty d\lambda\, (\lambda-1)^{D/2-1} \partial_\lambda \left[\lambda^{-3/2}\, r_{\text{TE}}(\lambda;\chi)\right]. \tag{22.342}$$

Now the expression is regular at $D = 2$, so we will take $D \longrightarrow 2^+$:

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} \int_1^\infty d\lambda\, \partial_\lambda \left[\lambda^{-3/2}\, r_{\text{TE}}(\lambda;\chi)\right]. \tag{22.343}$$

The remaining integration is easy, and so

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} r_{\text{TE}}(1;\chi) = -\frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} \left(\frac{\sqrt{1+\chi}-1}{\sqrt{1+\chi}+1}\right), \tag{22.344}$$

(TE Casimir–Polder potential, $D = 2$)

since only the lower limit stays. Note that the reflection coefficient is the normal-incidence ($\lambda = 1$) coefficient, and that the potential is overall negative because of the sign of the coefficient. Also the reflection coefficinet here approaches -1 as $\chi \longrightarrow \infty$, so this result is consistent with our earlier strong-coupling calculation (22.224). For small χ, $r_{\text{TE}} \approx -\chi/4$, and so our result here is also consistent with the weak-coupling result (22.269).

To see that this result is consistent with the earlier calculations of Section 22.4.3, note that the analogue of Eq. (22.243) for $D = 2$ is

$$\begin{aligned}
&\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^2} \left\langle\!\!\left\langle 1 - (1+\chi T_{\text{s}})^{-3/2} \right\rangle\!\!\right\rangle \\
&\quad = \int_0^1 dx\, [1-(1+\chi x)^{-3/2}] \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^2} \left[\sqrt{\frac{4(1-x)}{\pi x \mathcal{T}}}\, e^{-1/(1-x)\mathcal{T}} + \left(1 - \frac{2}{\mathcal{T}}\right) e^{-1/\mathcal{T}}\, \text{erfc}\left(\sqrt{\frac{x}{(1-x)\mathcal{T}}}\right)\right] \\
&\quad = \int_0^1 dx\, [1-(1+\chi x)^{-3/2}] \left[\frac{1}{\sqrt{x}} - 1\right] \\
&\quad = \frac{\sqrt{1+\chi}-1}{\sqrt{1+\chi}+1}.
\end{aligned} \tag{22.345}$$

Combining this with Eq. (22.240) for $D = 2$ gives

$$\begin{aligned}
V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= -\frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^2} \left\langle\!\!\left\langle 1 - \left(1 + \chi T_{\text{s}}\left[B(t); \frac{1}{\sqrt{2\mathcal{T}}}\right]\right)^{-3/2} \right\rangle\!\!\right\rangle_{B(t)} \\
&= -\frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} \left(\frac{\sqrt{1+\chi}-1}{\sqrt{1+\chi}+1}\right).
\end{aligned} \tag{22.346}$$

Thus, we obtain the same result with that method.

22.8 Evaluation of the TM Interface Potential

Recall from Sections 22.4.4 and 22.4.5 that the TM potential in the case of a dielectric,

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{2}\left[\left(\nabla \log\sqrt{\epsilon_{\text{r}}}\right)^2 - \nabla^2 \log\sqrt{\epsilon_{\text{r}}}\right], \tag{22.347}$$

becomes singular and difficult to work with for a dielectric interface. Here, we will demonstrate how to work with this potential for an arbitrary dielectric susceptibility. To keep things simple, we will work with

a one-dimensional geometry and work with a simple model path integral that involves *only* this potential (and hence will have no obvious physical interpretation).

Going to one dimension, the potential (22.347) becomes

$$V_{\text{TM}}(z) = \frac{1}{2}\left[\left(\partial_z \log\sqrt{\epsilon_{\text{r}}(z)}\right)^2 - \partial_z^2 \log\sqrt{\epsilon_{\text{r}}(z)}\right]. \tag{22.348}$$

We want to consider the simple case of a planar interface, where

$$\epsilon_{\text{r}}(z) = 1 + \chi\Theta(z-d) \tag{22.349}$$

where the interface is at $z = d$, and separates vacuum ($z < d$) and a dielectric of susceptibility χ (for $z > d$). In this case the potential becomes

$$V_{\text{TM}}(z) = \frac{1}{2}\left[\left(\log\sqrt{1+\chi}\right)^2 \delta^2(z-d) - \left(\log\sqrt{1+\chi}\right)\delta'(z-d)\right], \tag{22.350}$$

which is, again, a rather singular potential to be stuck with.

22.8.1 Regularized TM Potential

As a regularized form of this potential, let us take the *logarithm* of the permittivity to ramp linearly over a transition length a,

$$\log\sqrt{\epsilon_{\text{r}}(z)} = \begin{cases} 0 & (z < d) \\ (z-d)\,\dfrac{\Xi}{a} & (d < z < d+a) \\ \Xi & (z > d+a), \end{cases} \tag{22.351}$$

where we have introduced the coupling parameter

$$\Xi := \log\sqrt{1+\chi}. \tag{22.352}$$

(dielectric coupling parameter for TM potential)

It is not yet obvious, but the explicit form of the function

$$\tanh\Xi = \frac{\chi}{2+\chi} \tag{22.353}$$

(path-integral coupling)

will be useful in what follows, as this is what will appear in the resulting path integrals.

In this case, the regularized potential (22.348) becomes

$$V_{\text{TM}}(z) = \frac{\Xi^2}{2a^2}\mathbb{1}_{[d,d+a]}(z) + \frac{\Xi}{2a}\Big[\delta[z-(d+a)] - \delta(z-d)\Big], \tag{22.354}$$

(regularized TM dielectric potential)

where $\mathbb{1}_{\mathcal{A}}(x)$ is the indicator function for the set $\mathcal{A}$ (i.e., the function is unity if $x \in \mathcal{A}$ and zero otherwise). Note that this potential "converges" to the singular potential (22.350) as $a \longrightarrow 0$. (The convergence here is in the sense of *any* delta-function limit: the limit should only be taken after appropriate integrations have taken place that do away with the singularities.) Thus, a path integral over Brownian bridges $B_t(t')$ involving this potential requires consideration of the **regularized interface functional**

$$\begin{aligned}\mathfrak{M}[B_t; d, \Xi, a] &:= \int_0^t dt'\, V_{\text{TM}}\big[B_t(t')\big] \\ &= \frac{\Xi}{2a}\int_0^t dt' \left\{\frac{\Xi}{a}\mathbb{1}_{[d,d+a]}\Big[B_t(t')\Big] + \delta\Big[B_t(t') - (d+a)\Big] - \delta\Big[B_t(t') - d\Big]\right\} \\ &= \frac{\Xi}{2a}\left\{\frac{\Xi}{a}T_{\text{s}}\Big(B_t; [d, d+a]\Big) + \ell\big(B_t; d+a\big) - \ell\big(B_t; d\big)\right\}.\end{aligned} \tag{22.355}$$

(regularized interface functional)

In the last expression here, we have replaced the integral of the three terms with the sojourn time $T_{\rm s}$ of B_t in the interval $[d, d+a]$

$$T_{\rm s}[y;\mathcal{A}] := \int_0^t dt'\, \mathbb{1}_{\mathcal{A}}[y(t')] \tag{22.356}$$

(this generalizes the sojourn time across a boundary, as in Section 17.12), and the local time of the process B_t at d and $d+a$ (Section 17.13):

$$\ell[y;a] := \int_0^t dt'\, \delta\big[y(t')-a\big]. \tag{22.357}$$

The regularized interface functional is, of course, interesting as a route to evaluating path integrals involving the interface functional, which we will define via

$$\begin{aligned}\mathfrak{N}[B_t;d,\Xi] &:= \lim_{a\to 0}\mathfrak{M}[B_t;d,\Xi,a]\\ &= \frac{\Xi}{2}\int_0^t dt'\left\{\Xi\,\delta^2\Big[B_t(t')-d\Big]-\delta'\Big[B_t(t')-d\Big]\right\}.\end{aligned} \tag{22.358}$$

(interface functional)

Again, any calculations for $\mathfrak{N}$ will need to involve $\mathfrak{M}$, only then taking the defined limit when possible without causing problems.

22.8.2 Model Path Integral

Now having defined the interface functional, we will evaluate the model path integral

$$L(\lambda;d,\Xi) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-\mathfrak{N}[B_t;d,\Xi]\Big)\right\rangle\!\!\right\rangle_{B_t} \tag{22.359}$$

to demonstrate the TM potential and gain some intuition. Of course, to do this, we will need to consider the regularized path integral

$$L_{\rm R}(\lambda,s;d,\Xi,a) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-s\mathfrak{M}[B_t;d,\Xi,a]\Big)\right\rangle\!\!\right\rangle_{B_t}, \tag{22.360}$$

and then take the limit $a\longrightarrow 0$. We have also introduced the parameter s to highlight the delicate interaction of the δ^2 and δ' parts of the interface functional; as it turns out, that the $a\longrightarrow 0$ limit will *only* lead to sensible results if $s=1$.

22.8.3 Feynman–Kac Formula

22.8.3.1 Setup

To work out the path integral (22.360), we start with the ODE part of the FK formula (22.300),

$$f''(x) = 2\big[\lambda + sV(x)\big]f(x) - 2g(x), \tag{22.361}$$

choosing the potential

$$V(x) = \sigma'\mathbb{1}_{[d,d+a]}(x) + \sigma\delta(x-d-a) - \sigma\delta(x-d), \tag{22.362}$$

where

$$\sigma := s\frac{\Xi}{2a}, \qquad \sigma' := s\frac{\Xi^2}{2a^2}. \tag{22.363}$$

In this case, we will also take $g(x) = e^{ikx}$ to pin $W(t)=0$. [For an alternate choice for $g(x)$, see Problem 22.13. For $x<d$ or $x>d+a$, the ODE is

$$f'' = 2\lambda f - 2e^{ikx}, \tag{22.364}$$

with general solutions

$$f(x) = \alpha_+ e^{\sqrt{2\lambda}\,x} + \alpha_- e^{-\sqrt{2\lambda}\,x} + \frac{e^{ikx}}{\lambda + k^2/2} \tag{22.365}$$

for coefficients $\alpha_\pm$ to be determined in each region. On the other hand, for $d < x < d + a$, the ODE is

$$f'' = 2(\lambda + \sigma')f - 2e^{ikx}, \tag{22.366}$$

$$f(x) = \alpha_+ e^{\sqrt{2(\lambda+\sigma')}\,x} + \alpha_- e^{-\sqrt{2(\lambda+\sigma')}\,x} + \frac{e^{ikx}}{\lambda + \sigma' + k^2/2} \tag{22.367}$$

for coefficients $\alpha_\pm$ to be determined. Picking the bounded solutions in each domain, we have the solution

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (x > d + a) \\ Be^{-\sqrt{2(\lambda+\sigma')}\,x} + Ce^{\sqrt{2(\lambda+\sigma')}\,x} + \dfrac{e^{ikx}}{\lambda + \sigma' + k^2/2} & (d < x < d + a) \\ De^{\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{22.368}$$

for undetermined constants A, B, C, and D.

22.8.3.2 Boundary Matching

Demanding continuity of $f(x)$ at $x = d + a$ and $x = d$ gives the conditions

$$\begin{aligned} Ae^{-\sqrt{2\lambda}(d+a)} + \frac{e^{ik(d+a)}}{\lambda + k^2/2} &= Be^{-\sqrt{2(\lambda+\sigma')}(d+a)} + Ce^{\sqrt{2(\lambda+\sigma')}(d+a)} \\ &\quad + \frac{e^{ik(d+a)}}{\lambda + \sigma' + k^2/2} \\ Be^{-\sqrt{2(\lambda+\sigma')}\,d} + Ce^{\sqrt{2(\lambda+\sigma')}\,d} + \frac{e^{ikd}}{\lambda + \sigma' + k^2/2} &= De^{\sqrt{2\lambda}\,d} + \frac{e^{ikd}}{\lambda + k^2/2}. \end{aligned} \tag{22.369}$$

The δ functions in the ODE say that the derivative $f'(x)$ should jump by $2\sigma f(d + a)$ at $x = d + a$ and $-2\sigma f(d)$ at $x = d$, so that

$$\begin{aligned} &-\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}(d+a)} + \frac{ike^{ik(d+a)}}{\lambda + k^2/2} \\ &\quad - 2\sigma\left(Ae^{-\sqrt{2\lambda}(d+a)} + \frac{e^{ik(d+a)}}{\lambda + k^2/2}\right) = -\sqrt{2(\lambda+\sigma')}Be^{-\sqrt{2(\lambda+\sigma')}(d+a)} \\ &\qquad + \sqrt{2(\lambda+\sigma')}Ce^{\sqrt{2(\lambda+\sigma')}(d+a)} + \frac{ike^{ik(d+a)}}{\lambda + \sigma' + k^2/2} \\ &-\sqrt{2(\lambda+\sigma')}Be^{-\sqrt{2(\lambda+\sigma')}\,d} \\ &\quad + \sqrt{2(\lambda+\sigma')}Ce^{\sqrt{2(\lambda+\sigma')}\,d} \\ &\qquad + \frac{ike^{ik(d+a)}}{\lambda + \sigma' + k^2/2} = \sqrt{2\lambda}De^{\sqrt{2\lambda}\,d} + \frac{ike^{ikd}}{\lambda + k^2/2} - 2\sigma\left(De^{\sqrt{2\lambda}\,d} + \frac{e^{ikd}}{\lambda + k^2/2}\right), \end{aligned} \tag{22.370}$$

or that is, $f'(d + a + 0^+) - 2\sigma f(d + a + 0^+) = f'(d + a - 0^+)$ and $f'(d + 0^+) = f'(d - 0^+) - 2\sigma f(d - 0^+)$. Since we are only in interested in $f(0)$, we only need solution of these four equations for the coefficient D. As it turns out, this expression is cumbersome, but it is somewhat less so after integration over k:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, D = -\frac{1}{\sqrt{2\lambda}}\left(\frac{\left(2\sigma^2 - \sigma' + 2\sqrt{2\lambda}\sigma\right)\sinh\left[a\sqrt{2(\lambda+\sigma')}\right]e^{-2\sqrt{2\lambda}\,d}}{(2\sigma^2 - \sigma' - 2\lambda)\sinh\left[a\sqrt{2(\lambda+\sigma')}\right] - 2\sqrt{\lambda(\lambda+\sigma')}\cosh\left[a\sqrt{2(\lambda+\sigma')}\right]}\right). \tag{22.371}$$

Then note that

$$f(0) = \frac{1}{\lambda + k^2/2} + D, \tag{22.372}$$

and since we want to calculate

$$L_{\text{R}}(\lambda, s; d, \Xi, a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, f(0) = \int_0^{\infty} \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- s\mathfrak{M}[B_t; d, \Xi, a]\Big) \right\rangle\!\!\right\rangle, \tag{22.373}$$

from the FK formula (22.299), we can use the integral formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, \frac{e^{-ikc}}{\lambda + k^2/2} = \sqrt{\frac{\pi}{\lambda}}\, e^{-\sqrt{2\lambda}|c|} \tag{22.374}$$

to find

$$L_{\text{R}}(\lambda, s; d, \Xi, a) = \sqrt{\frac{\pi}{\lambda}} \left(1 - \frac{\Big(2\sigma^2 - \sigma' + 2\sqrt{2\lambda}\sigma\Big) \sinh\Big[a\sqrt{2(\lambda + \sigma')}\Big]\, e^{-2\sqrt{2\lambda}\, d}}{(2\sigma^2 - \sigma' - 2\lambda) \sinh\Big[a\sqrt{2(\lambda + \sigma')}\Big] - 2\sqrt{\lambda(\lambda + \sigma')} \cosh\Big[a\sqrt{2(\lambda + \sigma')}\Big]} \right), \tag{22.375}$$

which is our basic result for the regularized model integral (22.360).

22.8.3.3 Overall Scaling of the Potential

Now we turn to the dependence of this result on the parameter s. Recall that we have used the notation (22.363)

$$\sigma := s\frac{\Xi}{2a}, \qquad \sigma' := s\frac{\Xi^2}{2a^2} \tag{22.376}$$

to render the expression a bit more compact. In the case of $s = 1$,

$$2\sigma^2 - \sigma' = s(s-1)\frac{\Xi^2}{2a^2} = 0, \tag{22.377}$$

and so

$$L_{\text{R}}(\lambda, s; d, \Xi, a) = \sqrt{\frac{\pi}{\lambda}} \left(1 + \frac{\sqrt{2\lambda}\sigma \sinh\Big[a\sqrt{2(\lambda + \sigma')}\Big]\, e^{-2\sqrt{2\lambda}\, d}}{\lambda \sinh\Big[a\sqrt{2(\lambda + \sigma')}\Big] + \sqrt{\lambda(\lambda + \sigma')} \cosh\Big[a\sqrt{2(\lambda + \sigma')}\Big]} \right). \tag{22.378}$$

This is important in analyzing the subsequent limit $a \longrightarrow 0$, where the $2\sigma^2 - \sigma'$ terms in the numerator and denominator would dominate the other terms, leading to the Ξ-independent result

$$L(\lambda, s; d, \Xi, a) = \sqrt{\frac{\pi}{\lambda}} \left(1 - e^{-2\sqrt{2\lambda}\, d}\right), \tag{22.379}$$

This result is characteristic of a ''strong-coupling'' result. For example this is equivalent to the path-integral result (22.316) from the TE case, in the limit $s \longrightarrow \infty$. This shows that $s = 1$ is a special choice, inducing the right balance between local and occupation times to produce a nontrivial limit. This also shows that V_{TM} (alternately, $\mathfrak{N}$) is *not* a potential in the usual sense (for example, doubling it does not double the ''energy''); rather its function is to enforce a boundary condition at the interface, which has an arbitrarily large ''cost'' in terms of the energy.

22.8.3.4 Sharp-Interface Limit

Then in the limit $a \longrightarrow 0$,

$$a\sqrt{2(\lambda + \sigma')} = \sqrt{(2a^2\lambda + s\Xi^2)} \longrightarrow \sqrt{s}\Xi = \Xi \tag{22.380}$$

for $s = 1$, and so the path-integral result for $\mathfrak{M}$ (22.378) goes over to the path integral for $\mathfrak{N}$,

$$L(\lambda, s; d, \Xi) = \lim_{a \longrightarrow 0} L_{\mathrm{R}}(\lambda, s; d, \Xi, a) = \sqrt{\frac{\pi}{\lambda}} \left(1 + \tanh(\Xi)\, e^{-2\sqrt{2\lambda}\, d}\right), \tag{22.381}$$

(model TM path-integral result)

which is a nontrivial result. As $\Xi \longrightarrow 0$, this reduces to

$$L(\lambda, s; d, \Xi, a) \longrightarrow \sqrt{\frac{\pi}{\lambda}}, \tag{22.382}$$

which is reasonable, as it is the same as the $d \longrightarrow \infty$ decoupling limit. For small Ξ, the result is

$$L(\lambda) = \sqrt{\frac{\pi}{\lambda}} \left(1 + \Xi\, e^{-2\sqrt{2\lambda}\, d}\right). \tag{22.383}$$

On the other hand, as $\Xi \longrightarrow \infty$, this reduces to

$$L(\lambda) = \sqrt{\frac{\pi}{\lambda}} \left(1 + e^{-2\sqrt{2\lambda}\, d}\right), \tag{22.384}$$

which is also a reasonable result, in the sense of avoiding pathologies and depending sensibly on d.

22.8.4 Weak-Coupling Expansion

As a check for the small-Ξ limit, we can keep only the first-order terms in Ξ in the potential (22.358),

$$\begin{aligned}
\mathfrak{N}[B_t; d, \Xi] &= \lim_{a \to 0} \mathfrak{M}[B_t; d, \Xi, a] \\
&= \lim_{a \to 0} \frac{\Xi}{2a} \int_0^t dt' \left\{ \delta\Big[B_t(t') - (d + a)\Big] - \delta\Big[B_t(t') - d\Big] \right\} \\
&= -\frac{\Xi}{2} \int_0^t dt'\, \delta'\Big[B_t(t') - d\Big] \\
&= -\partial_d^2 \frac{\Xi}{2} \int_0^t dt'\, \Theta\Big[B_t(t') - d\Big].
\end{aligned} \tag{22.385}$$

In this regime, the path integral (22.359) becomes

$$\begin{aligned}
L(\lambda; d, \Xi) &= \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- \mathfrak{N}[B_t; d, \Xi] \Big) \right\rangle\!\!\right\rangle_{B_t} \\
&= \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle 1 - \mathfrak{N}[B_t; d, \Xi] \right\rangle\!\!\right\rangle_{B_t} \\
&= \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle 1 + \partial_d^2 \frac{\Xi}{2} \int_0^t dt'\, \Theta\Big[B_t(t') - d\Big] \right\rangle\!\!\right\rangle_{B_t} \\
&= \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left[1 + \partial_d^2 \frac{\Xi}{2} \Big\langle\!\!\Big\langle T_{\mathrm{s}}[B_t; d] \Big\rangle\!\!\Big\rangle \right],
\end{aligned} \tag{22.386}$$

where we have written this now as an integral over the first moment of the sojourn time. We can use the expression (17.588), generalized to pinning at time t,

$$\Big\langle\!\!\Big\langle T_{\mathrm{s}}[B_t, d] \Big\rangle\!\!\Big\rangle = t \left(\frac{e^{-2d^2/t}}{2} - \sqrt{\frac{\pi d^2}{2t}}\, \mathrm{erfc}\left[\sqrt{\frac{2d^2}{t}} \right] \right), \tag{22.387}$$

and then evaluate the integral, with the result

$$\begin{aligned}
L(\lambda; d, \Xi) &= \sqrt{\frac{\pi}{\lambda}} \left(1 + \frac{\Xi}{2}\, \partial_d^2 \frac{e^{-2\sqrt{2\lambda}\, d}}{4\lambda} \right) \\
&= \sqrt{\frac{\pi}{\lambda}} \left(1 + \Xi\, e^{-2\sqrt{2\lambda}\, d} \right).
\end{aligned} \tag{22.388}$$

This agrees with the earlier small-Ξ calculation (22.383). Note that there are no issues with obtaining the wrong sign in the expanded integral here.

22.8.5 Transfer-Layer Formalism

Now that we see that the $a \longrightarrow 0$ limit is well-defined (at least for $s = 1$), it makes sense to simplify the derivation of Section 22.8.3 to take this limit as early as possible, and to thereby do it "once and for all." First, we will consider the homogeneous case $g(x) = 0$, and modify this later to include the pinning delta function. Then for $x < d$ or $x > d + a$ we are considering the homogeneous ODE

$$f'' = 2\lambda f, \tag{22.389}$$

and thus referring back to the function $f(x)$ defined in Eq. (22.368), we should instead work with the solution

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} & (x > d + a) \\ Be^{-\sqrt{2(\lambda+\sigma')}\,x} + Ce^{\sqrt{2(\lambda+\sigma')}\,x} & (d < x < d + a) \\ De^{\sqrt{2\lambda}\,x} & (x < d). \end{cases} \tag{22.390}$$

We can then rewrite the boundary conditions (22.369) demanding continuity of $f(x)$ at $x = d+a$ and $x = d$ as

$$\begin{aligned} f(d + a + 0^+) &= Be^{-\sqrt{2(\lambda+\sigma')}(d+a)} + Ce^{\sqrt{2(\lambda+\sigma')}(d+a)} \\ Be^{-\sqrt{2(\lambda+\sigma')}\,d} + Ce^{\sqrt{2(\lambda+\sigma')}\,d} &= f(d - 0^+), \end{aligned} \tag{22.391}$$

and we can rewrite the boundary conditions (22.370) for $f'(x)$ as

$$\begin{aligned} f'(d + a + 0^+) - 2\sigma f(d + a + 0^+) &= -\sqrt{2(\lambda + \sigma')}Be^{-\sqrt{2(\lambda+\sigma')}(d+a)} \\ &\quad + \sqrt{2(\lambda + \sigma')}Ce^{\sqrt{2(\lambda+\sigma')}(d+a)} \\ -\sqrt{2(\lambda + \sigma')}Be^{-\sqrt{2(\lambda+\sigma')}\,d} + \sqrt{2(\lambda + \sigma')}Ce^{\sqrt{2(\lambda+\sigma')}\,d} &= f'(d - 0^+) - 2\sigma f(d - 0^+), \end{aligned} \tag{22.392}$$

or that is, $f'(d+a+0^+) - 2\sigma f(d+a+0^+) = f'(d+a-0^+)$ and $f'(d+0^+) = f'(d-0^+) - 2\sigma f(d-0^+)$. Then solving these four equations for $f(d - 0^+)$ and $f'(d - 0^+)$, while eliminating B and C, then substituting in

$$\sigma := s\frac{\Xi}{2a}, \qquad \sigma' := s\frac{\Xi^2}{2a^2}, \tag{22.393}$$

then setting $s = 1$ and taking the limit $a \longrightarrow 0$, we find the remarkably simple result that the interface functional defines a discontinuity in the solution $f(x)$ according to

$$\begin{aligned} f(d + 0^+) &= e^{-\Xi} f(d - 0^+) \\ f'(d + 0^+) &= e^{\Xi} f'(d - 0^+). \end{aligned}$$

(TM boundary condition in FK solution) (22.394)

These boundary conditions replace the usual continuity conditions at the interface.

For example, suppose we apply this boundary condition to

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} & (x > d) \\ Be^{-\sqrt{2\lambda}\,x} + Ce^{\sqrt{2\lambda}\,x} & (0 < x < d) \\ De^{\sqrt{2\lambda}\,x} & (x < d). \end{cases} \tag{22.395}$$

which is an alternative to the choice (22.368), but we are applying the new boundary condition at $x = d$, and the boundary condition of a delta function of amplitude -2 at $x = 0$, since we are now choosing $g(x) = \delta(x)$ to pin $W(t) = 0$. Then the boundary conditions (22.394) at $x = d$ read

$$\begin{aligned} Ae^{-\sqrt{2\lambda}\,d}e^{\Xi} &= Be^{-\sqrt{2\lambda}\,d} + Ce^{\sqrt{2\lambda}\,d} \\ -\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d}e^{-\Xi} &= -\sqrt{2\lambda}Be^{-\sqrt{2\lambda}\,d} + \sqrt{2\lambda}Ce^{\sqrt{2\lambda}\,d}, \end{aligned} \tag{22.396}$$

and the boundary conditions at $x = 0$ read

$$\begin{aligned} D &= B + C \\ \sqrt{2\lambda}D - 2 &= -\sqrt{2\lambda}B + \sqrt{2\lambda}C. \end{aligned} \tag{22.397}$$

Solving for D gives

$$D = \frac{1}{\sqrt{2\lambda}}\left(1 + \tanh(\Xi)\, e^{-2\sqrt{2\lambda}\,d}\right), \tag{22.398}$$

and since

$$L(\lambda) = \sqrt{2\pi}f(0) = \sqrt{2\pi}D = \sqrt{\frac{\pi}{\lambda}}\left(1 + \tanh(\Xi)\, e^{-2\sqrt{2\lambda}\,d}\right), \tag{22.399}$$

we have recovered the result (22.381), but now by directly applying the interface boundary conditions (22.394).

Note that we can also apply these boundary conditions in the $g(x) = e^{ikx}$ calculation. We set this up with the solution function

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{22.400}$$

which is basically the solution (22.368), but with no "middle" region $d < x < d + a$. The analysis of the thin boundary layer still applies, since the inhomogeneous part of the solution is ignorable in that region. The two boundary conditions (22.394) at $x = d$ then read

$$\begin{aligned} Ae^{-\sqrt{2\lambda}\,d} + \frac{e^{ikd}}{\lambda + k^2/2} &= \left(Be^{\sqrt{2\lambda}\,d} + \frac{e^{ikd}}{\lambda + k^2/2}\right)e^{-\Xi} \\ -\sqrt{2\lambda}Ae^{-\sqrt{2\lambda}\,d} + \frac{ike^{ikd}}{\lambda + k^2/2} &= \left(\sqrt{2\lambda}Be^{\sqrt{2\lambda}\,d} + \frac{ike^{ikd}}{\lambda + k^2/2}\right)e^{\Xi}, \end{aligned} \tag{22.401}$$

and solving for B gives

$$B = \frac{\left(\sqrt{2\lambda} - ike^{\Xi}\right)\left(e^{\Xi} - 1\right)e^{ikd - \sqrt{2\lambda}\,d}}{(\lambda + k^2/2)\sqrt{2\lambda}\,(e^{2\Xi} + 1)}. \tag{22.402}$$

Then using

$$L(\lambda, s; d, \Xi) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, f(0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, \left(\frac{1}{\lambda + k^2/2} + B\right), \tag{22.403}$$

we again recover the result (22.381), but with considerably less effort.

22.8.6 Mellin Transform

To continue an example with the TM interface potential, we will compute the path integral

$$I_D(z, \Xi) = \int_0^{\infty} \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle e^{-\Re[B_{\mathcal{T}}; z, \Xi]} - 1 \right\rangle\!\!\right\rangle_{B_{\mathcal{T}}}. \tag{22.404}$$

This has the form of the TM path integral (22.247), but simplified to only the exponential factor, as a ''warm up'' to doing the full path integral.

Here we will follow the same method as in Section 22.7.2. Beginning with Eq. (22.381),

$$L(\lambda, s; z, \Xi) = \sqrt{\frac{\pi}{\lambda}} \left(1 + \tanh(\Xi)\, e^{-2\sqrt{2\lambda}\, z}\right), \tag{22.405}$$

we then use Eq. (22.284) in the form

$$\begin{aligned} I_D(z) &= \frac{1}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{(D-1)/2} \Big[L(\lambda, s; z) - L(\lambda, 0; z)\Big] \\ &= \frac{\sqrt{\pi}\tanh\Xi}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{D/2-1} e^{-2\sqrt{2\lambda}\, z}, \end{aligned} \tag{22.406}$$

or evaluating the last integral, we have

$$I_D(z) = \frac{\sqrt{\pi}\Gamma(D)\tanh\Xi}{2^{3D/2-1}\Gamma\big[(D+1)/2\big]\, z^D}. \tag{22.407}$$

Using Eq. (22.335) in the form

$$\frac{\sqrt{\pi}\Gamma(D)}{\Gamma[(D+1)/2]} = 2^{D-1}\Gamma(D/2). \tag{22.408}$$

we have

$$I_D(z) = \frac{\Gamma(D/2)\tanh\Xi}{2^{D/2}\, z^D} = \frac{\Gamma(D/2)}{2^{D/2}\, z^D}\left(\frac{\chi}{2+\chi}\right), \tag{22.409}$$

(model TM path integral)

as the result for our model TM path integral, where we have used Eq. (22.353) to connect the result back to the susceptibility χ.

22.8.6.1 Comparison to Strong-Coupling

It is useful to compare this result with the strong-coupling limit of physical path integrals. For example, we can take the TE strong-coupling result (22.222), and remove the prefactor in the expression (22.219) to obtain the analogous result

$$\begin{aligned} I_D^{(\mathrm{TE})}(\mathbf{r}) &= \lim_{\chi\to\infty} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\mathrm{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} - 1 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= -\frac{\Gamma(D/2)}{2^{D/2} z^D}. \end{aligned} \tag{22.410}$$

Note that this is a *generic* result, occurring whenever there is a strong effect on a path touching the dielectric interface. For example, the sojourn-time path integral (22.315),

$$I_D(z) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle e^{-sT_{\mathrm{s}}[B_{\mathcal{T}}; z]} - 1 \right\rangle\!\!\right\rangle, \tag{22.411}$$

has the form of a Casimir–Polder path integral for a massless scalar field coupled to a potential step, where s governs the ''strength'' of the potential. The result (22.317) for this path integral was

$$I_D(z) = -\frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{D/2-1} e^{-2\sqrt{2\lambda}\, z} \left(\frac{\sqrt{\lambda+s} - \sqrt{\lambda}}{\sqrt{\lambda+s} + \sqrt{\lambda}}\right). \tag{22.412}$$

In the limit $s \longrightarrow \infty$, this becomes

$$\begin{aligned} I_D(z) &= -\frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{D/2-1} e^{-2\sqrt{2\lambda}\, z} \\ &= -\frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma\big[(D+1)/2\big] z^D}, \end{aligned} \tag{22.413}$$

which after applying Eq. (22.408) is equivalent to Eq. (22.410).

Thus, we will compare the TM model path integral (22.409) to the strong-coupling result (22.410). The factor $(-\tanh\Xi)$ appearing in (22.409) is an ''efficiency factor'' multiplying the expected strong-coupling limit, and note that the sign is *opposite* to what one might expect for an efficiency. This is because while most of the potential is positive, there is a negative component, which when exponentiated can spike to large, positive values, thus yielding a positive result even after subtracting away the $z \longrightarrow \infty$ part (normally, a negative potential keeps the exponential less than one, and when subtracting one, the result is strictly negative).

In the limit $\Xi \longrightarrow \infty$ (equivalently, $\chi \longrightarrow \infty$), this integral is equivalent to the strong-coupling result, except for the overall minus sign. This explains the sign discrepancy that we found in Section 22.4.4 when we tried to cheat to obtain the strong-coupling result by considering only the δ^2 part of the potential. There, we obtained the ''usual'' strong-coupling result, which is wrong by an overall sign compared to the ''correct'' result we are getting here.

22.8.7 Generalization for Open Bridges

A useful generalization of the path integral in Eqs. (22.359) and (22.381) is to change the path from ''loop'' Brownian bridges $B_t(t')$ to bridges $B_{t(0\to c)}(t')$ pinned at 0 and some other point c. In particular, our goal will be to compute the interface path integral

$$\mathcal{N}(d,\Xi,t,c) := \left\langle\!\!\left\langle e^{-\mathfrak{N}[B_{t(0\to c)};d,\Xi]} \right\rangle\!\!\right\rangle_{B_{t(0\to c)}}, \tag{22.414}$$
(interface path integral)

which is useful in numerically evaluating path integrals with the interface potential.

22.8.7.1 Feynman–Kac Formula

To start, we will derive an expression for the path integral

$$L(\lambda;d,\Xi,c) := \int_0^\infty dt\, e^{-\lambda t}\, \frac{e^{-c^2/2t}}{\sqrt{t}} \left\langle\!\!\left\langle \exp\Big(-\mathfrak{N}[B_{t(0\to c)};d,\Xi]\Big) \right\rangle\!\!\right\rangle_{B_{t(0\to c)}}, \tag{22.415}$$

where again $B_{t(0\to c)}(t')$ is a Wiener process (Brownian bridge) pinned to c at time $t' = t$ [i.e., $B_t(t) = c$], and the interface functional $\mathfrak{N}$ here is defined as

$$\begin{aligned}\mathfrak{N}[B_t;d,\Xi,c] &:= \lim_{a\to 0}\frac{\Xi}{2a}\int_0^t dt' \left\{\frac{\Xi}{a}\mathbb{1}_{[d,d+a]}\Big[B_{t(0\to c)}(t')\Big] + \delta\Big[B_{t(0\to c)}(t') - (d+a)\Big] - \delta\Big[B_{t(0\to c)}(t') - d\Big]\right\} \\ &= \lim_{a\to 0}\frac{\Xi}{2a}\left\{\frac{\Xi}{a}T_{\rm s}\Big(B_{t(0\to c)};[d,d+a]\Big) + \ell\big(B_{t(0\to c)};d+a\big) - \ell\big(B_{t(0\to c)};d\big)\right\} \\ &\qquad\qquad (d>0,\, \Xi\geq 0),\end{aligned} \tag{22.416}$$

in analogy to the path case in Eqs. (22.355) and (22.358). The path integral here generalizes the integral (22.359) for path bridges.

In solving this problem, we will employ the transfer-layer formalism in the limit $a \longrightarrow 0$, embodied by the boundary conditions (22.394), and generalize the derivation of the $c = 0$ case, starting with the solution function (22.400)

$$f(x) = \begin{cases} Ae^{-\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x > d) \\ Be^{\sqrt{2\lambda}\,x} + \dfrac{e^{ik(x-c)}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{22.417}$$

where we are taking $g(x) = e^{ik(x-c)}$ in the Feynman–Kac formula, and we are excluding the ''middle'' region $d < x < d + a$. The two boundary conditions (22.394) at $x = d$ then read

$$\begin{aligned} Ae^{-\sqrt{2\lambda}\,d} + \frac{e^{ik(d-c)}}{\lambda + k^2/2} &= \left(Be^{\sqrt{2\lambda}\,d} + \frac{e^{ik(d-c)}}{\lambda + k^2/2} \right) e^{-\Xi} \\ -\sqrt{2\lambda} Ae^{-\sqrt{2\lambda}\,d} + \frac{ike^{ik(d-c)}}{\lambda + k^2/2} &= \left(\sqrt{2\lambda} Be^{\sqrt{2\lambda}\,d} + \frac{ike^{ik(d-c)}}{\lambda + k^2/2} \right) e^{\Xi}, \end{aligned} \tag{22.418}$$

and solving for B gives

$$B = \frac{\left(\sqrt{2\lambda} - ike^{\Xi}\right)\left(e^{\Xi} - 1\right) e^{ik(d-c) - \sqrt{2\lambda}\,d}}{(\lambda + k^2/2)\sqrt{2\lambda}\,(e^{2\Xi} + 1)}. \tag{22.419}$$

Then using

$$L(\lambda; d, \Xi, c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, f(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \left(\frac{e^{-ikc}}{\lambda + k^2/2} + B \right), \tag{22.420}$$

we obtain

$$L(\lambda; d, \Xi, c) = \sqrt{\frac{\pi}{\lambda}} \left(e^{-\sqrt{2\lambda}\,|c|} + \frac{\left(1 - e^{-\Xi}\right)\left[\mathrm{sgn}(d-c) + e^{-\Xi}\right]}{1 + e^{-2\Xi}}\, e^{-\sqrt{2\lambda}(d + |c-d|)} \right). \tag{22.421}$$

Throughout, we have assumed $d > 0$, so we can also write this as

$$L(\lambda; d, \Xi, c) = \sqrt{\frac{\pi}{\lambda}} \left(e^{-\sqrt{2\lambda}\,|c|} + \frac{\left(1 - e^{-\Xi}\right)\left[\mathrm{sgn}(d-c) + e^{-\Xi}\right]}{1 + e^{-2\Xi}}\, e^{-\sqrt{2\lambda}(|d| + |c-d|)} \right). \tag{22.422}$$

The absolute value we introduced here is important in maintaining the choice of bounded solution when we generalize to $d < 0$. An equivalent result also arises if we make the simultaneous replacement $d \longrightarrow -d$, $c \longrightarrow -c$, and $\Xi \longrightarrow -\Xi$, given the reflection symmetry of the problem (which reverses the ''direction of crossing'' through the boundary, hence the change in Ξ),

$$L(\lambda; -d, -\Xi, -c) = \sqrt{\frac{\pi}{\lambda}} \left(e^{-\sqrt{2\lambda}\,|c|} + \frac{\left(1 - e^{\Xi}\right)\left[-\mathrm{sgn}(d-c) + e^{\Xi}\right]}{1 + e^{2\Xi}}\, e^{-\sqrt{2\lambda}(|d| + |c-d|)} \right), \tag{22.423}$$

so in general, we have

$$\begin{aligned} L(\lambda; d, \Xi, c) &= \sqrt{\frac{\pi}{\lambda}} \left(e^{-\sqrt{2\lambda}\,|c|} + \frac{\left(1 - e^{-\mathrm{sgn}(d)\Xi}\right)\left[\mathrm{sgn}(d)\,\mathrm{sgn}(d-c) + e^{-\mathrm{sgn}(d)\Xi}\right]}{1 + e^{-\mathrm{sgn}(d)2\Xi}}\, e^{-\sqrt{2\lambda}(|d| + |c-d|)} \right) \\ &= \sqrt{\frac{\pi}{\lambda}} \left(e^{-\sqrt{2\lambda}\,|c|} + \frac{\sinh(\Xi/2)\left[\mathrm{sgn}(d-c)\, e^{\mathrm{sgn}(d)\Xi/2} + \mathrm{sgn}(d)\, e^{-\mathrm{sgn}(d)\Xi/2}\right]}{\cosh(\Xi)}\, e^{-\sqrt{2\lambda}(|d| + |c-d|)} \right), \end{aligned} \tag{22.424}$$

which is now valid for any c and d (positive or negative)

22.8.7.2 Inversion of the Laplace Transform

Now to compute the path integral (22.414), we must invert the Laplace transform in Eq. (22.415). Using the inverse Laplace transform

$$\mathscr{L}^{-1}\left[\sqrt{\frac{\pi}{\lambda}}\, e^{-\sqrt{2\lambda}\,\alpha} \right](t) = \frac{1}{\sqrt{t}} e^{-\alpha^2/2t}, \tag{22.425}$$

we have

$$\mathscr{L}^{-1}\Big[L(\lambda; d, \Xi, c) \Big](t) = \frac{e^{-c^2/2t}}{\sqrt{t}} + \left(\frac{\sinh(\Xi/2)\left[\mathrm{sgn}(d-c)\, e^{\mathrm{sgn}(d)\Xi/2} + \mathrm{sgn}(d)\, e^{-\mathrm{sgn}(d)\Xi/2}\right]}{\cosh(\Xi)} \right) \frac{e^{-(|d| + |c-d|)^2/2t}}{\sqrt{t}}. \tag{22.426}$$

Equating this with the inverse transform of Eq. (22.415),

$$\mathscr{L}^{-1}\Big[L(\lambda;d,\Xi,c)\Big](t) = \frac{e^{-c^2/2t}}{\sqrt{t}}\Big\langle\Big\langle \exp\Big(-\mathfrak{N}[B_{t(0\to c)};d,\Xi]\Big)\Big\rangle\Big\rangle_{B_{t(0\to c)}}, \tag{22.427}$$

we then find

$$\mathcal{N}(d,\Xi,t,c) = 1 + \frac{\sinh(\Xi/2)\left[\mathrm{sgn}(d-c)\,e^{\mathrm{sgn}(d)\Xi/2} + \mathrm{sgn}(d)\,e^{-\mathrm{sgn}(d)\Xi/2}\right]}{\cosh(\Xi)}\, e^{\left[c^2-(|d|+|c-d|)^2\right]/2t}$$

(interface path integral) (22.428)

as our desired result for the path integral defined in Eq. (22.414). Note that the exponential factor in the second term is unity if d is between 0 and c (i.e., if the path endpoints straddle the boundary), and is the boundary-crossing probability $\exp[2d(c-d)/t] = \exp[-2|d(d-c)|/t]$ otherwise (irrespective of the signs of c and d). The Ξ-dependent factor reduces to $\mathrm{sech}\,(\Xi)-1$ again if d is between 0 and c (i.e., a boundary-crossing path), in which case $\mathcal{N}(d,\Xi,t,c) = \mathrm{sech}\,(\Xi)$; this factor is $\mathrm{sgn}(d)\tanh(\Xi)$ otherwise. Thus, the second term is restricted to the range $(-1,1)$, and so the functional (22.428) is strictly positive, as it should be, from the definition (22.414).

The expression for $\mathcal{N}(d,\Xi,t,c)$ has the interpretation of representing probabilities for reflection and transmission at the interface. We can see this by letting $d \longrightarrow 0$, so that the crossing probability reduces to unity, being careful to maintain the sign of d. In this case, we may write

$$\mathcal{N}(0^-,\Xi,t,c) = \begin{cases} \mathfrak{t}_{\mathcal{N}}(\Xi) & (c>0) \\ 1+\mathfrak{r}_{\mathcal{N}}(\Xi) & (c<0) \end{cases}$$
$$\mathcal{N}(0^+,\Xi,t,c) = \begin{cases} 1-\mathfrak{r}_{\mathcal{N}}(\Xi) & (c>0) \\ \mathfrak{t}_{\mathcal{N}}(\Xi) & (c<0), \end{cases}$$

(interface path integral, starting at interface) (22.429)

where the effective reflection and transmission coefficients are

$$\mathfrak{r}_{\mathcal{N}}(\Xi) = \tanh\Xi = \frac{\chi}{2+\chi}$$
$$\mathfrak{t}_{\mathcal{N}}(\Xi) = \mathrm{sech}\,\Xi = \frac{2\sqrt{1+\chi}}{2+\chi}.$$

(reflection/transmission coefficients for interface path integral) (22.430)

The interpretation here is as follows. Suppose we follow a (Wiener) stochastic path until it just bumps into the interface. The probability density for the path position x after a time t later is then given by the Gaussian measure $\exp[-(x-d)^2/2t]/\sqrt{2\pi t}$, multiplied by the appropriate factor in Eq. (22.429) for each side of the interface. An enhancement of probability on one side compared to the other is exactly what we expect from a reflection/transmission process at an interface. However, since $\mathrm{sech}^2 x + \tanh^2 x = 1$, it is *not* the case that $\mathfrak{r}_{\mathcal{N}} + \mathfrak{t}_{\mathcal{N}} = 1$; in fact, $\mathfrak{r}_{\mathcal{N}} + \mathfrak{t}_{\mathcal{N}} \geq 1$, with equality only achieved for $\chi = 0$ or in the limit $\chi \longrightarrow \infty$. Rather, the coefficients satisfy the ''probability-conservation relation'' $\mathfrak{r}_{\mathcal{N}}^2 + \mathfrak{t}_{\mathcal{N}}^2 = 1$. Thus, each ''collision'' of the path with the interface does not conserve the path amplitude. Of course, we do not necessarily require this, since the path integral must only reproduce some correct value on average. However, it is interesting to note that for a closed Brownian bridge, which must have an equal number of crossings in either direction, and thus transmission factors contribute only in powers of $\mathfrak{t}_{\mathcal{N}}^2$. On the other hand, evidently any power of $\mathfrak{r}_{\mathcal{N}}$ may arise. In the limit where $\chi \longrightarrow \infty$, we have $\mathfrak{r}_{\mathcal{N}} \longrightarrow 1$ and $\mathfrak{t}_{\mathcal{N}} \longrightarrow 0$, indicating perfect reflection (and amplitude conservation on the left-hand side of the interface). Note that for finite χ, in evaluating the TM potential, there are also losses on the right-hand side of the interface associated with the $\langle\chi\rangle^{-\alpha}$ factor in the Casimir path integral that compensate for the extra probability generated on each reflection.

22.8.8 Assembling Closed Bridges

The path integral (22.414), which takes on the value given in Eq. (22.428), is an average of an exponential functional over a Brownian bridge pinned from 0 to c in time t. Thus, if we split the Brownian bridge into two temporal "segments" $(0, t/2)$ and $(t/2, t)$, then note that the exponential factors, and by construction we should be able to recover the original result by summing over all possible values of the Brownian bridge at time $t/2$, with the appropriate weight. Specifically, we are saying that by construction $\mathcal{N}(d, \Xi, t, c)$ satisfies

$$\mathcal{N}(d, \Xi, t, 0) = \int_{-\infty}^{\infty} dc \mathcal{N}(d, \Xi, t/2, c)\, \mathcal{N}(d-c, \Xi, t/2, -c)\, f_{\rm G}(c; 0, t/4), \tag{22.431}$$

where

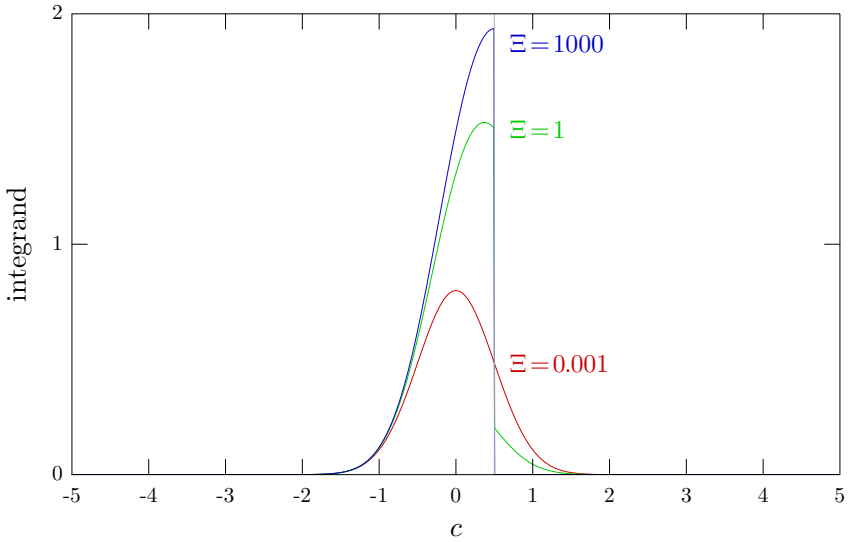
$$f_{\rm G}(x; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2}. \tag{22.432}$$

is the normal (Gaussian) probability density. Recall that from our analysis of finite-bridge generation [Eq. (17.310)], we can generate an $(N+1)$-point *standard* Brownian bridge (i.e., pinned to zero at $t = 1$) of points $B_0, \ldots B_N$, where $B_0 = B_N = 0$, by using the forward recurrence

$$\begin{aligned} B_0 &= B_N = 0 \\ B_n &= z_n \sqrt{\frac{N-n}{N(N-n+1)}} + \left(\frac{N-n}{N-n+1}\right) B_{n-1}. \end{aligned} \tag{22.433}$$

where z_n (from $n = 1, \ldots, N-1$) are standard normal random numbers (zero mean, unit variance). The variance $t/4$ for the density of c follows from the distribution of B_1 for the case $N = 2$.

As a numerical example, the plot below shows the integrand of Eq. (22.431) for three values of Ξ, with $d = 0.5$ and $t = 1$.



Note the discontinuity at $c = d$ in the integrand, where the source point of the Brownian bridge just touches the "dielectric surface." The values of the integral in the three cases are 1.00061 ($\Xi = 0.001$), 1.46193 ($\Xi = 1$), and 1.60653 ($\Xi = 1000$). The main feature to note in the plot is that for small Ξ, the integrand is well-approximated by the Gaussian factor, with increasingly large deviations above and below the Gaussian as Ξ increases. The impact of this observation is as follows. Suppose we evaluate the integral in a Monte-Carlo fashion, by choosing random values of c according to the Gaussian factor $f_{\rm G}(c; 0, t/4)$, and averaging the values taken on by the rest of the integrand for each chosen c. This would be fairly simple at small Ξ, but at large Ξ, this introduces relatively large fluctuations in the integrand value, and thus slowed convergence.

Of course, the slow convergence gets dramatically worse if we put in many intermediate integrals, developing a path integral. For example, the generalization of (22.431) for two intermediate points is

$$\mathcal{N}(d, \Xi, t, 0) = \int dc_1\, dc_2\, \mathcal{N}(d, \Xi, t/3, c_1)\, \mathcal{N}(d-c_1, \Xi, t/3, c_2-c_1)\, \mathcal{N}(d-c_2, \Xi, t/3, -c_2)\, f_1(c_1) f_2(c_2), \tag{22.434}$$

where from Eqs. (22.433), we have

$$f_n(c_n) = f_{\rm G}(c_n; \xi_n c_{n-1}, \xi_n \Delta t), \tag{22.435}$$

since the c_n corresponds to a particular value for B_n, and in this case $N = 3$. We are also here using the usual notations

$$\Delta t := \frac{t}{N}, \qquad \xi_n := \frac{N-n}{N-n+1}. \tag{22.436}$$

Note the recursive nature of the definition here, such that successive integrations have the form of convolutions.

For the general case of $N-1$ intermediate points, this generalizes to

$$\mathcal{N}(d,\Xi,t,0) = \int dc_1 \cdots dc_{N-1}\, \mathcal{N}(d-c_{N-1},\Xi,\Delta t, c_N - c_{N-1}) \prod_{j=1}^{N-1} \mathcal{N}(d-c_{j-1},\Xi,\Delta t, c_j - c_{j-1})\, f_j(c_j),$$

(path integral for TM potential) (22.437)

where $c_0 = c_N = 0$. Again, whenever Ξ is large and there is significant ''overlap'' of the integral with the TM boundary at d, the $\mathcal{N}$ factor will have large variation. For large N, there will be many such factors, leading to large fluctuations of the path-integral samples (with c-paths), and thus poor convergence, becoming increasingly worse for large N. Of course, this is an artifact of choosing the product of all $f_j(c_j)$ as the path-integral measure, when really we should absorb the fluctuations into the path measure to the greatest extent we can. In particular, we should define the ''tempered'' path-step distributions

$$\bar{f}_j(c_j) := \eta_j \mathcal{N}(d-c_{j-1},\Xi,\Delta t, c_j - c_{j-1})\, f_j(c_j), \tag{22.438}$$

where we have introduced the normalization factors η_j, defined by

$$\begin{aligned}
\eta_j^{-1} &:= \int_{-\infty}^{\infty} dc_j\, \mathcal{N}(d-c_{j-1},\Xi,\Delta t, c_j - c_{j-1})\, f_j(c_j) \\
&= \int_{-\infty}^{\infty} dc_j\, \mathcal{N}(d-c_{j-1},\Xi,\Delta t, c_j - c_{j-1})\, f_{\rm G}(c_j; \xi_j c_{j-1}, \xi_j \Delta t) \\
&= \int_{-\infty}^{\infty} dx\, \mathcal{N}(d-c_{j-1},\Xi,\Delta t, x)\, f_{\rm G}[x; (\xi_j - 1)c_{j-1}, \xi_j \Delta t].
\end{aligned} \tag{22.439}$$

This has the analytic solution

$$\begin{aligned}
\eta_j^{-1} \;=\; & \frac{\cosh^2(\Xi/2)}{\cosh\Xi} + \mathrm{sgn}(d-c_{j-1})\frac{\sinh^2(\Xi/2)}{\cosh\Xi}\,\mathrm{erf}\left(\frac{d-\xi_j c_{j-1}}{\sqrt{2\xi_j\Delta t}}\right) \\
& + \frac{\tanh\Xi}{2}\left[\mathrm{sgn}(d-c_{j-1}) - \mathrm{erf}\left(\frac{(2\xi_j-1)d - \xi_j c_{j-1}}{\sqrt{2\xi_j\Delta t}}\right)\right] e^{2(\xi_j-1)d(d-c_{j-1})/\Delta t} \\
\xi_j :=\; & \frac{N-j}{N-j+1},
\end{aligned} \tag{22.440}$$

at least for $c_{j-1} \neq d$, since this factor has a discontinuity at $c_{j-1} = d$. The precise value $c_{j-1} = d$ doesn't really matter for integration, but it is sensible to define the function at this point to have the mean of the values for $c_{j-1} = d + 0^{\pm}$. Note again that all these are defined recursively in terms of the previous point c_{j-1} of the path. Thus, we have the tempered form of the path integral (22.437):

$$\mathcal{N}(d,\Xi,t,0) = \int_{-\infty}^{\infty} dc_1 \cdots dc_{N-1}\, \mathcal{N}(d-c_{N-1},\Xi,\Delta t, c_N - c_{N-1}) \prod_{j=1}^{N-1} \eta_j^{-1} \bar{f}_j(c_j). \tag{22.441}$$

Then taking the product of the $\bar{f}_j(c_j)$ as the path measure, the tempered path integral becomes

$$\mathcal{N}(d,\Xi,t,0) = \left\langle\!\!\left\langle \mathcal{N}(d-\bar{c}_{N-1},\Xi,\Delta t, \bar{c}_N - \bar{c}_{N-1}) \prod_{j=1}^{N-1} \eta_j^{-1} \right\rangle\!\!\right\rangle_{\bar{c}(t')} \tag{22.442}$$

in Monte-Carlo form, which behaves much better than the Gaussian path integral (as the η_j factors are essentially smoothed versions of the $\mathcal{N}$ distributions), but at the expense of more complicated paths. Note in this last expression that we have switched the path to the $\bar{c}$ notation to emphasize that this path is generated by the $\bar{f}$ path measure.

22.9 TM Casimir–Polder Path Integral: Dielectric Interface

To apply the method for handling the interface functional in the previous section, we will evaluate the TM path integral (22.247)

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2}\nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.443}$$

for a planar, vacuum–dielectric interface. The derivation here will parallel the TE treatment in Section 22.7, but with additions from the TM-potential treatment in Section 22.8.

22.9.1 Feynman–Kac Formula

We will begin by deriving an expression for the path integral

$$L(\lambda, s; d, \Xi) := \int_0^\infty \frac{dt}{\sqrt{t}} \, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(-\mathfrak{N}[B_t; d, \Xi] - s T_{\text{s}}[B_t; d] \Big) \right\rangle\!\!\right\rangle_{B_t}, \tag{22.444}$$

where $B_t(t')$ is a Wiener process (Brownian bridge) pinned to 0 at time $t' = t$ [i.e., $B_t(t) = 0$], T_{s} is the sojourn time, and the functional $\mathfrak{N}$ is as defined in Eq. (22.358). This generalizes the corresponding TE expression (22.304) by introducing the interface functional at the boundary.

In working out this integral, we will employ the transfer-layer formalism of Section 22.8.5 (in the sharp-interface limit $a \longrightarrow 0$), embodied by the boundary conditions (22.394). We will thus generalize the derivation of the $s = 0$ case from Eq. (22.359) to (22.381). We start with the generalization of the solution function (22.400),

$$f(x) = \begin{cases} A e^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{e^{ikx}}{\lambda + s + k^2/2} & (x > d) \\ B e^{\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (x < d), \end{cases} \tag{22.445}$$

which modifies the denominator in the $x > d$ region to account for the sojourn time. Again, we are taking $g(x) = e^{ikx}$ in the Feynman–Kac formula to change the Wiener path into a closed bridge. This is the same as the setup for the sojourn-time calculation [cf. Eq. (17.556)], except that we will apply the modified boundary conditions (22.394) instead of simple continuity of the function and the derivative.

The two boundary conditions (22.394) at $x = d$ then read

$$\begin{aligned} A e^{-\sqrt{2(\lambda+s)}\,d} + \frac{e^{ikd}}{\lambda + s + k^2/2} &= \left(B e^{\sqrt{2\lambda}\,d} + \frac{e^{ikd}}{\lambda + k^2/2} \right) e^{-\Xi} \\ -\sqrt{2(\lambda+s)} A e^{-\sqrt{2(\lambda+s)}\,d} + \frac{ike^{ikd}}{\lambda + s + k^2/2} &= \left(\sqrt{2\lambda} B e^{\sqrt{2\lambda}\,d} + \frac{ike^{ikd}}{\lambda + k^2/2} \right) e^{\Xi}, \end{aligned} \tag{22.446}$$

and solving for B gives

$$B = e^{-\sqrt{2\lambda}\,d + ikd} \left(\frac{e^{\Xi}\left(\lambda + k^2/2\right)\left(\sqrt{\lambda+s} + ik/\sqrt{2}\right) - \left(\lambda + s + k^2/2\right)\left(\sqrt{\lambda+s} + ike^{2\Xi}/\sqrt{2}\right)}{\left(\lambda + k^2/2\right)\left(\lambda + s + k^2/2\right)\left(\sqrt{\lambda+s} + e^{2\Xi}\sqrt{\lambda}\right)} \right). \tag{22.447}$$

Then using

$$L(\lambda, s; d, \Xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \, f(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \, \left(\frac{1}{\lambda + k^2/2} + B \right), \tag{22.448}$$

we obtain

$$\begin{aligned} L(\lambda,s;d,\Xi) &= \sqrt{\frac{\pi}{\lambda}}\left[1-e^{-2\sqrt{2\lambda}\,d}\left(\frac{\sqrt{\lambda+s}-e^{2\Xi}\sqrt{\lambda}}{\sqrt{\lambda+s}+e^{2\Xi}\sqrt{\lambda}}\right)\right] \\ &= \sqrt{\frac{\pi}{\lambda}}\left[1+e^{-2\sqrt{2\lambda}\,d}\,\mathfrak{r}(\lambda;s,\Xi)\right], \end{aligned} \tag{22.449}$$

where we are using the shorthand

$$\mathfrak{r}(\lambda;s,\Xi) := \frac{e^{2\Xi}\sqrt{\lambda}-\sqrt{\lambda+s}}{e^{2\Xi}\sqrt{\lambda}+\sqrt{\lambda+s}}, \tag{22.450}$$

which will later become the usual TM reflection coefficient. Notice the similarity here of Eq. (22.449) to Eq. (22.304), to which it reduces as $\Xi \longrightarrow 0$.

22.9.2 Mellin Transform

Now we want to use the Mellin-Laplace-transform conversion formula (22.284) to compute the path integral

$$\tilde{I}_D(s,\Xi;z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle e^{-\mathfrak{N}[B_\mathcal{T};z,\Xi]-s(\mathcal{T}+\chi T_\mathrm{s}[B_\mathcal{T};z])} - e^{-s\mathcal{T}}\right\rangle\!\!\right\rangle, \tag{22.451}$$

in analogy to the TE integral (22.321). To do this, we rewrite (22.449) by replacing t with $\mathcal{T}$, let $s \longrightarrow s\chi$ and then let $\lambda \longrightarrow \lambda + s$:

$$\begin{aligned} L(\lambda+s,s\chi;d,\Xi) &= \int_0^\infty \frac{d\mathcal{T}}{\sqrt{\mathcal{T}}}\,e^{-\lambda\mathcal{T}}\left\langle\!\!\left\langle e^{-s(\mathcal{T}+\chi T_\mathrm{s}[B_\mathcal{T}])}\right\rangle\!\!\right\rangle \\ &= \sqrt{\frac{\pi}{\lambda+s}}\left[1+e^{-2\sqrt{2\lambda}\,d}\,\mathfrak{r}(\lambda+s;s\chi,\Xi)\right], \end{aligned} \tag{22.452}$$

as in Eq. (22.320). Then we can transform this result using Eq. (22.284),

$$\begin{aligned} \tilde{I}_D(s,\chi;z) &= \frac{1}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\lambda^{(D-1)/2}\Big[L(\lambda+s,s\chi;z)-L(\lambda+s,0;z)\Big] \\ &= \frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\frac{\lambda^{(D-1)/2}}{\sqrt{\lambda+s}}\,e^{-2\sqrt{2(\lambda+s)}\,z}\,\mathfrak{r}(\lambda+s;s\chi,\Xi), \end{aligned} \tag{22.453}$$

as in Eq. (22.322). Notice that our procedure here parallels exactly the TE calculation, except for the reflection coefficient, which has includes two factors of $e^{2\Xi}$ compared to $r_\mathrm{TE}(\lambda+s;s\chi)$. Recall that Ξ is determined completely by χ [Eq. (22.352)], so we won't need to note the dependence of the result on Ξ. (As far as the derivation was concerned, they *were* independent parameters, but we are now identifying them as being related.)

22.9.3 Inverse Moments

In converting the result (22.453) into the worldline path-integral result, the calculation here continues to closely parallel that of Section 22.7.3. We thus require the TM version of the TE integral (22.323), which is

$$\mathcal{I}_{D,\alpha}(z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \frac{e^{-\mathfrak{N}[B_\mathcal{T};z,\Xi]}}{(1+\chi T_\mathrm{s}[B_\mathcal{T};z]/\mathcal{T})^\alpha}-1\right\rangle\!\!\right\rangle. \tag{22.454}$$

To do this, we apply the inverse-moment formula (22.295) to Eq. (22.453). Again, the functional form of the integral here is essentially the same as what leads to the result (22.331), except for the (constant) factors of $\exp(2\Xi) = 1+\chi$. Thus, all the same algebra applies, with the modified result

$$\begin{aligned} \mathcal{I}_{D,\alpha}(z) &= \frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,(\lambda-1)^{(D-1)/2-\alpha}\lambda^{-(D+1)/2}\,\mathfrak{r}(\lambda;\chi,\Xi) \\ &= -\frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,(\lambda-1)^{(D-1)/2-\alpha}\lambda^{-(D+1)/2}\,r_\mathrm{TM}(\lambda;\chi), \end{aligned} \tag{22.455}$$

where we used

$$\mathfrak{r}(\lambda;\chi,\Xi) = \frac{\sqrt{\lambda}(1+\chi) - \sqrt{\lambda+\chi}}{\sqrt{\lambda}(1+\chi) + \sqrt{\lambda+\chi}} = -r_{\text{TM}}(\lambda;\chi), \tag{22.456}$$

with $r_{\text{TM}}(\lambda;\chi)$ being the usual Fresnel coefficient for TM polarization, as defined in Eq. (22.279). Again, to proceed, we will need to choose specific values of α and D, and to obtain Casimir–Polder energies, we will need to consider *combinations* of these integrals with different values of α and D.

22.9.4 Result: 3D Electromagnetism

Comparing the TM path integral (22.443) for the Casimir–Polder potential

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2}\nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.457}$$

to the integral definition (22.454), we see that we can write the (renormalized) TM potential as

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \left(\mathcal{I}_{D,3/2}(z) - \frac{1}{2}\partial_z^2 \mathcal{I}_{D-2,1/2}(z) \right) \tag{22.458}$$

for an atom at a distance z from a planar dielectric interface.

Thus, to begin, we will need

$$\begin{aligned} \mathcal{I}_{D,3/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-2}\Gamma(D/2-1)z^D} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D+1)/2} \, r_{\text{TM}}(\lambda;\chi) \\ &= -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{D/2-1}\sqrt{\pi} z^D} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D+1)/2} \, r_{\text{TM}}(\lambda;\chi), \end{aligned} \tag{22.459}$$

where we used the duplication formula in the form (22.335),

$$\frac{\sqrt{\pi}\Gamma(D)}{\Gamma(D/2)} = 2^{D-1}\Gamma[(D+1)/2], \tag{22.460}$$

which we can rewrite as

$$\frac{\Gamma(D)}{\Gamma(D/2-1)} = \frac{2^{D-1}(D/2-1)\Gamma[(D+1)/2]}{\sqrt{\pi}}. \tag{22.461}$$

Similarly, we will need

$$\mathcal{I}_{D-2,1/2}(z) = -\frac{\Gamma(D-2)}{2^{3D/2-4}\Gamma(D/2-1)z^{D-2}} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D-1)/2} \, r_{\text{TM}}(\lambda;\chi), \tag{22.462}$$

to obtain the differentiated form

$$\begin{aligned} \partial_z^2 \mathcal{I}_{D-2,1/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-4}\Gamma(D/2-1)z^D} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D-1)/2} \, r_{\text{TM}}(\lambda;\chi) \\ &= -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{D/2-3}\sqrt{\pi} z^D} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D-1)/2} \, r_{\text{TM}}(\lambda;\chi). \end{aligned} \tag{22.463}$$

Thus, we have the combination

$$\mathcal{I}_{D,3/2}(z) - \frac{1}{2}\partial_z^2 \mathcal{I}_{D-2,1/2}(z) = -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{D/2-1}\sqrt{\pi} z^D} \int_1^\infty d\lambda \, (\lambda-1)^{D/2-2} \lambda^{-(D+1)/2} (1-2\lambda) \, r_{\text{TM}}(\lambda;\chi), \tag{22.464}$$

and hence the potential from (22.458):

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0 z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi).$$

(TM Casimir–Polder potential) (22.465)

Recall that from from Eq. (22.279),

$$r_{\text{TM}}(\lambda;\chi) = \frac{\sqrt{\lambda+\chi}-\sqrt{\lambda}(1+\chi)}{\sqrt{\lambda+\chi}+\sqrt{\lambda}(1+\chi)}. \tag{22.466}$$

This is the analogue of the TE potential (22.336).

In 3D electromagnetism ($D = 4$), the result (22.465) becomes

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{3\hbar c\alpha_0}{128\pi^2\epsilon_0 z^4}\int_1^\infty d\lambda\,\lambda^{-5/2}(2\lambda-1)\left(\frac{\sqrt{\lambda+\chi}-\sqrt{\lambda}(1+\chi)}{\sqrt{\lambda+\chi}+\sqrt{\lambda}(1+\chi)}\right). \tag{22.467}$$

This integral is equivalent to the one that comes out of the Green-tensor analysis of the atom–wall problem [see Eq. (14.210); the variable change $\xi = \sqrt{\lambda}$ (and thus $d\lambda = 2\xi\,d\xi$) yields twice the second term in the integral]. Adapting the solution there, we have

$$V_{\text{CP}}^{(\text{TM})}(z) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left[\frac{7}{6}+\chi+\frac{2-(1+\chi)\sqrt{1+\chi}}{2\chi}-\frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\left[1+\chi+2(1+\chi)\chi^2\right]\right.$$
$$\left.+\frac{(1+\chi)^2}{\sqrt{2+\chi}}\left(\sinh^{-1}\sqrt{1+\chi}-\sinh^{-1}\frac{1}{\sqrt{1+\chi}}\right)\right].$$

(TM Casimir–Polder potential) (22.468)

The bracketed quantity has asymptotes $43\chi/120$ for small χ and $5/6$ for large χ, and acts as an ''efficiency'' η_{TM} for the TM energy compared to the full-electromagnetism, perfect-conductor result [see Eq. (14.214)]. The small-χ result agrees with our earlier calculation (22.268). The large-χ result also agrees with our naïve strong-coupling result (22.255), except of course for the overall sign.

22.9.4.1 Digression: Derivative-Free Path Integrals

Although we have shown that we can evaluate the path integral (22.457) without problem, from a numerical point of view the presence of derivatives is a nuisance, as they tend to amplify numerical and statistical fluctations. Thus, we will spend a bit of time deriving alternative path integrals for the TM polarization *without* any such derivatives. One obvious approach is to start, as we have already noted, that the derivative with respect to the atomic position is a derivative with respect to the coordinate of the interface, up to a minus sign. But varying the distance to the surface is has a similar effect to varying the total path time $\mathcal{T}$, because it varies the portion of the path that ''contacts'' the interface. Because a Brownian bridge running from 0 to $\mathcal{T}$ has an extent that scales as $\sqrt{\mathcal{T}}$, we expect the path average in the integrand of the $\mathcal{T}$ integral to depend on d and $\mathcal{T}$ *only* via the combination $d/\sqrt{\mathcal{T}}$. This is explicitly the case, for example, for the sojourn time that appears explicitly in the TE path integral as well as the TM case [see, e.g., Eq. (22.231)]. Thus, consider a derivative with respect to the atom–surface distance d of such an integrand, which we can write schematically as

$$\partial_d f\left(\frac{d}{\sqrt{T}}\right) = \frac{1}{\sqrt{T}}f'\left(\frac{d}{\sqrt{T}}\right). \tag{22.469}$$

The analogous $\mathcal{T}$ derivative reads

$$\partial_{\mathcal{T}} f\left(\frac{d}{\sqrt{T}}\right) = -\frac{d}{2^{3/2}}f'\left(\frac{d}{\sqrt{T}}\right). \tag{22.470}$$

Comparing these two expressions, we may then identify

$$\partial_d \equiv -\frac{2\mathcal{T}}{d}\partial_{\mathcal{T}} \tag{22.471}$$

whenever operating on such an integrand. Then in general, we have an integrand with an operator of the form

$$\begin{aligned} T^{-D/2}\partial_d^2 &= -T^{-D/2}\partial_d \frac{2\mathcal{T}}{d}\partial_{\mathcal{T}} \\ &= -2T^{-D/2}\partial_d \frac{\mathcal{T}}{d^3}(d^2\partial_{\mathcal{T}}) \\ &= -2T^{-(1+D)/2}\partial_d \frac{\mathcal{T}^{3/2}}{d^3}(d^2\partial_{\mathcal{T}}) \\ &= 4T^{-(1+D)/2}\frac{\mathcal{T}}{d}\partial_{\mathcal{T}}\frac{\mathcal{T}^{3/2}}{d^3}(d^2\partial_{\mathcal{T}}) \\ &= \frac{4}{d^2}T^{-(D-1)/2}\partial_{\mathcal{T}}\mathcal{T}^{3/2}\partial_{\mathcal{T}}, \end{aligned} \tag{22.472}$$

where note that we were careful to ensure the variables to the right only appeared in the combination $d/\sqrt{\mathcal{T}}$ before changing the derivative variable. Now we can integrate by parts twice under the $\mathcal{T}$ integral to eliminate the derivatives:

$$\begin{aligned} T^{-D/2}\partial_d^2 &= \frac{2(D-1)}{d^2}T^{-(D+1)/2}\mathcal{T}^{3/2}\partial_{\mathcal{T}} \\ &= \frac{2(D-1)}{d^2}T^{1-D/2}\partial_{\mathcal{T}} \\ &= \frac{(D-1)(D-2)}{d^2}T^{-D/2}. \end{aligned} \tag{22.473}$$

Under this substitution, the path integral (22.457) becomes

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}(D-1)(D-2)}{d^2}\right)\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}.$$

(TM path integral, derivative-free) (22.474)

Note, however, that while we have eliminated the derivatives, we have had to explicitly introduce the atom–surface distance d. While this is fine for the single-planar-interface calculation, it does not obviously generalize to more general geometries.

A better path integral without an explicit distance results if we start from Eq. (22.455),

$$\begin{aligned} \mathcal{I}_{D,1/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-1}\Gamma(D/2)z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-1}\lambda^{-(D+1)/2}\,r_{\text{TM}}(\lambda;\chi) \\ &= -\frac{\Gamma(D)}{2^{3D/2-1}(D/2-1)\Gamma(D/2-1)z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-1}\lambda^{-(D+1)/2}\,r_{\text{TM}}(\lambda;\chi), \end{aligned} \tag{22.475}$$

which we can rewrite as

$$\begin{aligned} (D/2-1)\,\mathcal{I}_{D,1/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-1}\Gamma(D/2-1)z^D}\int_1^\infty d\lambda\,(\lambda-1)\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r_{\text{TM}}(\lambda;\chi) \\ &= -\frac{\Gamma(D)}{2^{3D/2-1}\Gamma(D/2-1)z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D-1)/2}\,r_{\text{TM}}(\lambda;\chi) \\ &\qquad + \frac{\Gamma(D)}{2^{3D/2-1}\Gamma(D/2-1)z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r_{\text{TM}}(\lambda;\chi) \\ &= \frac{1}{8}\partial_z^2\mathcal{I}_{D-2,1/2}(z) - \frac{1}{2}\mathcal{I}_{D,3/2}(z), \end{aligned} \tag{22.476}$$

after comparison with Eqs. (22.459) and (22.463). Rearranging, we have

$$\frac{1}{2}\partial_z^2\mathcal{I}_{D-2,1/2}(z) = 2(D-2)\,\mathcal{I}_{D,1/2}(z) + 2\mathcal{I}_{D,3/2}(z). \tag{22.477}$$

Then we can eliminate the derivative term in Eq. (22.458), which becomes

$$V_{\text{CP}}^{(\text{TM})}(z) = -\frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\Big(\mathcal{I}_{D,3/2}(z) + 2(D-2)\,\mathcal{I}_{D,1/2}(z)\Big). \tag{22.478}$$

Evidently, we may rewrite Eq. (22.457) in derivative-free form as

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = -\frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2} + 2(D-2)\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\right) e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}.$$

(TM path integral, derivative- and distance-free) (22.479)

Tracing the algebra for the $\mathcal{I}_{D,\alpha}(z)$ back to Eq. (22.327), we can see that the replacements for the derivatives ∂_z^2 (the factor of $4(D-2)$ on the second term, and the inverted minus sign on the first term) ultimately came from the derivative changing the power of $(\lambda+s)$ in the integrand, due to the exponential factor that involved z. This type of factor is generic in Casimir–Polder calculations, so we expect this result to hold even in the general case of multiple objects [for two surfaces, for example, see the TE two-plane Casimir–Polder calculation, Eq. (22.619)].

22.9.5 Result: 1D Electromagnetism

For $D = 2$, the Casimir–Polder potential has a removable singularity as in the TE case (Section 22.7.5), which we will again remove via dimensional regularization. The procedure is the same as before. First setting $D = 2$ in Eq. (22.465) in all places where there won't be any problems, and keeping $D > 2$ elsewhere,

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{(D-2)\hbar c\alpha_0}{32\pi\epsilon_0 z^2}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-3/2}(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi). \tag{22.480}$$

Integrating by parts gives

$$\begin{aligned} V_{\text{CP}}^{(\text{TM})}(z) = {}& \frac{(D-2)\hbar c\alpha_0}{(D/2-1)32\pi\epsilon_0 z^2}\Big[(\lambda-1)^{D/2-1}\lambda^{-3/2}(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi)\Big]_1^\infty \\ & - \frac{(D-2)\hbar c\alpha_0}{(D/2-1)32\pi\epsilon_0 z^2}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-1}\partial_\lambda\Big[\lambda^{-3/2}(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi)\Big]. \end{aligned} \tag{22.481}$$

Again, since $D > 2$, the boundary terms vanish at both the lower limit (due to the $\lambda - 1$ factor) and upper limits (provided D is not much larger than 2). The expression is now regular at $D = 2$, and takes the form

$$V_{\text{CP}}^{(\text{TM})}(z) = -\frac{\hbar c\alpha_0}{16\pi\epsilon_0 z^2}\int_1^\infty d\lambda\,\partial_\lambda\Big[\lambda^{-3/2}(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi)\Big]. \tag{22.482}$$

Now we can evaluate the integral. The result evaluated at the upper limit vanishes. Therefore, we find

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{\hbar c\alpha_0}{16\pi\epsilon_0 z^2}\,r_{\text{TM}}(1;\chi) = -\frac{\hbar c\alpha_0}{16\pi\epsilon_0 z^2}\left(\frac{\sqrt{1+\chi}-1}{\sqrt{1+\chi}+1}\right),$$

(TM Casimir–Polder potential, $D = 2$) (22.483)

where we used

$$r_{\text{TM}}(1;\chi) = \frac{\sqrt{1+\chi}-(1+\chi)}{\sqrt{1+\chi}+(1+\chi)} = \frac{1-\sqrt{1+\chi}}{1+\sqrt{1+\chi}} = r_{\text{TE}}(1;\chi), \tag{22.484}$$

which follows from Eqs. (22.274) and (22.279). This agrees with the TE result (22.344), as it should, since in 1D electromagnetism (where all waves are effectively at normal incidence), there should be no distinction between the two polarizations.

22.9.6 Casimir–Polder Potential: Full Electromagnetism

Now that we have worked out the TM Casimir–Polder path integral, after doing the same for the TE case quite a while back in Section 22.7.4, we will wrap up this discussion by briefly summarizing the combined results of both sections to obtain the Casimir–Polder potential in full electromagnetism, especially in the mainly relevant case of 3D electromagnetism. First, combining Eqs. (22.336) and Eqs. (22.465), we see that we can write the total potential as

$$V_{\text{CP}}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0 z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\Big[r_{\text{TE}}(\lambda;\chi)+(2\lambda-1)\,r_{\text{TM}}(\lambda;\chi)\Big],$$

(Casimir–Polder potential) (22.485)

where the reflection coefficients are defined in Eqs. (22.274) and (22.279). We see here that the contributions add in the form of combined reflection coefficients, with an extra factor of $(2\lambda-1)$ in the TM case. Recall that, in terms of a mode sum, this is a geometric factor: for a dielectric particle interacting with the electric field, the TE field behaves as a scalar (the electric field is always parallel to the planar interface) while it behaves as a *vector* in the TM case, so that the geometric factor is related to the dot product of the incident and reflected wave vectors for vacuum-incident modes.

In the $(D=4)$ case of 3D electromagnetism, we can summarize the results from Eqs. (22.339) and (22.468) as potentials

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta_{\text{TE}}(\chi), \qquad V_{\text{CP}}^{(\text{TM})}(z) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta_{\text{TM}}(\chi),$$

(TE and TM Casimir–Polder potentials) (22.486)

where we have defined ''efficiencies'' relative to the strong-coupling ($\chi\longrightarrow\infty$) case of [cf. Eq. (14.214)]

$$\begin{aligned}\eta_{\text{TE}}(\chi) &:= \frac{1}{6}+\frac{1}{\chi}-\frac{\sqrt{1+\chi}}{2\chi}-\frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\\ \eta_{\text{TM}}(\chi) &:= \frac{7}{6}+\chi+\frac{2-(1+\chi)\sqrt{1+\chi}}{2\chi}-\frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\,\big[1+\chi+2(1+\chi)\chi^2\big]\\ &\quad+\frac{(1+\chi)^2}{\sqrt{2+\chi}}\left(\sinh^{-1}\sqrt{1+\chi}-\sinh^{-1}\frac{1}{\sqrt{1+\chi}}\right).\end{aligned}$$

(TE and TM Casimir–Polder efficiencies) (22.487)

We have already discussed the asymptotic behaviors of these efficiencies in Sections 22.7.4 and 22.9.4.

In the full-electromagnetism case, we can thus most compactly combine the individual efficiencies

$$\eta(\chi) := \eta_{\text{TE}}(\chi)+\eta_{\text{TM}}(\chi), \tag{22.488}$$

in which case the full potential is

$$V_{\text{CP}}(z) = -\frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta(\chi).$$

(TE + TM Casimir–Polder potential) (22.489)

Written out explicitly, the full efficiency is [cf. Eq. (14.212)]

$$\begin{aligned}\eta(\chi) &= \frac{4}{3}+\chi+\frac{4-(2+\chi)\sqrt{1+\chi}}{2\chi}-\frac{\sinh^{-1}\sqrt{\chi}}{2\chi^{3/2}}\,\big[2+\chi+2(1+\chi)\chi^2\big]\\ &\quad+\frac{(1+\chi)^2}{\sqrt{2+\chi}}\left(\sinh^{-1}\sqrt{1+\chi}-\sinh^{-1}\frac{1}{\sqrt{1+\chi}}\right).\end{aligned}$$

(TE + TM Casimir–Polder efficiency) (22.490)

This efficiency is asymptotically $23\chi/60$ for small χ, and for large χ, $\eta(\chi)\longrightarrow 1$ (which of course was the point of the definition).

22.10 Casimir–Polder Potentials Within Media

A relatively straightforward generalization of the above results is to compute the Casimir–Polder potential for an atom near a dielectric interface, but with the atom *within* the dielectric. This is a model for an atom in a gas near an interface, or an atom in a liquid near a bubble.

22.10.1 TE Polarization

22.10.1.1 Feynman–Kac Formula

For TE polarization, we return to the path integral (22.197):

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.491}$$

The calculation parallels the calculation with the atom in vacuum of Section 22.7. The first difference is that in the Feynman–Kac solution (22.308), we should choose the solution for $x > d$ to put the atom inside the dielectric. Then instead of Eq. (22.314), we have

$$\begin{aligned} L(\lambda, s; d) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk\, f(0) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dk \left(A + \frac{2}{2(\lambda+s)+k^2} \right) \\ &= \sqrt{\frac{\pi}{\lambda+s}} \left[1 + e^{2\sqrt{2(\lambda+s)}\,d} \left(\frac{\sqrt{\lambda+s}-\sqrt{\lambda}}{\sqrt{\lambda+s}+\sqrt{\lambda}} \right) \right] \\ &= \sqrt{\frac{\pi}{\lambda+s}} \left[1 - e^{2\sqrt{2(\lambda+s)}\,d}\, r_{\text{TE}}(\lambda; s) \right], \end{aligned} \tag{22.492}$$

where we are using the reflection coefficient (22.274) incident from the *vacuum* side, even though we are considering a particle on the dielectric side (although the sign of the reflection coefficient differs, as we expect for the dielectric side). The reason for this will become more clear shortly. Note that the result here can be obtained from the previous result (22.314) by formally interchanging λ and $\lambda + s$, and changing the sign of d.

22.10.1.2 Mellin Transform

Proceeding in analogy to Section 22.7.2, we can use the replacements $s \longrightarrow s\chi$ and $\lambda \longrightarrow \lambda + s$ to write

$$\begin{aligned} L(\lambda + s, s\chi; d) &= \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1/2}}\, e^{-\lambda\mathcal{T}} \left\langle\!\!\left\langle e^{-s(\mathcal{T} + \chi T_{\text{s}}[B_{\mathcal{T}}])} \right\rangle\!\!\right\rangle \\ &= \sqrt{\frac{\pi}{\lambda + s(1+\chi)}} \left[1 - e^{2\sqrt{2[\lambda+s(1+\chi)]}\,d}\, r_{\text{TE}}(\lambda + s; s\chi) \right]. \end{aligned} \tag{22.493}$$

We will use this to compute the integral

$$\tilde{I}_D(s, \chi; z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle e^{-s(\mathcal{T} + \chi T_{\text{s}}[B_{\mathcal{T}}; z])} - e^{-s(\mathcal{T} + \chi T_{\text{s}}[B_{\mathcal{T}}; z \to \infty])} \right\rangle\!\!\right\rangle, \tag{22.494}$$

which is the counterpart to Eq. (22.321). Note that here, we are renormalizing by subtracting the limit of large distance $z = -d > 0$ from the interface. In the previous case, this was equivalent to setting $\chi = 0$, but this time the ''background'' case is an atom in a uniform dielectric. Using the same procedure as for

(22.317), we can write

$$\begin{aligned}\tilde{I}_D(s,\chi;z) &= \frac{1}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\lambda^{(D-1)/2}\Big[L(\lambda+s,s\chi;z)-L(\lambda+s,s\chi;z\to\infty)\Big]\\ &= -\frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\frac{\lambda^{(D-1)/2}}{\sqrt{\lambda+s(1+\chi)}}\,r_{\text{TE}}(\lambda+s;s\chi)\,e^{-2\sqrt{2[\lambda+s(1+\chi)]}\,z}.\end{aligned} \tag{22.495}$$

Again, this is the same as the atom-in-vacuum case so far, except for the factors of $(1+\chi)$.

22.10.1.3 Inverse Moments

Now for the dielectric-interface path integral, the integral that we need to compute is of the form

$$\mathcal{I}_{D,\alpha}(z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \frac{1}{(1+\chi T_{\text{s}}[B_{\mathcal{T}};z]/\mathcal{T})^\alpha} - \frac{1}{(1+\chi)^\alpha}\right\rangle\!\!\right\rangle, \tag{22.496}$$

in analogy to Eq. (22.323). Note that the renormalization here is different, as appropriate for a particle on the dielectric side of the interface. This cuts off the divergence at $\mathcal{T}$, where the paths become arbitrarily small, and so $T_{\text{s}}[B_{\mathcal{T}};z]=\mathcal{T}$ on the dielectric side. Rearranging this expression to be more compatible with Eq. (22.495),

$$\mathcal{I}_{D,\alpha}(z) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2-\alpha}}\left\langle\!\!\left\langle \frac{1}{(\mathcal{T}+\chi T_{\text{s}}[B_{\mathcal{T}};z])^\alpha} - \frac{1}{(1+\chi)^\alpha\mathcal{T}^\alpha}\right\rangle\!\!\right\rangle. \tag{22.497}$$

Then using (22.326),

$$\mathcal{I}_{D,\alpha}(z) = \frac{1}{\Gamma(\alpha)}\int_0^\infty ds\,s^{\alpha-1}\tilde{I}_{D-2\alpha}(z), \tag{22.498}$$

with Eq. (22.495) this becomes

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}}{\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\,s^{\alpha-1}\int_0^\infty d\lambda\,\frac{\lambda^{(D-1)/2-\alpha}}{\sqrt{\lambda+s(1+\chi)}}\,r_{\text{TE}}(\lambda+s;s\chi)\,e^{-2\sqrt{2[\lambda+s(1+\chi)]}\,z}, \tag{22.499}$$

as in Eq. (22.327). Simplifying as before by letting $\lambda\longrightarrow\lambda/8z^2$ and $s\longrightarrow s/8z^2$, then letting $\lambda\longrightarrow\lambda-s$, and then letting $\lambda\longrightarrow\lambda s$,

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}\,r_{\text{TE}}(\lambda;\chi)}{\sqrt{\lambda+\chi}}\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{s(\lambda+\chi)}}, \tag{22.500}$$

which compares to Eq. (22.330). Then letting $\lambda\longrightarrow\lambda-\chi$,

$$\begin{aligned}\mathcal{I}_{D,\alpha}(z) = -&\frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\\ &\times\int_{1+\chi}^\infty d\lambda\,\frac{[\lambda-(1+\chi)]^{(D-1)/2-\alpha}\,r_{\text{TE}}(\lambda-\chi;\chi)}{\sqrt{\lambda}}\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{\lambda s}},\end{aligned} \tag{22.501}$$

In the next step, we will let $\lambda\longrightarrow(1+\chi)\lambda$, and we note that

$$r_{\text{TE}}[\lambda(1+\chi)-\chi;\chi] = \frac{\sqrt{\lambda(1+\chi)-\chi}-\sqrt{\lambda(1+\chi)}}{\sqrt{\lambda(1+\chi)-\chi}+\sqrt{\lambda(1+\chi)}} = -r'_{\text{TE}}(\lambda;\chi), \tag{22.502}$$

where we used Eqs. (22.274) and (22.276), and we see that we are switching via the natural transformations of the problem from the vacuum-side to the dielectric-side reflection coefficient. Thus with $\lambda\longrightarrow(1+\chi)\lambda$ in Eq. (22.501), we have

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}(1+\chi)^{D/2-\alpha}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}\,r'_{\text{TE}}(\lambda;\chi)}{\sqrt{\lambda}}\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{\lambda(1+\chi)s}}. \tag{22.503}$$

The integral over s now has the value $2\Gamma(D)[\lambda(1+\chi)]^{-D/2}$, so

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big](1+\chi)^\alpha z^D}\int_1^\infty d\lambda\,(\lambda-1)^{(D-1)/2-\alpha}\lambda^{-(D+1)/2}\,r'_{\text{TE}}(\lambda;\chi). \quad (22.504)$$

as in Eq. (22.331). Note, in fact, that the only differences here compared to Eq. (22.331) are the replacement of r_{TE} with r'_{TE} and the factor of $(1+\chi)^{-\alpha}$. Now we will proceed by being more specific about α and D.

22.10.1.4 Result: 3D Electromagnetism

Now since the Casimir–Polder path integral (22.302) contains the integral (22.496) after renormalization, we have

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\mathcal{I}_{D,3/2}(z), \quad (22.505)$$

just as in Eq. (22.332). With the result (22.331), this becomes

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\Gamma(D)\hbar c\alpha_0}{2^{3D/2}(2\pi)^{D/2}\Gamma\big[D/2-1\big]\epsilon_0(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r'_{\text{TE}}(\lambda;\chi). \quad (22.506)$$

With Eq. (22.335), this result simplifies to

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r'_{\text{TE}}(\lambda;\chi),$$

(TE Casimir–Polder potential, dielectric side) (22.507)

where again [Eq. (22.276)]

$$r'_{\text{TE}}(\lambda;\chi) = \frac{\sqrt{\lambda(1+\chi)}-\sqrt{\lambda(1+\chi)-\chi}}{\sqrt{\lambda(1+\chi)}+\sqrt{\lambda(1+\chi)-\chi}}. \quad (22.508)$$

Notice again that this potential has the same form as the vacuum-side potential, except for the factor $(1+\chi)^{-3/2}$, and the dielectric-side reflection coefficient r'_{TE} appears in place of the vacuum-side version r_{TE}.[21]

For $D=4$, we have

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{3\hbar c\alpha_0}{128\pi^2\epsilon_0(1+\chi)^{3/2}z^4}\int_1^\infty d\lambda\,\lambda^{-5/2}\left(\frac{\sqrt{\lambda(1+\chi)}-\sqrt{\lambda(1+\chi)-\chi}}{\sqrt{\lambda(1+\chi)}+\sqrt{\lambda(1+\chi)-\chi}}\right), \quad (22.509)$$

and evaluating the remaining integral, we find

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{5}{6}+\frac{1}{\chi}-\frac{\sqrt{1+\chi}}{2\chi}-\frac{(1+\chi)^{3/2}}{2\chi^{3/2}}\tan^{-1}\sqrt{\chi}\right)(1+\chi)^{-3/2}.$$

(TE Casimir–Polder potential, dielectric side) (22.510)

The χ-dependent factor here is strictly positive. Thus the force here is *repulsive*, unlike the vacuum-side potential (22.339). For small χ, the χ-dependent factor is $\chi/40$, which is the same as the vacuum-side case (except for the overall minus sign). However, as $\chi\longrightarrow\infty$, the potential here decays to 0 as $(5/6-\pi/4)\chi^{-3/2}$, which is quite different to the vacuum-side case where the χ-dependent part levels off at $1/6$.

[21] See Fei Zhou and Larry Spruch, "van der Waals and retardation (Casimir) interactions of an electron or an atom with multilayered walls," *Physical Review A* **52**, 297 (1995), Eq. (4.55) (doi: 10.1103/PhysRevA.52.297).

22.10.1.5 Result: 1D Electromagnetism

In the case of 1D electromagnetism ($D = 2$), the procedure is the same as in Section 22.10.1.5, with the result

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2} r'_{\text{TE}}(1;\chi)\,(1+\chi)^{-3/2} = \frac{\hbar c \alpha_0}{16\pi\epsilon_0 z^2}\left(\frac{\sqrt{1+\chi}-1}{\sqrt{1+\chi}+1}\right)(1+\chi)^{-3/2}.$$

(TE Casimir–Polder potential, $D = 2$, dielectric side) (22.511)

Again, this has the same form as the vacuum-side expression (22.344), except for the overall sign and the factor $(1+\chi)^{-3/2}$.

22.10.2 TM Polarization

For the TM case of the Casimir–Polder potential for an atom on the dielectric side of the interface, we return to the path integral (22.198),

$$V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}}\nabla^2\left[\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}}\right]\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.512}$$

which is more general than the path integral (22.247) that we used in the vacuum-side calculation.

22.10.2.1 Feynman–Kac Formula

The derivation parallels that of the vacuum-side case in Section 22.9. The first difference is that we need the solution to the A coefficient from Eqs. (22.446):

$$A = e^{\sqrt{2(\lambda+s)}\,d + ikd}\left(\frac{e^{-\Xi}\left(\lambda+s+k^2/2\right)\left(\sqrt{\lambda}-ik/\sqrt{2}\right) - \left(\lambda+k^2/2\right)\left(\sqrt{\lambda}-ike^{-2\Xi}/\sqrt{2}\right)}{\left(\lambda+k^2/2\right)\left(\lambda+s+k^2/2\right)\left(e^{-2\Xi}\sqrt{\lambda+s}+\sqrt{\lambda}\right)}\right). \tag{22.513}$$

Then from the form of the solution $f(x)$ in Eq. (22.445), we need the function

$$\begin{aligned} L(\lambda,s;d,\Xi) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, f(0) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\,\left(\frac{1}{\lambda+s+k^2/2}+A\right) \\ &= \sqrt{\frac{\pi}{\lambda+s}}\left[1+e^{2\sqrt{2(\lambda+s)}\,d}\left(\frac{\sqrt{\lambda+s}-e^{2\Xi}\sqrt{\lambda}}{\sqrt{\lambda+s}+e^{2\Xi}\sqrt{\lambda}}\right)\right] \\ &= \sqrt{\frac{\pi}{\lambda+s}}\left[1-e^{2\sqrt{2(\lambda+s)}\,d}\,\mathfrak{r}(\lambda;s,\Xi)\right], \end{aligned} \tag{22.514}$$

where $L(\lambda,s;d,\Xi)$ is defined by Eq. (22.444). This result is in analogy to (22.449), but $d<0$ here. We are also using the notation (22.450)

$$\mathfrak{r}(\lambda;s,\Xi) := \frac{e^{2\Xi}\sqrt{\lambda}-\sqrt{\lambda+s}}{e^{2\Xi}\sqrt{\lambda}+\sqrt{\lambda+s}}. \tag{22.515}$$

22.10.2.2 Mellin Transform

The next step is to compute the path integral

$$\tilde{I}_D(s,\Xi;z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle e^{-\mathfrak{N}[B_{\mathcal{T}};z,\Xi]-s(\mathcal{T}+\chi T_{\text{s}}[B_{\mathcal{T}};z])} - e^{-s(\mathcal{T}+\chi T_{\text{s}}[B_{\mathcal{T}};z\to\infty])}\right\rangle\!\!\right\rangle, \tag{22.516}$$

which parallels (22.451), but with the more appropriate renormalization for this case. The result, as in Eq. (22.453), is

$$\begin{aligned}\tilde{I}_D(s,\chi;z) &= \frac{1}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\lambda^{(D-1)/2}\Big[L(\lambda+s,s\chi;z)-L(\lambda+s,s\chi;z\to\infty)\Big]\\ &= -\frac{\sqrt{\pi}}{\Gamma\big[(D+1)/2\big]}\int_0^\infty d\lambda\,\frac{\lambda^{(D-1)/2}}{\sqrt{\lambda+s(1+\chi)}}\,e^{-2\sqrt{2[\lambda+s(1+\chi)]}\,z}\,\mathfrak{r}(\lambda+s;s\chi,\Xi),\end{aligned} \tag{22.517}$$

where we are now taking $z=-d$.

22.10.2.3 Inverse Moments

Next, we compute the integral (22.454),

$$\mathcal{I}_{D,\alpha}(z) := \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \frac{e^{-\Re[B_\mathcal{T};z,\Xi]}}{(1+\chi T_s[B_\mathcal{T};z]/\mathcal{T})^\alpha} - \frac{1}{(1+\chi)^\alpha}\right\rangle\!\!\right\rangle, \tag{22.518}$$

but modified here for a distant-interface renormalization. To do this, we take the inverse-moment formula (22.295), in the form of Eq. (22.326),

$$\mathcal{I}_{D,\alpha}(z) = \frac{1}{\Gamma(\alpha)}\int_0^\infty ds\,s^{\alpha-1}\tilde{I}_{D-2\alpha}(z). \tag{22.519}$$

and apply it to Eq. (22.517):

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}}{\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\,s^{\alpha-1}\int_0^\infty d\lambda\,\frac{\lambda^{(D-1)/2-\alpha}}{\sqrt{\lambda+s(1+\chi)}}\,e^{-2\sqrt{2[\lambda+s(1+\chi)]}\,z}\,\mathfrak{r}(\lambda+s;s\chi,\Xi). \tag{22.520}$$

Now absorbing a factor of $\sqrt{8}z^2$ into λ and s, letting $\lambda\longrightarrow\lambda-s$, and then $\lambda\longrightarrow\lambda s$,

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}}{\sqrt{\lambda+\chi}}\,\mathfrak{r}(\lambda;\chi,\Xi)\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{s(\lambda+\chi)}}. \tag{22.521}$$

Using Eq. (22.456) to eliminate $\mathfrak{r}$ in favor of the TM reflection coefficient,

$$\mathcal{I}_{D,\alpha}(z) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}}{\sqrt{\lambda+\chi}}\,r_{\text{TM}}(\lambda;\chi)\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{s(\lambda+\chi)}}, \tag{22.522}$$

and then letting $\lambda\longrightarrow\lambda-\chi$,

$$\begin{aligned}\mathcal{I}_{D,\alpha}(z) = &\frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\\ &\times\int_{1+\chi}^\infty d\lambda\,\frac{[\lambda-(1+\chi)]^{(D-1)/2-\alpha}}{\sqrt{\lambda}}\,r_{\text{TM}}(\lambda-\chi;\chi)\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{s\lambda}}.\end{aligned} \tag{22.523}$$

Then letting $\lambda\longrightarrow(1+\chi)\lambda$,

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}(1+\chi)^{D/2-\alpha}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]z^D}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}\,r'_{\text{TM}}(\lambda;\chi)}{\sqrt{\lambda}}\int_0^\infty ds\,s^{D/2-1}\,e^{-\sqrt{s\lambda(1+\chi)}}, \tag{22.524}$$

where we switched to the dielectric-side reflection coefficient using

$$r_{\text{TM}}[\lambda(1+\chi)-\chi;\chi] = \frac{\sqrt{\lambda}-\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}}{\sqrt{\lambda}+\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}} - r'_{\text{TM}}(\lambda;\chi), \tag{22.525}$$

which follows from Eqs. (22.279) and (22.281). The integral over s again has the value $2\Gamma(D)[\lambda(1+\chi)]^{-D/2}$, so

$$\mathcal{I}_{D,\alpha}(z) = -\frac{\sqrt{\pi}\Gamma(D)}{2^{3D/2-1}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big](1+\chi)^{\alpha}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{(D-1)/2-\alpha}\lambda^{-(D+1)/2}\,r'_{\text{TM}}(\lambda;\chi). \tag{22.526}$$

This is the analogue of Eq. (22.504) in the TE dielectric-side case, or Eq. (22.455) in the TM vacuum-side case.

22.10.2.4 Result: 3D Electromagnetism

Now to evaluate the path integral (22.512) for the Casimir–Polder potential, we see from the definition (22.518) that we can write the potential as

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\left(\mathcal{I}_{D,3/2}(z) - \frac{1}{2\epsilon_{\text{r}}}\partial_z^2\mathcal{I}_{D-2,1/2}(z)\right) \tag{22.527}$$

for an atom at a distance z from a planar dielectric interface, on the dielectric side. Notice the factor of $\epsilon_{\text{r}} = 1+\chi$ (evaluated at the atomic position) that occurs in the second term here, which was absent from the vacuum-side expression (22.458).

Then we can take the special case of Eq. (22.526):

$$\begin{aligned}\mathcal{I}_{D,3/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-2}\Gamma(D/2-1)(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r'_{\text{TM}}(\lambda;\chi)\\ &= -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{3D/2-1}\sqrt{\pi}(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\,r'_{\text{TM}}(\lambda;\chi),\end{aligned} \tag{22.528}$$

where we used Eq. (22.461) to transform the gamma functions. Similarly,

$$\begin{aligned}\partial_z^2\mathcal{I}_{D-2,1/2}(z) &= -\frac{\Gamma(D)}{2^{3D/2-4}\Gamma(D/2-1)(1+\chi)^{1/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D-1)/2}\,r'_{\text{TM}}(\lambda;\chi)\\ &= -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{D/2-3}\sqrt{\pi}(1+\chi)^{1/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D-1)/2}\,r'_{\text{TM}}(\lambda;\chi).\end{aligned} \tag{22.529}$$

Thus, we have the combination

$$\mathcal{I}_{D,3/2}(z) - \frac{1}{2\epsilon_{\text{r}}}\partial_z^2\mathcal{I}_{D-2,1/2}(z) = -\frac{(D/2-1)\Gamma\big[(D+1)/2\big]}{2^{D/2-1}\sqrt{\pi}(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{(1-D)/2}(1-2\lambda)\,r'_{\text{TM}}(\lambda;\chi), \tag{22.530}$$

in analogy to Eq. (22.464). Then from Eq. (22.527), we have the potential

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}(2\lambda-1)\,r'_{\text{TM}}(\lambda;\chi)$$

(TM Casimir–Polder potential, dielectric side) (22.531)

where from Eq. (22.281),

$$r'_{\text{TM}}(\lambda;\chi) = \frac{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}-\sqrt{\lambda}}{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}+\sqrt{\lambda}}. \tag{22.532}$$

This is the analogue of the TM vacuum-side potential (22.465), having the same form except for the factor $(1+\chi)^{-3/2}$ and the appropriately modified reflection coefficient (the same changes that occurred in the TE case).

For $D = 4$, this becomes

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{3\hbar c\alpha_0}{128\pi^2\epsilon_0(1+\chi)^{3/2}z^4}\int_1^\infty d\lambda\,\lambda^{-5/2}(2\lambda-1)\left(\frac{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}-\sqrt{\lambda}}{\sqrt{(1+\chi)[(1+\chi)\lambda-\chi]}+\sqrt{\lambda}}\right). \tag{22.533}$$

Evaluating the integral, we have

$$V_{\text{CP}}^{(\text{TM})}(z) = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\left[\frac{5}{6} + \frac{2-\sqrt{1+\chi}}{2\chi(1+\chi)} - \frac{1+\chi(2+3\chi)}{2\chi^{3/2}(1+\chi)^{3/2}}\tan^{-1}\sqrt{\chi} + \frac{1}{(1+\chi)^{3/2}\sqrt{2+\chi}}\left(\tanh^{-1}\sqrt{\frac{1+\chi}{2+\chi}} - \coth^{-1}\sqrt{2+\chi}\right)\right](1+\chi)^{-3/2}.$$

(TM Casimir–Polder potential, dielectric side) (22.534)

For small χ, the χ-dependent factor is $43\chi/120$, as in the vacuum-side case (except for the overall minus sign). For large χ, this again drops to zero as $(5/6)\chi^{-3/2}$, unlike the vacuum-side case, which levels off to $5/6$.

22.10.2.5 Casimir–Polder Efficiencies

In analogy to the vacuum-side analysis of Section 22.9.6, we will combine and summarize the TM and TE results of our calculations above for the dielectric-side Casimir–Polder potentials, and rewrite our results in terms of efficiencies relative to the strong-coupling result. First, the combined Casimir–Polder potential on the dielectric side of the interface is, from Eqs. (22.507) and (22.531),

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{(D-2)\Gamma\big[(D+1)/2\big]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0(1+\chi)^{3/2}z^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\Big[r'_{\text{TE}}(\lambda;\chi) + (2\lambda-1)\,r'_{\text{TM}}(\lambda;\chi)\Big].$$

(TE + TM Casimir–Polder potential, dielectric side) (22.535)

as in the vacuum-side result (22.485), the total involves the combination of reflection coefficients, with a geometric factor on the TM coefficient. The dielectric-side Fresnel coefficients here are defined in Eqs. (22.276) and (22.281).

Specializing to the $(D = 4)$ case of 3D electromagnetism, we can summarize the results from Eqs. (22.510) and (22.534) as potentials

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta'_{\text{TE}}(\chi), \qquad V_{\text{CP}}^{(\text{TM})}(z) = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta'_{\text{TM}}(\chi).$$

(TE and TM Casimir–Polder potentials, dielectric side) (22.536)

Note that we are comparing to the *vacuum-side*, strong coupling limit, but *without* the minus sign, so that the efficiencies are positive (and, as it turns out, they will never achieve unity, even when combined). We are using primes to notate the efficiencies here, compared to the vacuum-side efficiencies (22.487), to denote that these efficiencies quantify the dielectric-side potentials. Explicitly, the efficiencies are

$$\eta'_{\text{TE}}(\chi) := (1+\chi)^{-3/2}\left(\frac{5}{6} + \frac{1}{\chi} - \frac{\sqrt{1+\chi}}{2\chi} - \frac{(1+\chi)^{3/2}}{2\chi^{3/2}}\tan^{-1}\sqrt{x}\right)$$

$$\eta'_{\text{TM}}(\chi) := (1+\chi)^{-3/2}\left[\frac{5}{6} + \frac{2-\sqrt{1+\chi}}{2\chi(1+\chi)} - \frac{1+\chi(2+3\chi)}{2\chi^{3/2}(1+\chi)^{3/2}}\tan^{-1}\sqrt{\chi} + \frac{1}{(1+\chi)^{3/2}\sqrt{2+\chi}}\left(\tanh^{-1}\sqrt{\frac{1+\chi}{2+\chi}} - \coth^{-1}\sqrt{2+\chi}\right)\right].$$

(TE and TM Casimir–Polder efficiencies, dielectric side) (22.537)

as can be read off directly from Eqs. (22.510) and (22.534). We have already discussed the asymptotic behaviors of these efficiencies in Sections 22.10.1.4 and 22.10.2.4, and it is worth reiterating that $\eta'_{\text{TE}} > 0$ and $\eta'_{\text{TM}} > 0$ for all $\chi \geq 0$.

In the full-electromagnetism case, we can then combine the individual efficiencies

$$\eta'(\chi) := \eta'_{\text{TE}}(\chi) + \eta'_{\text{TM}}(\chi),$$

(full Casimir–Polder efficiency, dielectric side) (22.538)

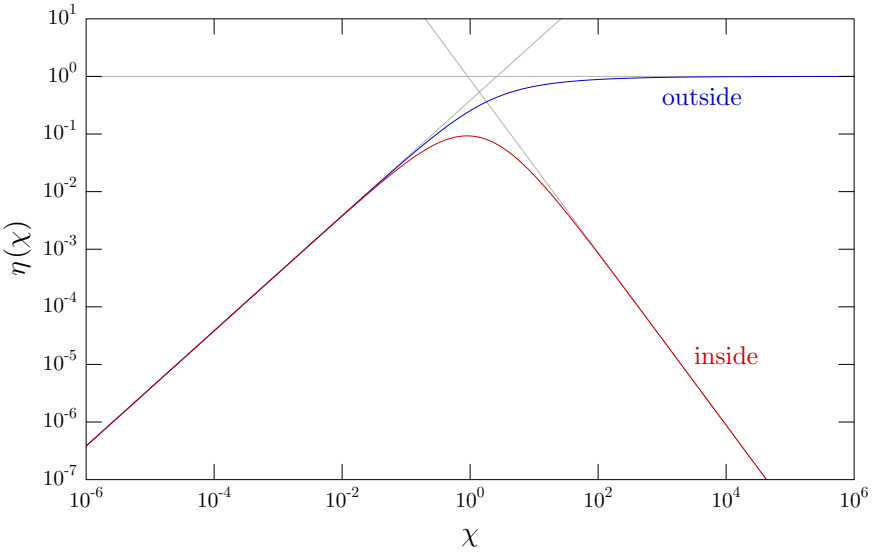
in which case the full potential is

$$V_{\text{CP}}(z) = \frac{3\hbar c\alpha_0}{32\pi^2\epsilon_0 z^4}\,\eta'(\chi).$$

(TE + TM Casimir–Polder potential, dielectric side.) (22.539)

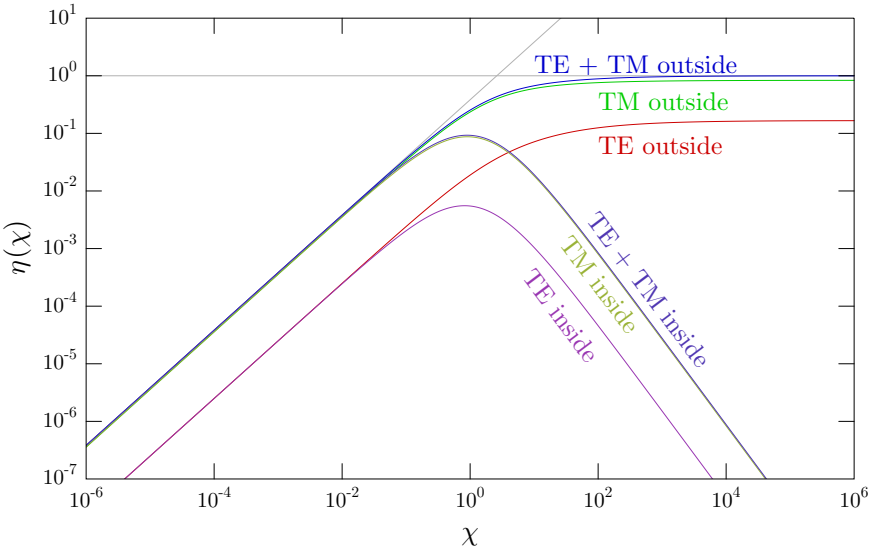
Since $\eta'(\chi)$ here is fairly cumbersome, we won't bother to write it out here. The efficiency is asymptotically $23\chi/60$ for small χ, just as in the vacuum-side case, and for large χ, $\eta'(\chi)$ scales as $(5/3 - \pi/4)\chi^{-3/2}$.

Plotting the "inside" (dielectric-side) efficiency $\eta'(\chi)$ from Eq. (22.538) along with the "outside" (vacuum-side) efficiency $\eta(\chi)$ from Eq. (22.490), we see the matching at small χ, but a dramatic difference in scaling at large χ, reflecting a sort of "screening" of the interface by the dielectric.



We are also displaying the asymptotes for the curves for small and large χ.

Plotting the individual TE and TM components for the "inside" (dielectric-side) efficiencies [Eq. (22.537)] and the "outside" (vacuum-side) efficiencies [Eq. (22.487)], we can see the relative contributions of each polarization.



The TE polarization contributes relatively little compared to the TM polarization, on either side of the interface.

22.11 Electric vs. Magnetic Casimir–Polder Interactions

Recall that the general forms (22.195) and (22.196) for the Casimir–Polder potential encapsulate both electric and magnetic interactions of particles with surfaces. So far, we have focused on the special case of a dielectric particle (atom) of polarizability α_0, interacting with a dielectric interface, where the dielectric function is constant in the neighborhood of the particle. This led to the path-integral expressions (22.197) and (22.198) for the TE and TM energies, respectively. For convencience we reproduce these here:

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2\epsilon_{\text{r}}} \nabla^2 \right) \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \, e^{-\mathcal{T}\langle V_{\text{TM}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \end{aligned} \tag{22.540}$$

In this section we will briefly review the analogous forms for other combinations of electric and magnetic interactions, and compare them to this one. The point of this exercise is to gain some more intuition for the general path integrals and for magnetic interactions in particular. Of course, the dielectric–dielectric interaction is the most important, since typically atoms and materials with large magnetic responses have even larger dielectric responses: in atoms, magnetic dipoles are comparable to electric quadrupoles, and in materials, magnetic materials are typically metallic conductors.

22.11.1 Dielectric Particle, Magnetic Surface

First, let's consider a dielectric particle (atom) interacting with a purely magnetic surface. We can do this by setting $\beta_0 = 0$ and $\epsilon_{\text{r}} = 1$ in Eqs. (22.195) and (22.196):

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \mu_{\text{r}} \langle \mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} \, e^{-\mathcal{T}\langle V_{\text{TE}} \rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \mu_{\text{r}} \langle \mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} - \frac{\mathcal{T}}{2} \nabla^2 \langle \mu_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \end{aligned}$$

(unrenormalized Casimir–Polder potential, magnetic body) (22.541)

We have also set the derivatives of μ_{r} to zero at the particle location, as appropriate, for example, for a particle in vacuum outside a purely magnetic body. Recall that the TE potential here is given by

$$V_{\text{TE}}(\mathbf{r}) := \frac{1}{2} \left[\left(\nabla \log \sqrt{\mu_{\text{r}}} \right)^2 - \nabla^2 \log \sqrt{\mu_{\text{r}}} \right], \tag{22.542}$$

as in Eq. (22.155).

22.11.1.1 Weak-Coupling Limit

These path integrals can be evaluated using the same techniques as in the dielectric-particle-dielectric-surface integrals. As an example, we will consider the weak-coupling limit of a dielectric particle at a planar vacuum–magnetic interface. That is, if

$$\mu_{\text{r}}(\text{r}) = 1 + \chi_{\text{m}}(\mathbf{r}), \tag{22.543}$$

then we will expand to lowest order in χ_{m}. Doing this in the path integrals (22.541), we find

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left(-\frac{3}{2} + \frac{\mathcal{T}}{4} \nabla^2 \right) \left\langle\!\!\left\langle \langle \chi_{\text{m}} \rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \tag{22.544}$$

after renormalizing against vacuum ($\chi_{\text{m}} = 0$). To evaluate the two integral terms, we can read off the dielectric result from Eqs. (22.226) and (22.235) as

$$\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \chi_{\text{m}} \rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} = \frac{\Gamma(D/2)\chi_{\text{m}}}{(D+1)2^{D/2+1} z^D}, \tag{22.545}$$

where the particle-surface distance is z. Using this in the path integrals (22.544),

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0}\left(-\frac{3}{2}\frac{\Gamma(D/2)\chi_{\text{m}}}{(D+1)2^{D/2+1}z^D} + \frac{1}{4}\nabla^2\frac{\Gamma(D/2-1)\chi_{\text{m}}}{(D-1)2^{D/2}z^{D-2}}\right) \\ &= \frac{\hbar c \alpha_0 \chi_{\text{m}}}{4(2\pi)^{D/2}\epsilon_0 z^D}\left(-\frac{3\Gamma(D/2)}{(D+1)2^{D/2+2}} + \frac{\Gamma(D/2-1)(D-2)}{2^{D/2+2}}\right) \\ &= \frac{\hbar c \alpha_0 \chi_{\text{m}}}{4(2\pi)^{D/2}\epsilon_0 z^D}\left(\frac{(2D-1)\Gamma(D/2)}{(D+1)2^{D/2+2}}\right) \\ &= \frac{(2D-1)\Gamma(D/2)\hbar c \alpha_0 \chi_{\text{m}}}{16(D+1)(4\pi)^{D/2}\epsilon_0 z^D}. \end{aligned} \tag{22.546}$$

For $D = 4$, the result is

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) = \frac{3\hbar c \alpha_0}{32\pi^2\epsilon_0 z^4}\left(\frac{7\chi_{\text{m}}}{120}\right). \tag{22.547}$$

This gives a total electromagnetic potential of $(7\chi_{\text{m}}/60)$ times the strong-coupling result, but with an overall *positive* result, so the potential in this case is *repulsive*.[22]

22.11.2 Magnetic Particle, Magnetic Surface

For a purely magnetic atom interacting with a purely magnetic body, we can take $\alpha_0 = 0$ and $\epsilon_{\text{r}} = 1$ in Eqs. (22.195) and (22.196), with the resulting path integrals

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= \frac{\hbar c \beta_0 \mu_0}{4(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \left(\langle\mu_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1} - \frac{\mathcal{T}}{2\mu_{\text{r}}}\nabla^2\right)\langle\mu_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{\text{TE}}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \beta_0 \mu_0}{4(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \langle\mu_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \end{aligned}$$

(unrenormalized Casimir–Polder potential, magnetic particle/magnetic body) (22.548)

where we have set any gradients of μ_{r} at this particle location to zero. Note that these are the same as the dielectric–dielectric path integrals (22.540), with the following changes: $\alpha_0 \longrightarrow \beta_0\mu_0$, $\epsilon_0 \longrightarrow \mu_0$, $\epsilon \longrightarrow \mu$, and the TE and TM integrals have interchanged. Thus, all the calculations for the electric-dipole atom interacting with a dielectric surface, in Sections 22.4–22.10 apply here with the same changes. In particular, for a magnetic response equivalent to the electric response, the total (TM+TE) Casimir–Polder potential in this case is the same as for the dielectric–dielectric case. [23]

22.11.3 Magnetic Particle, Dielectric Surface

For a purely magnetic atom interacting with a dielectric body, we can take $\alpha_0 = 0$ and $\mu_{\text{r}} = 1$ in Eqs. (22.195) and (22.196), to obtain

$$\begin{aligned} V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) &= \frac{\hbar c \beta_0 \mu_0}{4(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \epsilon_{\text{r}}\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2} - \frac{\mathcal{T}}{2}\nabla^2\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ V_{\text{CP}}^{(\text{TM})}(\mathbf{r}) &= \frac{\hbar c \beta_0 \mu_0}{4(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\left\langle\!\!\left\langle \epsilon_{\text{r}}\langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-3/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \end{aligned}$$

(unrenormalized Casimir–Polder potential, magnetic particle/dielectric body) (22.549)

[22] cf. S. Y. Buhmann, H. T. Dung, T. Kampf, and D.-G. Welsch, ''Casimir-Polder interaction of atoms with magnetodielectric bodies,'' *European Physical Journal D* **35**, 15 (2005), Eq. (99) (doi: 10.1140/epjd/e2005-00044-6).

[23] See Timothy H. Boyer, ''Van der Waals forces and zero-point energy for dielectric and permeable materials,'' *Physical Review A* **9**, 2078 (1974) (doi: 10.1103/PhysRevA.9.2078). See also Stefan Yoshi Buhmann, *Dispersion Forces I: Macroscopic Quantum Electrodynamics and Ground-State Casimir, Casimir–Polder and van der Waals Forces* (Springer, 2012). Compare Eq. (4.134) with $\mu = 1$ with Eq. (4.157) with $\epsilon = 1$.

where we have set any gradients of ϵ_r at this particle location to zero. Note that these have the form of the path integrals for a dielectric particle and a magnetic body as in Eqs. (22.541), but with the TE and TM polarizations interchanged, and the electric and magnetic matter functions (and parameters) interchanged. Thus, the weak-coupling result (22.11.1.1) also applies if we replace χ_m by χ and α_0 with $\beta_0\mu_0$, giving a total Casimir–Polder potential from (22.547) of

$$V_{CP}(\mathbf{r}) = V_{CP}^{(TE)}(\mathbf{r}) + V_{CP}^{(TM)}(\mathbf{r}) = \frac{3\hbar c\beta_0\mu_0}{32\pi^2 z^4}\left(\frac{7\chi}{60}\right). \tag{22.550}$$

That is, for positive χ_m and β_0, this configuration also produces a repulsive potential.[24]

We can confirm this intuition by considering a mode summation based on the mode diagrams in Sections 22.1.1 and 22.1.2. Recall that a *dielectric* particle interacting with the TE modes has path integrals given in Eqs. (22.540), and the overall *negative* sign of the (renormalized) potential is set by the Fresnel reflection coefficients, which are negative for a source outside the dielectric. Further, the vector nature of the electric field in the TM case is represented by the ∇^2 term in the TM path integral.

By contrast, for a *magnetic* particle interacting with the same modes, note from the TE diagram in Section 22.1.1 that now the incident and reflected magnetic fields have an *opposite* orientation compared to the electric fields, and they also have a vector character, since the field vectors are not parallel unless the mode is normally incident. Thus, the ∇^2 appears in the TE path integral in Eqs. (22.549) here. The overall minus sign is buried in the path integral, but is reflected in the positive, small-χ_m result (22.550). From the TM diagram in Section 22.1.2, we see that the magnetic field takes on a scalar character, and there is another relative minus sign in the incident vs. reflected fields, compared to the electric-field case. Hence, the ∇^2 does not appear in the TM path integral in Eqs. (22.549) here. However, the TM potential *does* appear, as it maintains the proper TM boundary conditions across the interface (thus implementing the proper reflection coefficient).

22.12 Casimir Potential

As an example of the Casimir interaction between two macroscopic bodies, we will evaluate the path integrals (22.177) for the two polarizations

$$\begin{aligned} E_{EM}^{(TE)} &= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\epsilon_r\mu_r\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{TE}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ E_{EM}^{(TM)} &= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\epsilon_r\mu_r\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{TM}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} . \end{aligned}$$

(Casimir energy, scalar EM) (22.551)

For convenience, we will reproduce the matter potentials (22.155) for the two polarizations here:

$$\begin{aligned} V_{TE}(\mathbf{r}) &:= \frac{1}{2}\left[\left(\nabla\log\sqrt{\mu_r}\right)^2 - \nabla^2\log\sqrt{\mu_r}\right] \\ V_{TM}(\mathbf{r}) &:= \frac{1}{2}\left[\left(\nabla\log\sqrt{\epsilon_r}\right)^2 - \nabla^2\log\sqrt{\epsilon_r}\right]. \end{aligned} \tag{22.552}$$

(matter-induced potentials)

[24] For the equivalence of the dielectric atom–magnetic surface with the magnetic atom–dielectric surface, see Stefan Yoshi Buhmann, *Dispersion Forces I: Macroscopic Quantum Electrodynamics and Ground-State Casimir, Casimir–Polder and van der Waals Forces* (Springer, 2012). In particular, compare Eq. (4.134) with $\epsilon = 1$ with Eq. (4.157) with $\mu = 1$; note also how the factors of ϵ_0 and μ_0 work out in the prefactors. For more about the repulsive nature of a magnetic atom and a conducting surface, see H. Haakh, F. Intravaia, C. Henkel, S. Spagnolo, R. Passante, B. Power, and F. Sols, "Temperature dependence of the magnetic Casimir-Polder interaction," *Physical Review A* **80**, 062905 (2009) (doi: 10.1103/PhysRevA.80.062905). The repulsive force between a dielectric body and a permeable body is discussed in Timothy H. Boyer, "Van der Waals forces and zero-point energy for dielectric and permeable materials," *Physical Review A* **9**, 2078 (1974) (doi: 10.1103/PhysRevA.9.2078). See also the online talk at `http://cnls.lanl.gov/casimir/PresentationsSF/henkel-sfe.pdf`.

Of course, the techniques we have used so far for Casimir–Polder potentials will apply here. The main difference is that here we will need to integrate over the source-point location $\mathbf{x}_0$ of the Brownian bridges.

To simplify the calculation, we will take the case of two identical dielectric interfaces (i.e., a vacuum gap between two dielectric half-spaces) of susceptibility χ, separated by a distance d. Thus, we will set $\mu_{\rm r} = 1$ throughout, which for example sets $V_{\rm TE}(\mathbf{r}) = 0$.

22.12.1 TE Path Integral

For the TE-polarized component, we will then need to evaluate the path integral

$$E_{\rm EM}^{(\rm TE)} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\rm r} \rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)},$$

(Casimir energy, scalar EM, TE polarization) (22.553)

where we have taken $\mu_{\rm r} = 1$. To evaluate this integral, we will follow the procedure of Section 22.7.

22.12.1.1 Feynman–Kac Formula

We will begin by considering the path integral

$$L(\lambda, s; d, d_0) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- sT_{\rm ds}\Big[B_t(t'); d, d_0\Big]\Big) \right\rangle\!\!\right\rangle_{B_t}, \tag{22.554}$$

where as usual, $B_t(t')$ is a Brownian bridge pinned to 0 at time $t' = t$ [i.e., $B_t(t) = 0$], and we have defined the ''double-sojourn time''

$$T_{\rm ds}[B_t; d, d_0] := \int_0^t dt'\, \mathbb{1}_{(-\infty, d_0]\cup[d_0+d,\infty)}\Big[B_t(t')\Big], \tag{22.555}$$

which is the combined time that $B_t(t')$ spends sojourning in either $(-\infty, d_0]$ or $[d_0 + d, \infty)$. Equivalently, the indicator function here counts the amount of time the bridge spends *outside* the ''gap'' $[d, d + d_0]$ (this interval may or may not include the point 0).

Starting with the Feynman–Kac ODE (22.300), we can obtain an expression for $f(0)$, which will yield the path integral (22.554) by solving

$$f''(x) = 2\big[\lambda + V(x)\big] f(x) - 2e^{ikx}, \tag{22.556}$$

where we have taken $g(x) = e^{ikx}$, and $V(x)$ is the (scaled) occupation function

$$V(x) = s\mathbb{1}_{(-\infty, d_0]\cup[d_0+d,\infty)}(x). \tag{22.557}$$

The steady-state solution is given by the Feynman–Kac formula (22.299) with $x = 0$ as

$$f(0) = \int_0^\infty dt\, e^{-\lambda t} \left\langle\!\!\left\langle e^{ikW(t)} \exp\left[- \int_0^t dt'\, V\big[W(t')\big] \right] \right\rangle\!\!\right\rangle, \tag{22.558}$$

which we will later integrate over k to obtain the desired path integral.

Then for $x < d_0$ or $x > d_0 + d$, the ODE is

$$f''(x) = 2(\lambda + s) f(x) - 2e^{ikx}, \tag{22.559}$$

with general solutions

$$f(x) = \alpha_+ e^{\sqrt{2(\lambda+s)}\, x} + \alpha_- e^{-\sqrt{2(\lambda+s)}\, x} + \frac{e^{ikx}}{\lambda + s + k^2/2}, \tag{22.560}$$

for coefficients $\alpha_\pm$ to be determined in each region. On the other hand, for $d_0 < x < d_0 + d$, the ODE is the same, but with $s = 0$. Picking the bounded solutions in each domain, we have the solution

$$f(x) = \begin{cases} Ae^{-\sqrt{2(\lambda+s)}\,x} + \dfrac{e^{ikx}}{\lambda + s + k^2/2} & (x > d_0 + d) \\ Be^{-\sqrt{2\lambda}\,x} + Ce^{\sqrt{2\lambda}\,x} + \dfrac{e^{ikx}}{\lambda + k^2/2} & (d_0 < x < d_0 + d) \\ De^{\sqrt{2(\lambda+s)}\,x} + \dfrac{e^{ikx}}{\lambda + s + k^2/2} & (x < d_0), \end{cases} \tag{22.561}$$

for undetermined constants A, B, C, and D. Demanding continuity of $f(x)$ at $x = d_0 + d$ and $x = d_0$ gives the conditions

$$\begin{aligned} Ae^{-\sqrt{2(\lambda+s)}(d_0+d)} + \frac{e^{ik(d_0+d)}}{\lambda + s + k^2/2} &= Be^{-\sqrt{2\lambda}(d_0+d)} + Ce^{\sqrt{2\lambda}(d_0+d)} + \frac{e^{ik(d_0+d)}}{\lambda + k^2/2} \\ Be^{-\sqrt{2\lambda}\,d} + Ce^{\sqrt{2\lambda}\,d_0} + \frac{e^{ikd_0}}{\lambda + k^2/2} &= De^{\sqrt{2(\lambda+s)}\,d_0} + \frac{e^{ikd_0}}{\lambda + s + k^2/2}, \end{aligned} \tag{22.562}$$

respectively, and continuity of the derivatives gives the conditions

$$\begin{aligned} -\sqrt{2(\lambda+s)}Ae^{-\sqrt{2(\lambda+s)}(d_0+d)} + \frac{ike^{ik(d_0+d)}}{\lambda + s + k^2/2} &= -\sqrt{2\lambda}Be^{-\sqrt{2\lambda}(d_0+d)} + \sqrt{2\lambda}Ce^{\sqrt{2\lambda}(d_0+d)} + \frac{ike^{ik(d_0+d)}}{\lambda + k^2/2} \\ -\sqrt{2\lambda}Be^{-\sqrt{2\lambda}\,d_0} + \sqrt{2\lambda}Ce^{\sqrt{2\lambda}\,d_0} + \frac{ike^{ik(d_0+d)}}{\lambda + k^2/2} &= \sqrt{2(\lambda+s)}De^{\sqrt{2(\lambda+s)}\,d_0} + \frac{ike^{ikd_0}}{\lambda + s + k^2/2}. \end{aligned} \tag{22.563}$$

The solution for the cofficients are fairly complicated, but they simplify somewhat after integration over k:

$$\begin{aligned} \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, A &= \frac{1}{\sqrt{2(\lambda+s)}} \left(\frac{s\sinh\left[d\sqrt{2\lambda}\right] e^{2\sqrt{2(\lambda+s)}\,(d_0+d)}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right] + 2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]} \right) \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, B &= \frac{1}{2\sqrt{2\lambda}} \left(\frac{-se^{\sqrt{2\lambda}\,(2d_0+d)} + \left(2\lambda + s - 2\sqrt{\lambda(\lambda+s)}\right) e^{-\sqrt{2\lambda}\,d}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right] + 2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]} \right) \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, C &= \frac{1}{2\sqrt{2\lambda}} \left(\frac{-se^{-\sqrt{2\lambda}\,(2d_0+d)} + \left(2\lambda + s - 2\sqrt{\lambda(\lambda+s)}\right) e^{-\sqrt{2\lambda}\,d}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right] + 2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]} \right) \\ \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, D &= \frac{1}{\sqrt{2(\lambda+s)}} \left(\frac{s\sinh\left[d\sqrt{2\lambda}\right] e^{-2\sqrt{2(\lambda+s)}\,d_0}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right] + 2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]} \right). \end{aligned} \tag{22.564}$$

Then since

$$f(0) = \begin{cases} A + \dfrac{1}{\lambda + s + k^2/2} & (0 > d_0 + d) \\ B + C + \dfrac{1}{\lambda + k^2/2} & (d_0 < 0 < d_0 + d) \\ D + \dfrac{1}{\lambda + s + k^2/2} & (0 < d_0), \end{cases} \tag{22.565}$$

and we want to calculate

$$L(\lambda, s; d, d_0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, f(0) = \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- s\,\mathbb{1}_{(-\infty,d_0]\cup[d_0+d,\infty)}\Big[B_t(t')\Big]\Big)\right\rangle\!\!\right\rangle_{B_t}, \tag{22.566}$$

and so

$$L(\lambda,s;d,d_0)=\begin{cases}\sqrt{\dfrac{\pi}{\lambda+s}}\left(1+\dfrac{s\sinh\left[d\sqrt{2\lambda}\right]e^{2\sqrt{2(\lambda+s)}\,(d_0+d)}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right]+2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]}\right) & (d_0+d<0)\\ \sqrt{\dfrac{\pi}{\lambda}}\left(1-\dfrac{s\cosh\left[(2d_0+d)\sqrt{2\lambda}\right]-\left(2\lambda+s-2\sqrt{\lambda(\lambda+s)}\right)e^{-\sqrt{2\lambda}\,d}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right]+2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]}\right) & (d_0<0<d_0+d)\\ \sqrt{\dfrac{\pi}{\lambda+s}}\left(1+\dfrac{s\sinh\left[d\sqrt{2\lambda}\right]e^{-2\sqrt{2(\lambda+s)}\,d_0}}{(2\lambda+s)\sinh\left[d\sqrt{2\lambda}\right]+2\sqrt{\lambda(\lambda+s)}\cosh\left[d\sqrt{2\lambda}\right]}\right) & (d_0>0).\end{cases} \tag{22.567}$$

Eliminating the hyperbolic functions in favor of exponentials, we have

$$L(\lambda,s;d,d_0)=\begin{cases}\sqrt{\dfrac{\pi}{\lambda+s}}\left(1+\dfrac{s\left(1-e^{-2\sqrt{2\lambda}\,a}\right)e^{2\sqrt{2(\lambda+s)}\,(d_0+d)}}{\Lambda_+-\Lambda_-e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0+d<0)\\ \sqrt{\dfrac{\pi}{\lambda}}\left(1-\dfrac{s\left(1+e^{-2\sqrt{2\lambda}(2d_0+d)}\right)e^{2\sqrt{2\lambda}\,d_0}-2\Lambda_-e^{-2\sqrt{2\lambda}\,d}}{\Lambda_+-\Lambda_-e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0<0<d_0+d)\\ \sqrt{\dfrac{\pi}{\lambda+s}}\left(1+\dfrac{s\left(1-e^{-2\sqrt{2\lambda}\,d}\right)e^{-2\sqrt{2(\lambda+s)}\,d_0}}{\Lambda_+-\Lambda_-e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0>0),\end{cases} \tag{22.568}$$

where the symbols

$$\begin{aligned}\Lambda_\pm&:=(2\lambda+s)\pm2\sqrt{\lambda(\lambda+s)}\\&=\left(\sqrt{\lambda+s}\pm\sqrt{\lambda}\right)^2\end{aligned} \tag{22.569}$$

encapsulate much of the dependence on λ and s. Recalling the TE reflection coefficient (22.274)

$$r_{\text{TE}}(\lambda;\chi)=\frac{\sqrt{\lambda}-\sqrt{\lambda+\chi}}{\sqrt{\lambda}+\sqrt{\lambda+\chi}}, \tag{22.570}$$

we can see that

$$r_{\text{TE}}^2(\lambda;s)=\frac{\Lambda_-}{\Lambda_+}, \tag{22.571}$$

and

$$\frac{s}{\Lambda_+}=\frac{\left(\sqrt{\lambda+s}+\sqrt{\lambda}\right)\left(\sqrt{\lambda+s}-\sqrt{\lambda}\right)}{\Lambda_+}=-r_{\text{TE}}(\lambda;s). \tag{22.572}$$

Then we can rewrite Eq. (22.568) as

$$L(\lambda,s;d,d_0)=\begin{cases}\sqrt{\dfrac{\pi}{\lambda+s}}\left(1-\dfrac{r_{\text{TE}}(\lambda;s)\left(1-e^{-2\sqrt{2\lambda}\,d}\right)e^{2\sqrt{2(\lambda+s)}\,(d_0+d)}}{1-r_{\text{TE}}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0+d<0)\\ \sqrt{\dfrac{\pi}{\lambda}}\left(1+\dfrac{r_{\text{TE}}(\lambda;s)\left(1+e^{-2\sqrt{2\lambda}(2d_0+d)}\right)e^{2\sqrt{2\lambda}\,d_0}+2r_{\text{TE}}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}}{1-r_{\text{TE}}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0<0<d_0+d)\\ \sqrt{\dfrac{\pi}{\lambda+s}}\left(1-\dfrac{r_{\text{TE}}(\lambda;s)\left(1-e^{-2\sqrt{2\lambda}\,d}\right)e^{-2\sqrt{2(\lambda+s)}\,d_0}}{1-r_{\text{TE}}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}}\right) & (d_0>0),\end{cases} \tag{22.573}$$

now in terms of Fresnel coefficients.

22.12.1.2 Mellin Transform

Our next goal here is to compute the integral

$$\tilde{I}_D(s,\chi;d,d_0) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle e^{-s\{\mathcal{T}+\chi T_{\rm ds}[B_\mathcal{T}]\}} - e^{-s\{\mathcal{T}+\chi\mathcal{T}\mathbb{1}_{(-\infty,d_0]\cup[d_0+d,\infty)}[0]\}} \right\rangle\!\!\right\rangle, \tag{22.574}$$

where the second term is a renormalization against separated plates, which in each region means considering the case where the interfaces are displaced far away from the point of consideration (i.e., the source point of the Brownian bridges). We will proceed in analogy to Section 22.7.2. First, we will make the replacements $s \longrightarrow s\chi$ and $\lambda \longrightarrow \lambda + s$, which from Eq. (22.566) means we have now computed

$$L(\lambda+s,s\chi;d,d_0) = \int_0^\infty \frac{d\mathcal{T}}{\sqrt{\mathcal{T}}}\, e^{-\lambda\mathcal{T}} \left\langle\!\!\left\langle e^{-s\{\mathcal{T}+\chi T_{\rm ds}[B_\mathcal{T};d,d_0]\}} \right\rangle\!\!\right\rangle_{B_\mathcal{T}}. \tag{22.575}$$

Note that we must compare this to the reference case

$$L_0(\lambda+s,s\chi;d,d_0) = \int_0^\infty \frac{d\mathcal{T}}{\sqrt{\mathcal{T}}}\, e^{-\lambda\mathcal{T}} \left\langle\!\!\left\langle e^{-s\{\mathcal{T}+\chi\mathcal{T}\mathbb{1}_{(-\infty,d_0]\cup[d_0+d,\infty)}[0]\}} \right\rangle\!\!\right\rangle_{B_\mathcal{T}}, \tag{22.576}$$

which again corresponds to taking the limit as the interfaces move far away from the source point (the origin) in $L(\lambda+s,s\chi;d,d_0)$. In each region this corresponds to the limit where all of the exponential factors that depend on d or d_0 vanish. This allows us to drop the leading terms in Eq. (22.568):

$$\begin{aligned} L_-(\lambda,s;d,d_0) &:= L(\lambda,s;d,d_0) - \sqrt{\frac{\pi}{\lambda+s[\Theta(d)+\Theta(-d_0-d)]}} \\ &= \begin{cases} -\sqrt{\dfrac{\pi}{\lambda+s}} \left(\dfrac{r_{\rm TE}(\lambda;s)\left(1-e^{-2\sqrt{2\lambda}\,d}\right)e^{2\sqrt{2(\lambda+s)}\,(d_0+d)}}{1-r_{\rm TE}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}} \right) & (d_0+d<0) \\ \sqrt{\dfrac{\pi}{\lambda}} \left(\dfrac{r_{\rm TE}(\lambda;s)\left(1+e^{-2\sqrt{2\lambda}(2d_0+d)}\right)e^{2\sqrt{2\lambda}\,d_0} + 2r_{\rm TE}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}}{1-r_{\rm TE}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}} \right) & (d_0<0<d_0+d) \\ -\sqrt{\dfrac{\pi}{\lambda+s}} \left(\dfrac{r_{\rm TE}(\lambda;s)\left(1-e^{-2\sqrt{2\lambda}\,d}\right)e^{-2\sqrt{2(\lambda+s)}\,d_0}}{1-r_{\rm TE}^2(\lambda;s)\,e^{-2\sqrt{2\lambda}\,d}} \right) & (d_0>0). \end{cases} \end{aligned} \tag{22.577}$$

Then, as in Eq. (22.321), we can compute the integral (22.574) as

$$\tilde{I}_D(s,\chi;d,d_0) = \frac{1}{\Gamma\big[(D+1)/2\big]} \int_0^\infty d\lambda\, \lambda^{(D-1)/2} L_-(\lambda+s,s\chi;d,d_0), \tag{22.578}$$

which for now we will leave in terms of the cumbersome expression (22.577).

22.12.1.3 Inverse Moments

In analogy to Eq. (22.324), we will now compute the integral

$$\mathcal{I}_{D,\alpha}(d,d_0) = \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2-\alpha}} \left\langle\!\!\left\langle \frac{1}{(\mathcal{T}+\chi T_{\rm ds}[B_\mathcal{T};d,d_0])^\alpha} - \frac{1}{\big(\mathcal{T}+\chi\mathcal{T}\mathbb{1}_{(-\infty,d]\cup[d+a,\infty)}\big[0\big]\big)^\alpha} \right\rangle\!\!\right\rangle, \tag{22.579}$$

using the transformation (22.326):

$$\mathcal{I}_{D,\alpha}(d,d_0) = \frac{1}{\Gamma(\alpha)} \int_0^\infty ds\, s^{\alpha-1} \tilde{I}_{D-2\alpha}(d,d_0). \tag{22.580}$$

With Eq. (22.578), this becomes

$$\mathcal{I}_{D,\alpha}(d,d_0) = \frac{1}{\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\, s^{\alpha-1}\int_0^\infty d\lambda\, \lambda^{(D-1)/2-\alpha} L_-(\lambda+s, s\chi; d, d_0). \tag{22.581}$$

We then carry out the other transformations in Section 22.7.3. First letting $\lambda \longrightarrow \lambda - s$ and then $\lambda \longrightarrow \lambda s$,

$$\mathcal{I}_{D,\alpha}(d,d_0) = \frac{1}{\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\, s^{D/2-1}\int_1^\infty d\lambda\, (\lambda-1)^{(D-1)/2-\alpha}\Big[\sqrt{s}L_-(\lambda s, s\chi; d, d_0)\Big]. \tag{22.582}$$

Then letting $s \longrightarrow s/8$,

$$\begin{aligned}\mathcal{I}_{D,\alpha}(d,d_0) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\, s^{D/2-1}\int_1^\infty d\lambda\, (\lambda-1)^{(D-1)/2-\alpha} \\ \times\Big[\sqrt{(s/8\pi)}\, L_-(\lambda s/8, s\chi/8; d, d_0)\Big].\end{aligned} \tag{22.583}$$

The last factor is

$$\begin{aligned}&\Big[\sqrt{(s/8\pi)}\, L_-(\lambda s/8, s\chi/8; d, d_0)\Big] \\ &\quad= \begin{cases} -\dfrac{1}{\sqrt{\lambda+\chi}}\left(\dfrac{r_{\text{TE}}(\lambda;\chi)\left(1-e^{-\sqrt{s\lambda}\,d}\right)e^{\sqrt{s(\lambda+\chi)}\,(d_0+d)}}{1-r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right) & (d_0+d<0) \\ \dfrac{1}{\sqrt{\lambda}}\left(\dfrac{r_{\text{TE}}(\lambda;\chi)\left(1+e^{-\sqrt{s\lambda}(2d_0+d)}\right)e^{\sqrt{s\lambda}\,d_0}+2r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}{1-r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right) & (d_0<0<d_0+d) \\ -\dfrac{1}{\sqrt{\lambda+\chi}}\left(\dfrac{r_{\text{TE}}(\lambda;\chi)\left(1-e^{-\sqrt{s\lambda}\,d}\right)e^{-\sqrt{s(\lambda+\chi)}\,d_0}}{1-r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right) & (d_0>0), \end{cases}\end{aligned} \tag{22.584}$$

by making the replacements in Eq. (22.577).

22.12.1.4 Spatial Integration and Renormalization

The next step is to compute the integral of the function $\mathcal{I}_{D,\alpha}(d,d_0)$ in Eq. (22.583) over all d_0, which corresponds to the component of the x_0 integral normal to the interfaces. All the d_0-dependent components are in the factor in Eq. (22.584), so we will focus on this last part. We will call the integrals over the respective regions of the three piecewise parts I_{I}, I_{II}, and I_{III}, in the order the components are listed in Eq. (22.584). Beginning with the third section,

$$\begin{aligned}I_{\text{III}} &= -\frac{r_{\text{TE}}(\lambda;\chi)\left(1-e^{-\sqrt{s\lambda}\,d}\right)}{\sqrt{\lambda+\chi}\left[1-r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\int_0^\infty dd_0\, e^{-\sqrt{s(\lambda+\chi)}\,d_0} \\ &= -\frac{r_{\text{TE}}(\lambda;\chi)\left(1-e^{-\sqrt{s\lambda}\,d}\right)}{\sqrt{s}(\lambda+\chi)\left[1-r_{\text{TE}}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}.\end{aligned} \tag{22.585}$$

By symmetry, the first section is equivalent:

$$I_{\text{I}} = I_{\text{III}}. \tag{22.586}$$

Finally, in the middle region we have

$$\begin{aligned} I_{\rm II} &= \frac{r_{\rm TE}(\lambda;\chi)}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\int_{-d}^{0}dd_0\left[\left(1+e^{-\sqrt{s\lambda}(2d_0+d)}\right)e^{\sqrt{s\lambda}\,d_0}+2r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right] \\ &= \frac{r_{\rm TE}(\lambda;\chi)}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\int_{-d}^{0}dd_0\left[\left(e^{\sqrt{s\lambda}\,d_0}+e^{-\sqrt{s\lambda}(d_0+d)}\right)+2r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right] \\ &= \frac{r_{\rm TE}(\lambda;\chi)}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\frac{2}{\sqrt{s\lambda}}\left(1-e^{-\sqrt{s\lambda}\,d}\right)+2d\,r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right] \\ &= \frac{2r_{\rm TE}(\lambda;\chi)}{\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\frac{1}{\sqrt{s}\,\lambda}\left(1-e^{-\sqrt{s\lambda}\,d}\right)+\frac{d}{\sqrt{\lambda}}\,r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]. \end{aligned} \tag{22.587}$$

The sum over all three regions is therefore

$$\begin{aligned} I_{\rm I}+I_{\rm II}+I_{\rm III} &= \frac{2r_{\rm TE}(\lambda;\chi)}{\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\left(\frac{1}{\sqrt{s}\,\lambda}-\frac{1}{\sqrt{s}(\lambda+\chi)}\right)\left(1-e^{-\sqrt{s\lambda}\,d}\right)+\frac{d}{\sqrt{\lambda}}\,r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right] \\ &= \frac{2r_{\rm TE}(\lambda;\chi)}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\sqrt{\frac{\lambda}{s}}\left(\frac{1}{\lambda}-\frac{1}{\lambda+\chi}\right)\left(1-e^{-\sqrt{s\lambda}\,d}\right)+d\,r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]. \end{aligned} \tag{22.588}$$

Anticipating the final result, note that we normally define the Casimir energy (22.553) to be zero when the separation between objects becomes large. Note, though, that Eq. (22.588), which encapsulates all the spatial dependence of the energy (22.553), does *not* have this property, since it does not vanish as $d\longrightarrow\infty$. The renormalization is only necessary for the first term, and has the form

$$\frac{1-e^{-\sqrt{s\lambda}\,d}}{1-r_{\rm TE}^2\,e^{-\sqrt{s\lambda}\,d}}-1=\frac{\left[r_{\rm TE}^2-1\right]e^{-\sqrt{s\lambda}\,d}}{1-r_{\rm TE}^2\,e^{-\sqrt{s\lambda}\,d}}, \tag{22.589}$$

Thus, after renormalization, Eq. (22.588) becomes

$$\begin{aligned} &(I_{\rm I}+I_{\rm II}+I_{\rm III})_{\rm renorm} \\ &\quad= \frac{2r_{\rm TE}(\lambda;\chi)}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\sqrt{\frac{\lambda}{s}}\left(\frac{1}{\lambda}-\frac{1}{\lambda+\chi}\right)\left[r_{\rm TE}^2(\lambda;\chi)-1\right]e^{-\sqrt{s\lambda}\,d}+d\,r_{\rm TE}(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]. \end{aligned} \tag{22.590}$$

We can simplify this further by using

$$\begin{aligned} \frac{\sqrt{\lambda}[r_{\rm TE}^2(\lambda;\chi)-1]}{r_{\rm TE}(\lambda;\chi)}\left(\frac{1}{\lambda}-\frac{1}{\lambda+\chi}\right) &= \left[r_{\rm TE}(\lambda;\chi)-r_{\rm TE}^{-1}(\lambda;\chi)\right]\left(\frac{\chi}{\sqrt{\lambda}(\lambda+\chi)}\right) \\ &= \left[\frac{\sqrt{\lambda}-\sqrt{\lambda+\chi}}{\sqrt{\lambda}+\sqrt{\lambda+\chi}}-\frac{\sqrt{\lambda}+\sqrt{\lambda+\chi}}{\sqrt{\lambda}-\sqrt{\lambda+\chi}}\right]\left(\frac{\chi}{\sqrt{\lambda}(\lambda+\chi)}\right) \\ &= \left[\frac{-4\sqrt{\lambda(\lambda+\chi)}}{-\chi}\right]\left(\frac{\chi}{\sqrt{\lambda}(\lambda+\chi)}\right) \\ &= \frac{4}{\sqrt{\lambda+\chi}}, \end{aligned} \tag{22.591}$$

where we used the explicit form (22.570) for the reflection coefficient. Thus, we have

$$(I_{\rm I}+I_{\rm II}+I_{\rm III})_{\rm renorm}=\frac{2r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}{\sqrt{\lambda}\left[1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}\right]}\left[\frac{4}{\sqrt{s(\lambda+\chi)}}+d\right] \tag{22.592}$$

as the simplified form of Eq. (22.590). Then we can write out the original result that we were trying to calculate

$$\int_{-\infty}^{\infty} dd_0\, \mathcal{I}_{D,\alpha}(d,d_0) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]} \int_0^\infty ds\, s^{D/2-1} \int_1^\infty d\lambda\, (\lambda-1)^{(D-1)/2-\alpha} \times \frac{2r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}}{\sqrt{\lambda}\left[1-r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right]} \left[d + \frac{4}{\sqrt{s(\lambda+\chi)}}\right] \tag{22.593}$$

from Eq. (22.583), again after renormalization.

22.12.1.5 Casimir Energy

Now we will assemble all the parts to obtain the Casimir energy. We started with the path-integral expression (22.553), which we can write as

$$E_{\text{EM}}^{(\text{TE})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int d^{D-2}x_0 \int dd_0\, \mathcal{I}_{D,1/2}(d,d_0), \tag{22.594}$$

in terms of the integral (22.579). Note that we have split the original x_0 integral into the d_0 integration normal to the interfaces, and the rest of the $D-2$ integrals, where the integrand is constant. Thus, we may define the (divergent) $(D-2)$-dimensional "area"

$$A := \int d^{D-2}x_0, \tag{22.595}$$

and then write the Casimir energy density as

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_{-\infty}^{\infty} dd_0\, \mathcal{I}_{D,1/2}(d,d_0). \tag{22.596}$$

Then using Eq. (22.593) with $\alpha = 1/2$, we have

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = -\frac{\hbar c}{(16\pi)^{D/2}\Gamma(D/2)} \int_0^\infty ds\, s^{D/2-1} \int_1^\infty d\lambda\, \frac{(\lambda-1)^{D/2-1}}{\sqrt{\lambda}} \left[\frac{r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}}{1-r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}}\right] \left[d + \frac{4}{\sqrt{s(\lambda+\chi)}}\right]. \tag{22.597}$$

which has already been renormalized (shifted) such that this energy vanishes as $d \longrightarrow \infty$. Then using

$$\partial_\lambda r_{\text{TE}}(\lambda;\chi) = \partial_\lambda \frac{\sqrt{\lambda}-\sqrt{\lambda+\chi}}{\sqrt{\lambda}+\sqrt{\lambda+\chi}} = -\frac{r_{\text{TE}}}{\sqrt{\lambda(\lambda+\chi)}} \tag{22.598}$$

and

$$\begin{aligned} \partial_\lambda \left[r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right] &= 2r_{\text{TE}}(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d} \partial_\lambda r_{\text{TE}}(\lambda;\chi) - \frac{d}{2}\sqrt{\frac{s}{\lambda}}\, r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d} \\ &= -\sqrt{\frac{s}{\lambda}}\, \frac{r_{\text{TE}}^2(\lambda;\chi)}{2}\, e^{-\sqrt{s\lambda}\, d} \left(d + \frac{4}{\sqrt{s(\lambda+\chi)}}\right), \end{aligned} \tag{22.599}$$

Eq. (22.597) becomes

$$\begin{aligned} \frac{E_{\text{EM}}^{(\text{TE})}}{A} &= \frac{2\hbar c}{(16\pi)^{D/2}\Gamma(D/2)} \int_0^\infty ds\, s^{D/2-3/2} \int_1^\infty d\lambda\, (\lambda-1)^{D/2-1} \frac{\partial_\lambda \left[r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right]}{1-r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}} \\ &= -\frac{2\hbar c}{(16\pi)^{D/2}\Gamma(D/2)} \int_0^\infty ds\, s^{D/2-3/2} \int_1^\infty d\lambda\, (\lambda-1)^{D/2-1} \partial_\lambda \log\left[1-r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right]. \end{aligned} \tag{22.600}$$

Integrating by parts, we then have

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = \frac{(D-2)\hbar c}{(16\pi)^{D/2}\Gamma(D/2)} \int_0^\infty ds\, s^{D/2-3/2} \int_1^\infty d\lambda\, (\lambda-1)^{D/2-2} \log\left[1 - r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right].$$

(TE dielectric Casimir energy) (22.601)

Note that we can rescale $s \longrightarrow s/d^2$ to separate out the distance dependence as

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = \frac{(D-2)\hbar c}{(16\pi)^{D/2}\Gamma(D/2) d^{D-1}} \int_0^\infty ds\, s^{D/2-3/2} \int_1^\infty d\lambda\, (\lambda-1)^{D/2-2} \log\left[1 - r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}}\right].$$

(TE dielectric Casimir energy) (22.602)

In the general, dispersive case, however, the reflection coefficient also depends on s, and this rescaling can't be done (indicating a crossover between different power-law potentials). Often these results are differentiated with respect to d to give a Casimir force (pressure).

For 3D electromagnetism ($D = 4$), this becomes

$$\begin{aligned}\frac{E_{\text{EM}}^{(\text{TE})}}{A} &= \frac{\hbar c}{128\pi^2} \int_0^\infty ds\, \sqrt{s} \int_1^\infty d\lambda\, \log\left[1 - r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}\, d}\right] \\ &= \frac{\hbar c}{128\pi^2 d^3} \int_0^\infty ds\, \sqrt{s} \int_1^\infty d\lambda\, \log\left[1 - r_{\text{TE}}^2(\lambda;\chi)\, e^{-\sqrt{s\lambda}}\right].\end{aligned}$$

(TE dielectric Casimir energy, 3D) (22.603)

To check the obvious special cases, for strong coupling ($\chi \longrightarrow \infty$), we have $r_{\text{TE}} = -1$ and thus

$$\begin{aligned}\frac{E_{\text{EM}}^{(\text{TE})}}{A} &= \frac{\hbar c}{128\pi^2 d^3} \int_0^\infty ds\, \sqrt{s} \int_1^\infty d\lambda\, \log\left[1 - e^{-\sqrt{s\lambda}}\right] \\ &= \frac{\hbar c}{128\pi^2 d^3} \left(-\frac{4\pi^4}{45}\right),\end{aligned} \tag{22.604}$$

and thus we have the strong-coupling result[25]

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = -\frac{\pi^2 \hbar c}{1440 d^3}.$$

(TE dielectric Casimir energy, 3D strong coupling) (22.605)

For small χ, we have $r_{\text{TE}} = -\chi/4\lambda$, and using this in Eq. (22.603) after expanding the logarithm,

$$\begin{aligned}\frac{E_{\text{EM}}^{(\text{TE})}}{A} &= -\frac{\hbar c \chi^2}{2048\pi^2 d^3} \int_0^\infty ds\, \sqrt{s} \int_1^\infty \frac{d\lambda}{\lambda^2}\, e^{-\sqrt{s\lambda}} \\ &= -\frac{\hbar c \chi^2}{2048\pi^2 d^3} \left(\frac{8}{5}\right),\end{aligned} \tag{22.606}$$

and so the weak-coupling result is[26]

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = -\frac{\hbar c \chi^2}{1280\pi^2 d^3}.$$

(TE dielectric Casimir energy, 3D weak coupling) (22.607)

Note that this goes as χ^2 to lowest order, with one factor of χ to leading order for each of the two surfaces in the interaction. The $O(\chi)$ terms—corresponding to the one-body energies—were lost in the renormalization when we subtracted away the $d \longrightarrow \infty$ limit.

[25] See, e.g., Julian Schwinger, "Casimir energy for dielectrics," *Proceedings of the National Academy of Sciences* **89**, 4091 (1992) (doi: 10.1073/pnas.89.9.4091).

[26] Julian Schwinger, *op. cit.*

22.12.1.6 Small-χ Limit

The above calculation was fairly cumbersome, so to get a more intuitive look, we will consider the dilute-medium limit. In the limit of small χ, the path integral (22.553) becomes

$$E_{\text{EM}}^{(\text{TE})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle 1 - \frac{1}{2}\langle\chi\rangle_{\mathbf{x}(\tau)} + \frac{3}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}^2 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.608}$$

where we have expanded to *second* order in χ. Again, this is because we are interested in an *interaction* energy between two bodies. This is perhaps more clear if we label the susceptibilities of the two bodies as χ_1 and χ_2; then the term we seek is of the form $\chi_1\chi_2$, while the $O(\chi_1)$ and $O(\chi_2)$ terms are one-body energies that drop out in the renormalization. Thus, dropping the (obvious) background and one-body contributions, we have the weak-coupling result

$$E_{\text{EM}}^{(\text{TE})} = -\frac{3\hbar c}{16(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\chi\rangle_{\mathbf{x}(\tau)}^2 \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}.$$

(Casimir energy, scalar EM, TE polarization, small χ) (22.609)

The factor of 3/8 here is similar to the analogous factor of 3/4 that we obtained in the Casimir–Polder potential: the first factor of 1/2 from the functional derivative in Eq. (22.182), and the remaining 3/2 from making the small-χ expansion as in Eq. (22.226).

In fact, in evaluating the path integral, it is best to emphasize the two-body interaction by letting $\chi = \chi_1 + \chi_2$, and again drop the one-body contributions of the form χ_1^2 and χ_2^2 to obtain

$$E_{\text{EM}}^{(\text{TE})} = -\frac{3\hbar c}{8(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)} \langle\chi_2\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}.$$

(Casimir energy, scalar EM, TE polarization, small χ) (22.610)

Note that since there are two terms in χ^2 corresponding to the two-body interaction, we obtain the factor of 2 that brings this path integral in line with the analogous Casimir–Polder path integral (22.226). Now this path integral is linear in both χ_1 and χ_2 separately. Thus, we can take a Green-function approach, letting

$$\chi_2(\mathbf{r}) = \chi_2 \int_{\chi_2} d^{D-1}r'\, \delta(\mathbf{r} - \mathbf{r}'), \tag{22.611}$$

where we assume χ_2 models a dielectric body that is uniform over some region in space, and the integral extends over this region. Then the path integral (22.610) becomes

$$E_{\text{EM}}^{(\text{TE})} = -\frac{3\hbar c\chi_2}{8(2\pi)^{D/2}} \int_{\chi_2} d^{D-1}r' \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)} \langle\delta(\mathbf{x} - \mathbf{r}')\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}. \tag{22.612}$$

This has precisely the same form as the Casimir–Polder path integral (22.206), once expanded to lowest order in χ, where χ refers to χ_1, and χ_2 is equivalent to α_0/ϵ_0. Thus, the delta function is removed in the same way as in the Casimir–Polder calculation, and we have

$$\frac{E_{\text{EM}}^{(\text{TE})}}{A} = -\frac{3\hbar c\chi_2}{8(2\pi)^{D/2}} \int_d^\infty dz \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.613}$$

where we are taking χ_1 to correspond to the region $z < 0$, and χ_2 to $z > d$ for the two-plane interaction, and we have already divided through by the divergent transverse-integral factor to give the energy per unit area. Then we can set $D = 4$ and adapt the weak-coupling Casimir–Polder result by comparing Eqs. (22.226) and (22.236) to write

$$\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^3} \left\langle\!\!\left\langle \langle\chi\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} = \frac{\chi}{40z^4}. \tag{22.614}$$

Thus, Eq. (22.613) becomes

$$\frac{E_{\rm EM}^{(\rm TE)}}{A} = -\frac{3\hbar c\chi_1\chi_2}{1280\pi^2}\int_d^\infty \frac{dz}{z^4} = -\frac{\hbar c\chi_1\chi_2}{1280\pi^2 d^3}. \tag{22.615}$$

Finally setting $\chi_1 = \chi_2 = \chi$, we see that we have obtained the small-χ limit (22.607) of our earlier calculation.

22.12.1.7 Casimir–Polder Potential: Two Parallel Planes

As a by-product of the Casimir-energy analysis above, we can also immediately obtain the Casimir–Polder potential for an atom in the vicinity of the same two dielectric planes. For example, to obtain the potential *between* two dielectric planes, we can start with Eqs. (22.583) and (22.584) in the middle region to write the integral

$$\mathcal{I}_{D,\alpha}(d,d_0) = \frac{\sqrt{\pi}}{2^{3D/2}\Gamma(\alpha)\Gamma\big[(D+1)/2-\alpha\big]}\int_0^\infty ds\, s^{D/2-1}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{(D-1)/2-\alpha}}{\sqrt{\lambda}} \times\left(\frac{r_{\rm TE}(\lambda;\chi)\left(1+e^{-\sqrt{s\lambda}(2d_0+d)}\right)e^{\sqrt{s\lambda}\,d_0}+2r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}{1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right). \tag{22.616}$$

Then using Eq. (22.332),

$$V_{\rm CP}^{(\rm TE)}(z) = \frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\mathcal{I}_{D,3/2}(z), \tag{22.617}$$

we can write-the Casimir–Polder potential as

$$V_{\rm CP}^{(\rm TE)}(z) = \frac{\hbar c\alpha_0}{2^{3D/2+1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0}\int_0^\infty ds\, s^{D/2-1}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}} \times\left(\frac{r_{\rm TE}(\lambda;\chi)\left(e^{\sqrt{s\lambda}\,d_0}+e^{-\sqrt{s\lambda}(d_0+d)}\right)+2r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}{1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right). \tag{22.618}$$

Note here that $d > 0$ and $d_0 < 0$. Thus, it is more natural to set $z = -d_0 > 0$ as the atomic distance to the left-hand interface, with d the interface separation, and $0 < z < d$:

$$V_{\rm CP}^{(\rm TE)}(z) = \frac{\hbar c\alpha_0}{2^{3D/2+1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0} \times\int_0^\infty ds\, s^{D/2-1}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}}\left(\frac{r_{\rm TE}(\lambda;\chi)\left(e^{-\sqrt{s\lambda}\,z}+e^{-\sqrt{s\lambda}(d-z)}\right)}{1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right). \tag{22.619}$$

Note that we have dropped the $r_{\rm TE}^2$ term in the numerator of the last factor, which is independent of the atomic position z, and only contributes to the Casimir energy between the interfaces. Also, notice that in the limit $d \longrightarrow \infty$, only the first term in the numerator of the last factor stays, and we can do the integral over s; the result is just the one-interface energy (22.336).

In general the integrals here are difficult to carry out, but they simplify in the case of perfectly conducting walls, where $r_{\rm TE}(\lambda;\chi) = -1$. Then Eq. (22.619) becomes

$$V_{\rm CP}^{(\rm TE)}(z) = -\frac{\hbar c\alpha_0}{2^{3D/2+1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}}\int_0^\infty ds\, s^{D/2-1}\left(\frac{e^{-\sqrt{s\lambda}\,z}+e^{-\sqrt{s\lambda}(d-z)}}{1-e^{-\sqrt{s\lambda}\,d}}\right), \tag{22.620}$$

where we have changed the order of integration. Carrying out the s integration,

$$V_{\rm CP}^{(\rm TE)}(z) = -\frac{\hbar c\alpha_0}{2^{3D/2+1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}}\left[\frac{2\Gamma(D)}{d^D\lambda^{D/2}}\Big(\zeta(D,z/d)+\zeta(D,1-z/d)\Big)\right], \tag{22.621}$$

where $\zeta(z,a)$ is the generalized zeta function. Cleaning this up, we have the result

$$V_{\rm CP}^{\rm (TE)}(z) = -\frac{(D-2)\Gamma[(D+1)/2]\hbar c\alpha_0}{4(4\pi)^{D/2}\sqrt{\pi}\epsilon_0 d^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-(D+1)/2}\Big[\zeta(D,z/d)+\zeta(D,1-z/d)\Big], \quad (22.622)$$

where we used Eq. (22.335) to simplify the gamma functions. Carrying out the remaining integral,

$$V_{\rm CP}^{\rm (TE)}(z) = -\frac{(D-2)\Gamma(D/2-1)\hbar c\alpha_0}{8(4\pi)^{D/2}\epsilon_0 d^D}\Big[\zeta(D,z/d)+\zeta(D,1-z/d)\Big].$$

(TE Casimir–Polder potential between dielectric planes) (22.623)

Now for the 3D case, we take $D=4$:

$$V_{\rm CP}^{\rm (TE)}(z) = -\frac{\hbar c\alpha_0}{64\pi^2\epsilon_0 d^4}\Big[\zeta(4,z/d)+\zeta(4,1-z/d)\Big]. \quad (22.624)$$

The generalized zeta functionis related to the polygamma function $\psi^{(n)}(z)$ by

$$\psi^{(n)}(z) = (-1)^{n+1}n!\,\zeta(n+1,z), \quad (22.625)$$

or in the present case,

$$\psi^{(3)}(z) = 6\,\zeta(4,z). \quad (22.626)$$

Thus,

$$V_{\rm CP}^{\rm (TE)}(z) = -\frac{\hbar c\alpha_0}{384\pi^2\epsilon_0 d^4}\Big[\psi^{(3)}(z/d)+\psi^{(3)}(1-z/d)\Big]. \quad (22.627)$$

Finally, using the formula[27]

$$\psi^{(n)}(1-z)+(-1)^{n+1}\psi^{(n)}(z) = (-1)^n\pi\partial_z^n\cot\pi z, \quad (22.628)$$

or specifically,

$$\psi^{(3)}(z)+\psi^{(3)}(1-z) = -\pi\partial_z^3\cot\pi z = 2\pi^4\frac{(2+\cos 2\pi z)}{\sin^4\pi z}, \quad (22.629)$$

we have

$$V_{\rm CP}^{\rm (TE)}(z) = -\frac{\pi^2\hbar c\alpha_0}{192\epsilon_0 d^4}\left[\frac{2+\cos(2\pi z/d)}{\sin^4(\pi z/d)}\right].$$

(TE Casimir–Polder potential between dielectric planes, 3D) (22.630)

Note that for small z/d, the last factor reduces to $3(d/\pi z)^4$, and then this result reduces to the one-interface, strong-coupling result (22.223).

Of course, we can also include the z-independent "Lamb-shift" component (the final term scaling as $r_{\rm TE}^2$) from (22.618),

$$V_{\rm L}^{\rm (TE)}(z) = \frac{\hbar c\alpha_0}{2^{3D/2+1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0}\int_0^\infty ds\,s^{D/2-1}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}}\left(\frac{2r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}{1-r_{\rm TE}^2(\lambda;\chi)\,e^{-\sqrt{s\lambda}\,d}}\right). \quad (22.631)$$

In the perfect-conductor limit, we have

$$V_{\rm L}^{\rm (TE)}(z) = \frac{\hbar c\alpha_0}{2^{3D/2}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0}\int_1^\infty d\lambda\,\frac{(\lambda-1)^{D/2-2}}{\sqrt{\lambda}}\int_0^\infty ds\,s^{D/2-1}\left(\frac{e^{-\sqrt{s\lambda}\,d}}{1-e^{-\sqrt{s\lambda}\,d}}\right). \quad (22.632)$$

The s integral gives the factor $2d^{-D}\lambda^{-D/2}\Gamma[D]\zeta[D]$,

$$V_{\rm L}^{\rm (TE)}(z) = \frac{\hbar c\alpha_0\Gamma[D]\zeta[D]}{2^{3D/2-1}(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0 d^D}\int_1^\infty d\lambda\,(\lambda-1)^{D/2-2}\lambda^{-D/2-1}, \quad (22.633)$$

[27] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions* (Dover, 1965), p. 260, Eq. (6.4.7).

and the λ integral gives a factor $4/D(D-2)$:

$$V_{\text{L}}^{(\text{TE})}(z) = \frac{\hbar c \alpha_0 \Gamma[D-1]\zeta[D]}{2^{3D/2-3}(D-2)(2\pi)^{D/2}\Gamma(D/2-1)\epsilon_0 d^D}. \tag{22.634}$$

For $D = 4$, we have

$$V_{\text{L}}^{(\text{TE})}(z) = \frac{\pi^2 \hbar c \alpha_0}{2880\epsilon_0 d^4}, \tag{22.635}$$

and thus Eq. (22.630) becomes

$$V_{\text{CP}}^{(\text{TE})}(z) = \frac{\pi^2 \hbar c \alpha_0}{192\epsilon_0 d^4}\left[\frac{1}{15} - \frac{2+\cos(2\pi z/d)}{\sin^4(\pi z/d)}\right], \tag{22.636}$$

(TE Casimir–Polder potential between dielectric planes, 3D)

including the z-independent potential shift between the plates.

22.12.2 TM Path Integral

For the TM polarization, we will return to Eq. (22.551), which is

$$E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\epsilon_{\text{r}}\rangle_{\mathbf{x}(\tau)}^{-1/2}\, e^{-\mathcal{T}\langle V_{\text{TM}}\rangle_{\mathbf{x}(\tau)}} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.637}$$

(Casimir energy, scalar EM)

for the pure-dielectric case ($\mu_{\text{r}} = 1$), and where the TM potential is

$$V_{\text{TM}}(\mathbf{r}) := \frac{1}{2}\left[\left(\partial_z \log\sqrt{\epsilon_{\text{r}}}\right)^2 - \partial_z^2 \log\sqrt{\epsilon_{\text{r}}}\right], \tag{22.638}$$

(matter-induced potentials)

if we assume the only variation in the dielectric occurs along the z-direction. The calculation here is analogous to the Casimir–Polder calculation of Section 22.9.

22.12.2.1 Small-χ Expansion

We will start with a weak-coupling expansion analogous to the TE calculation in Section 22.12.1.6 and the TM Casimir–Polder calculation of Section 22.4.5. Again, we will expand the energy to second order in χ, with $\epsilon_{\text{r}} = 1 + \chi$. Starting with the potential,

$$V_{\text{TM}}(\mathbf{r}) = \frac{1}{8}\left[\left(\partial_z\chi\right)^2 - 2\partial_z^2\chi + \partial_z^2\chi^2\right] + O(\chi^3), \tag{22.639}$$

the energy (22.551) becomes

$$\begin{aligned}
E_{\text{EM}}^{(\text{TM})} &= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \left(1 - \frac{1}{2}\langle\chi\rangle_{\mathbf{x}(\tau)} + \frac{3}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}^2\right)\right.\right. \\
&\qquad \left.\left. \times\left(1 + \frac{\mathcal{T}}{8}\left[\left\langle(\partial_z\chi)^2\right\rangle_{\mathbf{x}(\tau)} - 2\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)} + \partial_z^2\langle\chi^2\rangle_{\mathbf{x}(\tau)}\right] + \frac{\mathcal{T}^2}{32}\left[\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)}\right]^2\right)\right\rangle\!\!\right\rangle \\
&= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \\
&\quad \times\left\langle\!\!\left\langle 1 - \frac{1}{2}\langle\chi\rangle_{\mathbf{x}(\tau)} + \frac{3}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}^2 + \frac{\mathcal{T}}{8}\left[\left\langle(\partial_z\chi)^2\right\rangle_{\mathbf{x}(\tau)} - 2\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)} + \partial_z^2\langle\chi^2\rangle_{\mathbf{x}(\tau)}\right]\right.\right. \\
&\qquad\qquad \left.\left. + \frac{\mathcal{T}^2}{32}\left[\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)}\right]^2 - \frac{\mathcal{T}}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}\,\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)}\right\rangle\!\!\right\rangle.
\end{aligned} \tag{22.640}$$

To obtain the lowest-order interaction between the surfaces, we will again set $\chi = \chi_1 + \chi_2$, discarding all terms that are not of the form $\langle\chi_1\rangle\langle\chi_2\rangle$. We will also discard the $\langle(\partial_z\chi)^2\rangle$ and $\langle\chi^2\rangle$ terms, since these will vanish provided the two bodies do not overlap. First tossing out these latter terms and the first-order terms in χ,

$$E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \frac{3}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}^2 + \frac{\mathcal{T}^2}{32}\left[\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)}\right]^2 - \frac{\mathcal{T}}{8}\langle\chi\rangle_{\mathbf{x}(\tau)}\,\partial_z^2\langle\chi\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle, \tag{22.641}$$

and now labeling the individual bodies, we have

$$E_{\text{EM}}^{(\text{TM})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \frac{3}{4}\langle\chi_1\rangle_{\mathbf{x}(\tau)}\langle\chi_2\rangle_{\mathbf{x}(\tau)} + \frac{\mathcal{T}^2}{16}\Big[\partial_z^2\langle\chi_1\rangle_{\mathbf{x}(\tau)}\Big]\Big[\partial_z^2\langle\chi_2\rangle_{\mathbf{x}(\tau)}\Big] - \frac{\mathcal{T}}{8}\langle\chi_1\rangle_{\mathbf{x}(\tau)}\,\partial_z^2\langle\chi_2\rangle_{\mathbf{x}(\tau)} - \frac{\mathcal{T}}{8}\langle\chi_2\rangle_{\mathbf{x}(\tau)}\,\partial_z^2\langle\chi_1\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle.$$

(Casimir energy, scalar EM, TM polarization, small χ) (22.642)

Note that the first term is simply the TE energy, which we have already computed. For convenience we will repeat here the calculation following Eq. (22.610), where we compute the interaction of a small element of the χ_2 body with the χ_1 body, and then integrate over the profile of the χ_2 body. Thus we have the integral (here with $\alpha = 0$)

$$\begin{aligned} I(D,\alpha;d) &:= \partial_{z_1}^\alpha \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int dz_0 \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)}\langle\chi_2\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= \chi_2 \int_d^\infty dz\,\partial_z^\alpha \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \end{aligned} \tag{22.643}$$

where z_1 is the location of the boundary of the χ_1 region. Now following the small-χ calculation of the TM atom–wall potential in Section 22.4.5,

$$\begin{aligned} I(D,\alpha;d) &= \chi_1\chi_2 \int_d^\infty dz\,\partial_z^\alpha \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle\Theta(z)\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= \chi_1\chi_2 \int_d^\infty dz\,\partial_z^\alpha \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{z}{\sqrt{\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= \chi_1\chi_2 \int_d^\infty dz\,\partial_z^\alpha z^{-D} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{1}{\sqrt{\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= (-1)^\alpha \frac{\Gamma(D+\alpha)}{\Gamma(D)}\chi_1\chi_2 \int_d^\infty dz\,z^{-(D+\alpha)} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{1}{\sqrt{\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= (-1)^\alpha \frac{\Gamma(D+\alpha-1)\chi_1\chi_2}{\Gamma(D)d^{D+\alpha-1}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{1}{\sqrt{\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= (-1)^\alpha \frac{\Gamma(D+\alpha-1)\chi_1\chi_2}{2^{D/2}\Gamma(D)d^{D+\alpha-1}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{1}{\sqrt{2\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= (-1)^\alpha \frac{\Gamma(D+\alpha-1)\chi_1\chi_2}{2^{D/2}\Gamma(D)d^{D+\alpha-1}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle T_{\text{s}}\Big[B(t);\frac{1}{\sqrt{2\mathcal{T}}}\Big] \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \\ &= (-1)^\alpha \frac{\Gamma(D/2)\Gamma(D+\alpha-1)\chi_1\chi_2}{2^{D/2+1}(D+1)\Gamma(D)d^{D+\alpha-1}}. \end{aligned} \tag{22.644}$$

For $D = 4$ and $\alpha = 0$, this is

$$I(D=4,\alpha=0;d) = \frac{\chi_1\chi_2}{120d^3}, \tag{22.645}$$

so that the first term in the Casimir energy (22.646) leads to

$$\begin{aligned}\frac{E_{\rm EM}^{\rm (TE)}}{A} &= -\frac{3\hbar c}{8(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \langle\chi_1\rangle_{\mathbf{x}(\tau)}\langle\chi_2\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle \\ &= -\frac{3\hbar c}{8(2\pi)^{D/2}} I(D,\alpha;d) \\ &= -\frac{\hbar c\chi_1\chi_2}{1280\pi^2 d^3},\end{aligned} \tag{22.646}$$

in agreement with the prior calculation (22.615). The last two terms in Eq. (22.642) then have the form

$$\begin{aligned}E_{3,4} &= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle -\frac{\mathcal{T}}{8}\langle\chi_2\rangle_{\mathbf{x}(\tau)}\,\partial_z^2\langle\chi_1\rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle \\ &= -\frac{\hbar c}{2(2\pi)^{D/2}}\left(-\frac{1}{8}I(D-2,2;d)\right) \\ &= -\frac{\hbar c}{2(2\pi)^{D/2}}\left(-\frac{1}{8}\frac{\Gamma[(D-2)/2]\Gamma(D-1)\chi_1\chi_2}{2^{D/2}(D-1)\Gamma(D-2)d^{D-1}}\right),\end{aligned} \tag{22.647}$$

and the second term in Eq. (22.642) has the form

$$\begin{aligned}E_2 &= -\frac{\hbar c}{2(2\pi)^{D/2}}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}}\int d^{D-1}x_0 \left\langle\!\!\left\langle \frac{\mathcal{T}^2}{16}\Big[\partial_z^2\langle\chi_1\rangle_{\mathbf{x}(\tau)}\Big]\Big[\partial_z^2\langle\chi_2\rangle_{\mathbf{x}(\tau)}\Big] \right\rangle\!\!\right\rangle \\ &= -\frac{\hbar c}{2(2\pi)^{D/2}}\left(\frac{1}{16}I(D-4,4;d)\right) \\ &= -\frac{\hbar c}{2(2\pi)^{D/2}}\left(\frac{1}{16}\frac{\Gamma[(D-4)/2]\Gamma(D-1)\chi_1\chi_2}{2^{D/2-1}(D-3)\Gamma(D-4)d^{D-1}}\right),\end{aligned} \tag{22.648}$$

where we integrated by parts in the second step, and we have implicitly carried out dimensional regularization in the last step. Assembling all the pieces,

$$\begin{aligned}E_{\rm EM}^{\rm (TM)} &= -\frac{\hbar c\chi_1\chi_2}{2(2\pi)^{D/2}d^{D-1}}\left(\frac{3}{4}\frac{\Gamma(D/2)\Gamma(D-1)}{2^{D/2+1}(D+1)\Gamma(D)} + \frac{1}{16}\frac{\Gamma[(D-4)/2]\Gamma(D-1)}{2^{D/2-1}(D-3)\Gamma(D-4)} - \frac{1}{4}\frac{\Gamma[(D-2)/2]\Gamma(D-1)}{2^{D/2}(D-1)\Gamma(D-2)}\right) \\ &= -\frac{\hbar c\chi_1\chi_2}{2(2\pi)^{D/2}d^{D-1}}\left(\frac{\sqrt{\pi}\Gamma(D-1)[4D(D-1)-5]}{\Gamma[(D+3)/2]}\right)\end{aligned} \tag{22.649}$$

The quantity in parenthesis is $43/480$ for $D=4$, so the small-χ Casimir energy is

$$E_{\rm EM}^{\rm (TM)} = -\frac{43\hbar c\chi_1\chi_2}{3840\pi^2 d^3}$$

(Casimir energy, scalar EM, TM polarization, small χ) (22.650)

Note that this effect is substantially stronger than the TE contribution in Eq. (22.615). Combining the two polarizations, we obtain

$$E_{\rm EM} = E_{\rm EM}^{\rm (TE)} + E_{\rm EM}^{\rm (TM)} = -\frac{23\hbar c\chi_1\chi_2}{1920\pi^2 d^3}$$

(Casimir energy, small χ) (22.651)

for the total Casimir potential in the weak-coupling limit.

22.12.2.2 Feynman–Kac Formula

To proceed with the calculation for arbitrary dielectric strength, will begin in analogy to Section 22.12.1.1 by considering the path integral

$$L(\lambda,s;d,d_0) := \int_0^\infty \frac{dt}{\sqrt{t}}\,e^{-\lambda t}\left\langle\!\!\left\langle \exp\Big(-sT_{\rm ds}\Big[B_t(t');d,d_0\Big]\Big)\right\rangle\!\!\right\rangle_{B_t}, \tag{22.652}$$

as in Eq. (22.652), but also involving the TM potential as in ???????, Here, we are again using the double-sojourn time, defined in Eq. (22.555) by

$$T_{\rm ds}[B_t; d, d_0] := \int_0^t dt' \, \mathbb{1}_{(-\infty,d_0]\cup[d_0+d,\infty)}\Big[B_t(t')\Big]. \tag{22.653}$$

22.13 Exercises

Problem 22.1

Beginning with the electric-field energy (22.47)

$$H_{\rm E} = \frac{1}{2}\int d^3r \, \phi(\mathbf{r})\,\rho(\mathbf{r}), \tag{22.654}$$

separate the free and bound charge densities, $\rho = \rho_{\rm free} + \rho_{\rm P}$, and use

$$\rho_{\rm P} := -\nabla\cdot\mathbf{P} \tag{22.655}$$

to introduce the dielectric polarization. Show that the result you obtain is not the correct energy for a linear dielectric.

Problem 22.2

Consider a scalar field that obeys the wave equation

$$\nabla^2\phi = \mu_0\epsilon\partial_t^2\phi. \tag{22.656}$$

Write down the plane-wave solutions and dispersion relation for this wave equation.

Problem 22.3

Consider again the scalar wave equation (22.656). We will identify the ''electromagnetic fields''

$$E := -\partial_t\phi, \qquad \mathbf{B} := \nabla\phi. \tag{22.657}$$

Now start with the Lagrangian

$$L = \frac{\epsilon_0}{2}\int d^3r \left[\frac{\epsilon}{\epsilon_0}(\partial_t\phi)^2 - c^2(\nabla\phi)^2\right], \tag{22.658}$$

and work out the Hamiltonian, both in canonical coordinates and in terms of the fields.

Problem 22.4

Quantize the wave equation in Problem 22.2, by decomposing the Hamiltonian from Problem 22.3 into normal modes. In doing this, you should develop the conjugate normal-mode operators

$$\begin{aligned}\phi_k &= \sqrt{\frac{\hbar}{2\omega_k\epsilon_0}}\,[f_k(\mathbf{r})\,a_k + \text{H.c.}]\\ \pi_k &= -i\frac{\epsilon(\mathbf{r})}{\epsilon_0}\sqrt{\frac{\hbar\omega_k\epsilon_0}{2}}\,[f_k(\mathbf{r})\,a_k - \text{H.c.}]\,,\end{aligned} \tag{22.659}$$

with second-quantized field operators

$$\phi = \sum_k \phi_k, \qquad \pi = \sum_k \pi_k, \tag{22.660}$$

and where the mode functions $f_k(\mathbf{r})$ satisfy the orthonormality condition

$$\int d^3r \, \frac{\epsilon(\mathbf{r})}{\epsilon_0} f_k(\mathbf{r}) f_{k'}^*(\mathbf{r}) = \delta_{kk'}, \tag{22.661}$$

as is analogous to the full electromagnetic case.

Problem 22.5

For the quantized scalar field coupled to a dielectric, as in Problem 22.4, show that the vacuum energy is given by the expression

$$E_{\text{EM}}(\mathbf{r}) = \langle 0|H_{\text{EM}}|0\rangle = \sum_k \int d^3r \, \frac{\hbar\omega_k}{2} \frac{\epsilon(\mathbf{r})}{\epsilon_0} \left|f_k(\mathbf{r})\right|^2 . \tag{22.662}$$

Problem 22.6

A dielectric film (slab) of thickness d surrounded by vacuum is described in ordinary electromagnetism by the field reflection coefficient[28]

$$r_{\text{film}}^{(\text{TM,TE})}(\theta,\omega) = \frac{r_{\text{TM,TE}}(\theta,\omega)\,(1-e^{i\phi})}{1-r_{\text{TM,TE}}^2(\theta,\omega)\,e^{i\phi}} \tag{22.663}$$

for the two polarizations in terms of the appropriate Fresnel coefficients for the incident reflection, where the round-trip phase ϕ in the film is given by

$$\phi = 2kd\sqrt{n^2(\omega) - \sin^2\theta}, \tag{22.664}$$

and k is the optical wave number and n is the refractive index. The other parameters here are the refractive index $n(\omega)$ of the film and the film thickness d.

Show that, in the limit of a very thin film, and at normal incidence, that the reflection coefficient is given by

$$r_{\text{film}} = i\frac{kd\chi}{2} \tag{22.665}$$

Problem 22.7

Using the reflection coefficient from Problem 22.6, use a mode summation for the 1D version of the scalar field in Problem 22.4 and the formula

$$V_{\text{CP}}^{(\text{TE})} = -\frac{1}{2}\alpha_0 \langle E\rangle^2 \tag{22.666}$$

in terms of the vacuum expectation value to compute the (TE component of the) Casimir–Polder potential for an atom a distance z from a very thin dielectric thin film of thickness d. Your result should be

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{\hbar c\alpha_0 d\chi}{32\pi\epsilon_0 z^3}. \tag{22.667}$$

[28] Daniel A. Steck, *Classical and Modern Optics*, available online at `http://steck.us/teaching`.

Problem 22.8

In analogy with Problem 22.7, use a mode-summation approach to compute the (TE component of the) Casimir–Polder potential in 1D electromagnetism for an atom a distance z away from a dielectric half space of susceptibility χ, in the limit $\chi \ll 1$. Your result should be

$$V_{\text{CP}}^{(\text{TE})}(z) = -\frac{\hbar c \alpha_0 \chi}{64\pi\epsilon_0 z^2}. \tag{22.668}$$

Problem 22.9

In analogy with Problem 22.7, use a mode-summation approach to compute the (TE component of the) Casimir–Polder potential in 1D electromagnetism for an atom a distance z away from a planar, perfect conductor.

Problem 22.10

Use the weak-coupling path integral [Eq. (22.226)]

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = -\frac{3\hbar c \alpha_0}{8(4\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \chi \rangle_{\mathbf{x}(\tau)} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.669}$$

to compute the (TE component of the) Casimir–Polder potential for an atom a distance z from a very thin dielectric thin film of thickness d and susceptibility χ. Show that your results are consistent with Problem 22.7 in one dimension.

Problem 22.11

Use the TE-polarization path integral [Eq. (22.197)]

$$V_{\text{CP}}^{(\text{TE})}(\mathbf{r}) = \frac{\hbar c \alpha_0}{4(2\pi)^{D/2}\epsilon_0} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-3/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.670}$$

to compute the TE component of the Casimir–Polder potential for an atom at a distance z from a thin dielectric plane of the form

$$\epsilon_{\text{r}}(z) = 1 + \hat{\chi}\delta(z) \tag{22.671}$$

for the case $D = 4$. Show that your results are consistent in the appropriate limit with the results of Problem 22.10.

Problem 22.12

Using the path integral (22.553) for the TE component of the Casimir energy

$$E_{\text{EM}}^{(\text{TE})} = -\frac{\hbar c}{2(2\pi)^{D/2}} \int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \int d^{D-1}x_0 \left\langle\!\!\left\langle \langle \epsilon_{\text{r}} \rangle_{\mathbf{x}(\tau)}^{-1/2} \right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)}, \tag{22.672}$$

for a pair of thin dielectric planar membranes separated by distance d:

$$\epsilon_{\text{r}}(z) = 1 + \hat{\chi}\delta(z) + \hat{\chi}\delta(z-d). \tag{22.673}$$

Begin by defining the "double-local time"

$$\ell_{\text{dl}}[B_t; d, d_0] := \int_0^t dt' \Big(\delta\big[B_t(t') - d_0\big] + \delta\big[B_t(t') - d - d_0\big] \Big), \tag{22.674}$$

and then use the Feynman–Kac formula to compute the path integral

$$L(\lambda, s; d, d_0) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- s\ell_{\rm dl}\Big[B_t(t'); d, d_0\Big]\Big)\right\rangle\!\!\right\rangle_{B_t}, \tag{22.675}$$

which can be transformed into the Casimir energy (22.672), following the procedure of Section 22.12.1. Analyze the energy in the limits of large and small d, noting any pathologies that you find.

Problem 22.13

Recalculate the path integral

$$L(\lambda; d, \Xi) := \int_0^\infty \frac{dt}{\sqrt{t}}\, e^{-\lambda t} \left\langle\!\!\left\langle \exp\Big(- \mathfrak{N}[B_t; d, \Xi]\Big)\right\rangle\!\!\right\rangle_{B_t}, \tag{22.676}$$

where

$$\begin{aligned}
\mathfrak{N}[B_t; d, \Xi] &:= \lim_{a\longrightarrow 0} \mathfrak{M}[B_t; d, \Xi, a] \\
&= \frac{\Xi}{2}\int_0^t dt' \left\{\Xi\delta^2\Big[B_t(t') - d\Big] - \delta'\Big[B_t(t') - d\Big]\right\} \\
\mathfrak{M}[B_t; d, \Xi, a] &:= \int_0^t dt'\, V_{\rm TM}\big[B_t(t')\big] \\
&= \frac{\Xi}{2a}\int_0^t dt' \left\{\frac{\Xi}{a}\mathbb{1}_{[d,d+a]}\Big[B_t(t')\Big] + \delta\Big[B_t(t') - (d+a)\Big] - \delta\Big[B_t(t') - d\Big]\right\} \\
&= \frac{\Xi}{2a}\left\{\frac{\Xi}{a}T_{\rm s}\Big(B_t; [d, d+a]\Big) + \ell\big(B_t; d+a\big) - \ell\big(B_t; d\big)\right\},
\end{aligned} \tag{22.677}$$

by using the Feynman–Kac formula (22.299), but taking $g(x) = \delta(x)$ to directly pin $W(t)$ to the origin.

Problem 22.14

Give integral expressions for the Casimir–Polder potential (both polarizations) in 2D electromagnetism for an atom on the vacuum side of a planar, dielectric interface. Obtain explicit expressions for the strong- and weak-coupling limits.

Problem 22.15

Evaluate the derivative-free TM path integral (22.479)

$$V_{\rm CP}^{\rm (TM)}(\mathbf{r}) = -\frac{\hbar c\alpha_0}{4(2\pi)^{D/2}\epsilon_0}\int_0^\infty \frac{d\mathcal{T}}{\mathcal{T}^{1+D/2}} \left\langle\!\!\left\langle \left(\langle\epsilon_{\rm r}\rangle_{\mathbf{x}(\tau)}^{-3/2} + 2(D-2)\langle\epsilon_{\rm r}\rangle_{\mathbf{x}(\tau)}^{-1/2}\right) e^{-\mathcal{T}\langle V_{\rm TM}\rangle_{\mathbf{x}(\tau)}}\right\rangle\!\!\right\rangle_{\mathbf{x}(\tau)} \tag{22.678}$$

for an atom near one planar, dielectric interface in the limit of small χ. Do this by expanding the path integral to first order in χ right at the beginning.

Problem 22.16

The goal of this problem is to review the techniques for analytically summing the worldline path integral for a perfectly conducting plane, for the TE-polarization field.

(a) Starting with the Green-function form of the Feynman–Kac diffusion equation [Eq. (22.300),

$$f''(x) = 2\big[\lambda + sV(x)\big]f(x) - 2g(x), \tag{22.679}$$

set $g(x) = \delta(x)$ and $V(x) = 0$ to find the Green function in free space.

(b) Then find the Green function in the presence of a perfectly conducting boundary at $x = d$, using the image method to satisfy the Dirichlet boundary condition.

(c) Use the result to compute the Casimir–Polder potential, assuming TE polarization, following the general method for computing the potential at a dielectric interface from Section 22.7.

(d) Repeat the calculation in steps (a)–(c), but using the full 3D Green function solution (i.e., keep the ''ignorable'' transverse dimensions explicitly).

Part IV

Numerical Methods in Quantum Optics

Chapter 23

Welcome to the Machine

23.1 Finite Representations of Real Numbers

Almost all digital computers store numbers as binary strings of zeros and ones. Representing an integer in binary form is straightforward assuming you understand binary vs. decimal counting. The only subtle point is whether the integer is *signed* or *unsigned*. The only difference is in the interpretation of the sign bit. For example, a 32-bit unsigned integer can range from 0 to $2^{32} - 1 = 4\ 294\ 967\ 295$. By contrast, in a 32-bit *signed* integer, which ranges from $-2^{31} = -2\ 147\ 483\ 648$ to $2^{31} - 1 = 2\ 147\ 483\ 647$, the idea is the same, but the first bit is interpreted as the **sign bit** (with a one representing a negative integer). That is, any integer greater than $2^{31} - 1$ is simply *wrapped* by subtracting 2^{32}. Indeed, integer arithmetic is always performed modulo the range of the integer type, so that in unsigned 32-bit arithmetic, $4\ 294\ 967\ 295 + 1 = 0$. Note that modern Fortran has its own processor-independent integer model that is somewhat more restrictive than this (i.e., integers are always signed), but Fortran integers can effectively always be used in the same ways as C integers, with the right interpretation.

Fixed-point arithmetic is the simplest way to represent real numbers. Basically, fixed-point numbers are integers with a decimal place stuck in some fixed position. Thus, they are in some sense a reinterpretation of integers. Addition is straightforward, whereas multiplication is a bit more complicated. Even though fixed-point arithmetic is typically very fast, the limited dynamic range of fixed-point numbers usually leads to serious problems with accuracy, and consequently they are only used in specialized applications.

The **floating-point number** is the standard on modern computing machines. A floating point number is decomposed into two parts, a **mantissa** and an **exponent**. For example, the real number 328.925 is represented in floating-point arithmetic in usual scientific notation as

$$3.28925 \times 10^2, \tag{23.1}$$

or in computer-style notation as `3.28925E2` (in Fortran the `E` can be changed to a `D` to denote a double-precision number, which we will discuss below). The mantissa and exponent are then simply represented as integers, with a fixed decimal place associated with the mantissa; note that for a binary representation, the exponent is chosen such that the mantissa is in the range $[1, 2)$. There are many ways to decide how to associate binary data for floating-point numbers, but by far the most common standard is dictated by the IEEE Standard for Binary Floating-Point Arithmetic (ANSI/IEEE Std 754-1985). IEEE single precision allocates 32 bits for a floating-point number. Without going into too much detail, one bit represents the sign, 8 bits represent the exponent, and the rest represent the mantissa. IEEE single precision is characterized by the following values:

- ‘‘machine epsilon’’ (`epsilon(1.0)` in Fortran 90) of 1.1920929E-7
- smallest positive number (`tiny(1.0)` in Fortran 90) of 1.1754944E-38
- largest number (`huge(1.0)` in Fortran 90) of 3.4028235E+38

Here **machine epsilon** is a measure of the precision of the representation, and is defined as the smallest number that, when added to 1, returns a number different from 1. Thus, single precision gets about 7 significant digits, and ranges through roughly $10^{\pm 38}$. (Actually, nonzero values closer to zero can be represented, but are not usually accessible in Fortran.)

IEEE double precision extends this range considerably, and is characterized by the following values:

- machine epsilon (`epsilon(1.0D0)` in Fortran 90) of 2.220446049250313E-16
- smallest positive number (`tiny(1.0D0)` in Fortran 90) of 2.2250738585072014E-308
- largest number (`huge(1.0D0)` in Fortran 90) of 1.7976931348623157E+308

Thus, we get about 16 significant digits and a range through about $10^{\pm 308}$. There are two perfectly reasonable but opposite philosophies considering the choice of precision. The first goes: ''modern computers are so fast that you may as well just use double precision just to be safe.'' Actually, as 64-bit processors become more common, the speed penalty for using double precision instead of single precision is shrinking. However, the other philosophy is ''if you *need* double precision, your algorithm is probably flawed; single precision is almost always sufficient.'' You'll just have to decide for yourself, but it is generally useful to write your code such that it is easy to switch precision. In Fortran 90, you do this by defining a parameter `wp` (for ''working precision'' as

```
integer, parameter :: wp = selected_real_kind(p=14)
```

to select double precision. Then, whenever you declare a variable, use this parameter, as in declaring variables such as `real(wp) :: a, b`, in specifying numbers such as `a = 1.0_wp`, and in rounding, such as `a = real(b, kind=wp)`. To switch to single precision, simply change `p=14` to `p=4` (meaning ''I want at least 14 digits of precision'' and ''at least 4 digits of precision, please'') as the argument to `selected_real_kind` above.

23.2 Machine Structure and Optimization

Here we will discuss just a few of the basic features of modern processors that relate to high-performance numerical computing. This is an area that is complicated, varies widely among processor vendors, and evolves rapidly, and so a general and introductory discussion of this sort must necessarily sacrifice detail and completeness.[1] However, knowledge of some basic concepts is invaluable in tuning your codes for high performance.

23.2.1 Memory Hierarchy

The first thing to understand is how the processor accesses, stores, and manipulates information. There is a hierarchy of locations in which information can be stored. Here, we list these in decreasing order of access speed, and conversely, increasing order of size.

- **Registers.** The **registers** are the memory locations for the data on which the processor is currently working. There are typically of the order of 10 registers on a processor, and obviously they must operate at the nominal execution speed of the processor. A typical machine instruction for a floating-point operation might involve, say, taking the contents of registers 0 and 1, multiplying them together, and storing the result in register 0. In the past, a special register called the *accumulator* was typically the register that was always involved in any particular operation and also the destination for the result of any operation. Now the term is less common, but one or two registers are still typically of more importance than the rest.

  For numerical calculations, it is useful to note that on some processors, the registers handle more data than regular memory addresses. For example, even when processing 64-bit data, the registers

[1] For more detailed discussions, see Kevin Dowd and Charles Severance, *High-Performance Computing*, 2nd ed. (O'Reilly, 1998). While this reference is somewhat dated, it is clear and readable, and still contains a great deal of relevant information.

might be designed be 80-bit or 128-bit "extended precision" registers (as happens, for example, on modern Intel processors). Thus, if you compute the sum of many numbers, the processor can keep the intermediate sums with extra precision to reduce roundoff error, particularly in cases where the sum is sensitive to the order of addition due to roundoff errors. This also helps reduce unexpected effects due to unintended reordering of operations at the compiler or processor levels.

- **Cache.** Next comes a small area of high-speed memory called **cache memory**. The idea is that the main memory, where large quantities of data can be stored is quite slow, but the processor needs to take in and push out data very quickly. Cache is the intermediate area that the processor can use on short time scales. Think of it as a small pile of papers on your desk that you want quick access to, compared to main memory, which is more like the filing cabinet. Because this memory is so fast, it is quite expensive, and thus its size is quite limited. Modern designs also use multiple levels of cache: L1 ("level 1") cache is the fastest and smallest, typically being on the order of a few to a few hundred KB, and talks directly to the cpu; L2 cache is larger and slower, typically on the order of a few hundred KB or larger; and some designs even incorporate an L3 cache.

  Cache is a particularly important concept in modern computers, because the processors must *not* be kept waiting for the information they need. Of course, the cache can only keep a small subset of the total data in main memory on hand, but if it doesn't have a particular piece of data when the processor requests it, the computation stalls while the data are fetched from a higher-level cache or even main memory. One fairly obvious strategy that helps here is to process data in fairly small-sized chunks such that "blocks" of the calculation can fit entirely in cache. Of course, most calculations where you care about speed will *not* fit in cache, and manually breaking up calculations into cache-sized chunks is difficult and guaranteed to render your code unreadable or at least ugly.

  The other strategy requires a bit more understanding about how cache works. Essentially, the various elements of cache are copies of various elements of main memory. However, if each location in cache were a copy of a completely independent location in main memory, we would need *another* bank of fast memory, the same size as the cache, just so we would know which main memory entry each cache entry referred to. To reduce this memory overhead, cache elements are grouped into **cache lines**, so that when one datum from memory is needed, a whole line's worth of data are actually fetched from memory. Thus, as long as you have to fetch all the data together, you may as well make the best possible use of them. The basic strategy is this: stick to **unit-stride access** as much as possible. That is, if you are processing long arrays of data, try to access the elements *only sequentially*, if at possible. The canonical example here is in computing the sum of a matrix. In Fortran, a two-dimensional array `A` is stored in memory *by column*, or such that `A(1,1)` and `A(2,1)` are adjacent in memory, while `A(1,1)` and `A(1,2)` are separated in memory by at least the length of the first array dimension. Thus, for example, what might seem a reasonable method for computing the array sum,

```
s = 0
do j = 1, m
  do k = 1, n
    s = s + A(j, k)
  end do
end do
```

  is actually a *bad* idea, because the access in memory in the inner loop has a stride of (at least) `m`. In the worst, case, an entire line of cache must be fetched for *each* addition operation, slowing things down considerably. Fortunately, this problem is easily fixed by switching the order of the loops:

```
s = 0
do k = 1, n
  do j = 1, m
    s = s + A(j, k)
  end do
end do
```

The access here is now unit-stride, and this makes optimal use of the cache since (almost) *all* the fetched data are used in the sum. Most compilers will detect the problem in the former code sample and change it to the latter, depending on the level of optimization you request (after all, you might *want* the former calculation due to some issue with roundoff errors, since the results of the two codes are not guaranteed to be identical). In a more modern approach, as in Fortran 90, you can simply use an intrinsic such as `sum`, as in

```
s = sum(A)
```

or

```
s = sum( A(1:m, 1:n) )
```

to accomplish the same calculation, but explicitly giving the compiler freedom to choose the best way to perform the sum. If an intrinsic does not exist to do what you want, there are other constructs such as the `forall` loop in Fortran 95, which can be used to indicate that there is no dependency among the iterations of multiple loops, so that the compiler can perform the operations in any (presumably optimal) order.

The problems with cache can be even a bit more insidious than what we have indicated. To see why, consider a slightly different code that performs a sum over the second dimension of an array:

```
dimension A(2048, 128)
do j = 1, 2048
  do k = 1, 128
    A(j,1) = A(j,1) + A(j, k)
  end do
end do
```

If `A` is stored contiguously in memory, then the elements `A(j,k)` for the same `j` but different `k` are separated by powers of two. But to perform the calculation, lines of cache are fetched, corresponding to `A(j,k)` for the same `k` but different `j`, which is not so useful. In a **fully associative cache**, a line in cache can be associated with *any* line in memory, and many lines of cache can be fetched on the first iteration of the `j` loop, so that they will be reused later on subsequent iterations. However, to reduce complexity and cost, most cache is **set-associative**, which means that a line in cache can only map to particular lines in memory, and conversely a line in memory can only map to a few (say two or four) cache lines. Typically, a cache line maps to locations in memory that are widely spaced by some number of lines given by a power of two. The problem in the above example is that the full array itself has as its first dimension a large power of two. All of the `A(j,k)` for fixed `j` are needed at the same time, but many will overlap to the same few lines of cache. A line of cache has to be fetched for each `k`, but they all can't be stored in cache. So the same lines need to be refetched on subsequent iterations of the `j` loop, and in the worst case, a line of cache must be fetched for *each* addition operation. So even if the relevant data could have fit in cache, the power-of-two associativity caused fetched data to be flushed before it was needed, so that the cache needed to fetch it again. This behavior is called **cache thrashing** and can be highly detrimental to the performance of your code. In addition to unit stride access, cache thrashing can sometimes be avoided (as in this example) by padding the first array dimensions to some larger (non-power-of-two) value, to avoid problems with associativity. (An appropriate change in the example would be $2048 \longrightarrow 2048 + 128$.) In general, you should avoid accessing data with large, power-of-two strides (some traditional FFT algorithms are notorious for doing just this).

- **Main Memory.** Most of your data reside in **main memory** during a calculation, which can be large (in the range of tens of GB on the best single-box machines). However, it is slow, which is the point of having cache. The strategies we discussed for cache also apply here, since main memory is often also **banked**. That is, sequential chunks in memory can come from different banks of memory to reduce

latency when fetching data quickly. The number of banks is typically a power of two, so power-of-two strides are bad here as well, and clearly unit stride is best: after accessing one bank for data, you would then access the next bank for the next piece of data, giving the first some time to recover and prepare for the next fetch before you bother it again.

- **Storage.** Of course, memory for even more data in the longer term comes in the form of disks, tapes, and so on. These are really slow, and if you have to use slow storage on the fly because your computation is so large then you're really in trouble.

23.2.2 Pipeline

One crucially important concept in understanding modern, ultrahigh-speed processors is the processor **pipeline**. Let's consider an analogy to motivate the pipeline. Suppose you have a ''clothes-cleaning machine.'' It's quite nice, you just pop in a load of clothes, and then it proceeds to wash, dry, iron, and fold them for you. Suppose each of these tasks takes the machine 15 minutes. Then the rate at which the machine washes clothes is 1 load/hour. Of course, that's not very efficient, since while the clothes are being dried, the machinery related to washing the clothes sits idle. But if you have a *lot* of laundry to do, you'd want a more clever design for a machine would divide the machine into 4 units, one for each task. Each task still takes 15 minutes, and after you put in your first load of laundry, it's an hour before you see a set of spanking-fresh linens. But after the first load finishes washing and goes to the dryer, you can start the *second* load of clothes right away. In all you can have a total of 4 loads of laundry ''in flight,'' and after a delay of one hour, you effectively are finishing one load of laundry every 15 minutes, or four times faster than the first machine. This is the essence of pipelining: dividing the work up into stages, so that several different instructions can be processed in different parts of the ''assembly line'' at once. It should be apparent that dividing up tasks into smaller pieces allow for longer pipelines, and thus for faster processors. Indeed, the recent offerings from Intel bear this out, with pipelines of 10 stages for the Pentium III, 20 stages for the Pentium 4, 31 stages for the Xeon Nocona (with processor speeds currently in the high 3 GHz range). By contrast, the older 8086 used no pipeline at all, and executed one instruction per cycle with no latency.

The crucial point here is this: the pipeline is your friend *only* when full. With a full 31-stage pipeline, it appears (after a delay of 31 clock cycles) that one instruction is being executed per cycle. However, suppose that successive instructions depend on each other, so that the next instruction can't be started until the previous one is finished (say, as in iterating a map of the form $x_{n+1} = (ax_n) (\mathrm{mod}\, b)$, which you might do to implement a random-number generator). In this case, it takes 31 clock cycles to execute each operation, because no pipelining is possible, and you've effectively just cut the processor speed down by a factor of 31. That's bad. In fact, you can see why cache thrashing is even worse on a fast processor: not only does the processor have to wait for the memory fetch, but the problem could be compounded if no other instructions are ready to go, since the pipeline will drain.

Modern processors implement a bunch of tricks, many of which are aimed at keeping the pipeline packed with instructions, to keep the performance high. We will discuss them below. But it is important to note that with speed limitations on memory (especially main memory), and with constraints on instruction sets, and so on, it is typically very hard to keep a processor working at its ''theoretical peak.'' For example, Intel Xeon processors can theoretically execute 1 floating-point operation (flop) per processor cycle, with register, cache, and other overhead it is often difficult to execute flops on more than, say, 30% of the processor cycles. You should definitely keep this in mind when tuning your codes with hardware performance counters (see below), so you don't have unrealistic goals for your flop counts.

There is a particular class of processor, the **vector processor** (as opposed to the above cache-based, or scalar, processors), that is optimized for just one task: take contiguous chunks of data, and perform the same mathematical operation on all the data, and do this quickly. These are very good for scientific computation, again provided you use unit strides in your calculation. In these processors, it is much more realistic to achieve flop rates nearing the 100% ideal. However, these days, such processors are expensive and normally relegated to the best supercomputers (i.e., computers that aren't sitting on your desktop). So it's important to learn to deal with the constraints of the cheaper (and ubiquitous) scalar processor.

However, it is difficult to do much in a code to avoid ''bubbles'' in the pipeline, beyond what you would already do to make a ''cache-friendly'' code. This is especially true since pipelines vary greatly among processors, with most processors having multiple pipelines (say, to handle integer and floating-point instructions separately). You have to rely heavily on good compilers to provide a mix of instructions to the processor without a lot of dependencies to keep the pipelines working. Some rather simple tricks, like Intel's ''hyperthreading technology,'' rely on executing multiple codes at once, so the different codes fill in each others' pipeline bubbles, so at least the processor is staying more busy overall, even if each code is not executing faster.

23.2.2.1 Out-of-Order Execution

One of the tricks implemented by all of the fastest modern processors is **out-of-order execution**. The concept is fairly simple: as the machine instructions are being sent to the pipelines for execution, they are first held in a buffer (of something like 100 instructions), and then analyzed for dependencies. If the processor detects an instruction that depends on the result of another instruction (either also in the buffer or already in the pipeline), it is free to dynamically reorder the instructions so that the dependent instruction is moved back in the queue so the pipeline isn't idle while the instruction waits. This helps the execution speeds greatly, but makes it difficult to analyze how a set of machine instructions will actually be executed (especially if you're trying to hand-code an optimized routine in machine language, or trying to disassemble some compiled code to analyze its performance). Again, there isn't much for you to do here, you have to hope your compiler is good enough to provide a good mix of instructions for the out-of-order buffer to work with.

23.2.2.2 Loop Unrolling

Very often, you need to perform a repeated computation on multiple elements on an array:

```
do j = 1, n
  a(j) = a(j) * b(j)
end do
```

The key idea here is that at each loop iteration, the processor must execute a branch (if/then) instruction to decide if the next iteration should be performed. This is particularly bad if the processor waits for each branch instruction to be carried out before starting the next loop multiplication, since effectively all benefits of having a pipeline are lost. (Branch prediction, described below, helps this somewhat.) Thus, it would help to rewrite the loop as

```
do j = 1, n, 4
  a(j)   = a(j)   * b(j)
  a(j+1) = a(j+1) * b(j+1)
  a(j+2) = a(j+2) * b(j+2)
  a(j+3) = a(j+3) * b(j+3)
end do
```

This loop is said to have been ''unrolled four times.'' Now there is no branch separating the instructions in the loop, and four iterations of the original loop can be pipelined right away without any branch overhead. Of course, the code here assumes that `n` is a multiple of 4; the unrolled loop is more complicated for arbitrary `n`. The tradeoff here is that the resulting code is larger than the original. The benefits also decrease as the loop is unrolled more, so excessive loop unrolling is not useful. Generally, this is handled by the compiler optimizer, and not in your code, so your code stays readable. However it is useful to know that a loop can be unrolled only if its iterations are independent. In certain constructions, for example with pointers, the compiler may not be able to ''prove'' that the iterations are independent, and thus not optimize it. For example, consider this loop:

```
real, dimension(n) :: a, b
integer, dimension(n) :: c
```

```
do j = 1, n
  a(c(j)) = a(c(j)) * b(j)
end do
```

If you know that the array `c` is a permutation of the set $\{1, \ldots, n\}$, then the loop can still be unrolled. However, a compiler would likely assume that values in the `c` array could have been repeated, and thus not optimized the loop. In this case, a *compiler directive* would help by telling the compiler that it can assume the loop iterations to be independent. In the HPF (High-Performance Fortran) language, this would look like this:

```
!hpf$ independent
  do j = 1, n
    a(c(j)) = a(c(j)) * b(j)
  end do
```

This also occurs in the original loop, if it is in a subroutine (this is valid Fortran 90):

```
subroutine foo(a, b, n)
  integer :: n, j
  real, dimension(n), intent(inout) :: a, b
  do j = 1, n
    a(j) = a(j) * b(j)
  end do
end subroutine foo
```

The problem could arise in some languages if the two arrays overlap in memory, for example if the two arguments are overlapping parts of the *same* array, as in `call foo(c(2:6), c(1:5), 5)`. In this case, the results depend on the order in which the loop iterations are executed, since the value of elements of `b` are changing. Actually, this problematic call is explicitly disallowed in Fortran 90: array arguments must not overlap if they are defined or modified by the subroutine. Fortran 90 tends to make choices to favor optimization over flexibility, and the loop could be unrolled by the compiler in this example. However, in most other languages (like C), the call would be acceptable and thus the compiler would not unroll the loop.

23.2.2.3 Branch Prediction

As we mentioned above, branch instructions (i.e., instructions to jump to different parts of a program depending on some condition, as in an `if` or `case` statement) are problematic for pipelined machines, since in principle instructions after the branch can't be executed until the result from the branch condition is known. Actually, you *could* start executing instructions before the branch result is known: just pick one outcome, and start executing the appropriate instructions, hoping to win if the processor guesses the right outcome in advance. This trick is called **speculative execution**. The problem is that canceling (or **retiring**) the finished and in-flight instructions in the case of a wrong guess involves a lot of overhead, and unless the processor has a good way to accurately guess the result of the branch condition, speculative execution could actually slow things down. Thus enters the art of **branch prediction**. Again, the prediction in modern processors must be good, since in some processors the cost of a mispredicted branch is stalling and flushing the entire pipeline to clear the false branch.

Branch-prediction algorithms are numerous and can be quite complex[2] As a simple example, we'll consider branch prediction with a dynamically updated **Moore machine**.[3] We'll do this for a three-bit machine, so consider the following table of "addresses" three bits long:

[2]For nice coverage see the Wikipedia entry "Branch Predictor," `http://en.wikipedia.org/wiki/Branch_predictor`.

[3]Edward F. Moore, "Gedanken-experiments on Sequential Machines," **Automata Studies (Annals of Mathematical Studies) 34**, 129 (1956).

b_1	b_2	b_3	output
0	0	0	x
0	0	1	x
0	1	0	x
0	1	1	x
1	0	0	x
1	0	1	x
1	1	0	x
1	1	1	x

The outputs are initially undetermined (corresponding to some default value of, say, 0). The 0's correspond to ''branch true,'' and the 1's correspond to ''branch false.'' Then the output is the prediction given the last three branch results. That is, if the last three branches were true, false, and false, then our prediction for the next branch would be the ''100'' output. Correspondingly, after the real branch result is known, the result is recorded at the same place in the table. Clearly, if the branch result is always the same thing, this algorithm will predict it perfectly after four possible mispredictions (settling into either the 000 or 111 output). Suppose now that branches come in some more complicated pattern like 1000100010001000..., and suppose we always default with 0. Shortly, the table will settle down to this:

b_1	b_2	b_3	output
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	x
1	0	0	0
1	0	1	x
1	1	0	x
1	1	1	x

That is, after a transient of 4 branches, some of which are default mispredictions, the predictor predicts the pattern perfectly. Larger predictors will obviously predict longer-period patterns, but may take longer to ''lock on.'' Basically, regular branch patterns are easy to predict, whereas any change in pattern (e.g., at the end of a loop) or branching on a random number can seriously hurt your performance if done too often.

23.2.2.4 Addition, Multiplication, and Division

One more thing we can discuss are the basic *types* of floating-point operations that you want to do. Additions are typically the easiest, with multiplications next, and divisions are by far the hardest. Most modern processors are set up to churn out (with pipelining) one floating-point operation per cycle, such as one addition. Many also can do one floating-point multiplication per cycle (per pipeline). However, there are many variations on this theme. Modern Intel Xeon processors can only pipeline one multiplication every *other* cycle; however, on the cycles between multiplications, you are allowed to pipeline an addition with no extra cost. That's a sense in which additions are ''cheaper'' than multiplications (additions can be pipelined on every cycle). Other processors are set up to do one multiplication per cycle, but they can also do an addition at the same time as a multiplication in a single ''multadd'' operation. Thus, mixing floating-point multiplications and additions together can take advantage of hardware capabilities and result in very good efficiency.

Divisions are extremely bad: for example, a double-precision division on an Intel Xeon can be finished every 38 processor cycles, compared to 2 for multiplication. Many optimizing compilers have options to substitute computing b^{-1} and then multiplying by a to compute a/b, although the result may be less accurate than the divide. (Similarly, it is useful to know that it is usually possible to compute inverse square roots very quickly, and some processors have special instructions to compute the sine and cosine of a number at the same time.) It is usually best to keep your code readable and let the compiler make the appropriate transformations: most compilers can easily change a statement like `a/2` to `0.5*a`. However, compilers occasionally miss operations, and these concepts can be useful in speeding things up.

23.2.3 Avoiding Overhead

As we mentioned above in loop unrolling, we can greatly increase performance by decreasing overhead operations. Here we will briefly discuss a couple of situations where it is possible to profitably decrease overhead.

23.2.3.1 Procedure Inlining

One place where much overhead can be eliminated is in the calling of a procedure (a function or a subroutine). Whenever a procedure is called, the code must jump to a new section, with some provisions for where to return at the completion of the procedure, as well as possible setup of temporary variables and arrays. If the program spends a lot of time in the procedure, then the overhead may not be a big deal. However, consider the loop in this code example:

```
do j = 1, n
  a(j) = a(j) + foo(a(j))
  b(j) = b(j) + bar(a(j), b(j))
end do
```

If `foo` and `bar` are relatively small functions, then the function-call overhead can be quite substantial. Furthermore, the computations in the two procedures cannot be rearranged to improve performance (i.e., instructions from one procedure could fill a pipeline bubble in the other), if the procedures are compiled separately.

An easy solution is to simply take the contents of the two procedures and paste them directly into the loop to eliminate the function calls. This trick is called **procedure inlining**. Obviously, manually inlining procedures will make your code a whole lot less readable, and inlining almost always increases the size of your code. Inlining is best done at the compiler level, and even then it should be limited to certain cases, such as procedure calls in the innermost loops of your code. For example, if inlining expands the code for a loop to the point where it no longer fits into instruction cache (if there are many copies of a procedure), then inlining may slow things down. It is also worth noting that compilers have trouble inlining under certain conditions. For example, it is difficult for a compiler to inline a procedure when it is defined in a different file from which it is called (e.g., when it is in a library). To inline such separated procedures, it must defer most of the optimizations until the linking phase, and a number of compilers now do this.

23.2.3.2 Compiler Issues

Compiler vendors must work very hard to get their codes to implement the tricks we have discussed here (as well as many, many more). If your primary goal is speed at any cost you should note this: *simple, old* features of a language will be the best supported in a compiler in terms of optimization or just plain *working*, while using new, advance, fancy-schmancy features in your code will tend to inhibit optimization or even cause you to send a bug report to the vendor! For example, suppose you are integrating the Schrödinger equation, and in your subroutine to evaluate the time derivative of the wave-function array `psi`, you try to keep things organized by using a function call to compute the result of applying an operator A on ψ in one of the terms of the Schrödinger equation $\partial_t \psi = A\psi + \cdots$:

```
psidot = psidot + A_on_psi(psi)
```

In Fortran 90, this is an allowed construction: a function can return an array as a result, which here is then added to the `psidot` array. However, even otherwise excellent compilers can miss the fact that a temporary array to store the result of the function call to `A_on_psi` can in fact be eliminated. If this procedure is called often (as is likely in a code that evolves the Schrödinger equation), it is likely to speed things up by using a much less elegant subroutine call, which adds the term $A\psi$ to `psidot`:

```
call add_A_on_psi(psidot)
```

The point is that the procedure should be inlined in either case, but when the new feature (array-valued functions), the optimizer support won't be as good. The same goes for user-defined data types, pointers, and virtually anything object-oriented. Of course, there are good reasons for the existence of these features: they

make your code more readable and elegant. There is usually a tradeoff between speed and general elegance when programming.

23.2.4 Parallel Programming

Optimizing parallel codes is an advanced topic, and way beyond the scope of this discussion, even as it becomes more relevant with multicore computers and wider availability of computing clusters. We'll just stick to crude, obvious observations here. Basically, parallel programming involves communication between processors. In the best case of *symmetric multiprocessor* (SMP) computing, the processors are all part of the same computer, or "box," and share the same memory. Communication is simple, since the processors have access to the same pool of memory. In the worst case of parallel processing, as in many clusters, the computers must talk over standard networks. In any case, communication is much slower than the crunching that happens on a single computer. Your goal is to minimize the communication, or at least organize it to make it as efficient as possible (e.g., into occasional, large "bursts" to minimize effects of network latency). Parallel programming is complex and subtle. The quantum optician who doesn't wish to spend all of his/her time writing code would do well to learn a data-parallel language such as High-Performance Fortran (HPF), where in many cases a well-written Fortran 90 code can be changed to a parallel code only by adding some compiler directives as comments in the Fortran 90 file.

23.2.5 Tuning Your Code

So, how do you increase the speed of your code? We'll only say a couple of superficial things here to get you started. The most important thing to do is to get a good compiler, and investigate all of its options. Good compilers support the above optimizations and many more, but usually only if you explicitly enable them. Good compilers will also let you write nice-looking, readable code without worrying about speed. For example, in this loop,

```
do j = 1, n
  a(j) = 2 * n * a(j)
end do
```

it is possible to save on almost half of the multiplications by precomputing the constant product `2*n` (the **loop invariant**):

```
two_n = 2 * n
do j = 1, n
  a(j) = two_n * a(j)
end do
```

However, this slightly obfuscates things, and anyway almost any worthwhile compiler will take care of this for you.

The basic strategy for writing a good, fast code is this:

- Write a *readable*, *working* code, not worrying about performance.
- Use a **profiler**, a program that runs your code and tells you where your code is spending its execution time. Usually it will give you something like a percentage of time in each subroutine. A free and commonly available program is `gprof`.
- If you identify one or a few routines in which the code is spending most of its time, then concentrate on optimizing those. If you can't identify any such routine, it may not be worth the effort. (If you work very hard to double the speed of a routine that accounts for 1% of the total execution time, the net speedup will be truly unimpressive!)
- If you have to resort to obfuscating a routine to speed it up, consider maintaining two versions: a readable one and a fast one. Or at least put the readable version in the comments.

- Make sure to test *each* change you make, to make sure it is really helping.

If possible, you should also use **hardware performance counters** to diagnose and characterize the performance of your code. Most processors have counters for performance-related events such as floating-point operations, cache misses, and so on. Unfortunately, tools for accessing these counters are often less convenient and less available than their profiling counterparts (in some operating systems, accessing the counters requires kernel-level access and thus recompiling the kernel to enable them).

Chapter 24

Ordinary Differential Equations

Here we'll cover some of the basic ideas for numerically solving ordinary differential equations, particularly for solving equations of motion for the dynamical evolution of a system. We won't be aiming for completeness in this vast area, but rather we will just cover some of the important and useful methods. First, however, we need a few mathematical preliminaries that will allow us to quantify how much of an error we're making when generating a numerical—and hence approximate—solution to an ODE.

24.1 Convergence

24.1.1 Sequences

First, let's start off with the notion of a sequence. A real-valued *sequence* is a function $x : \mathbb{Z}^+ \longrightarrow \mathbb{R}$, where $\mathbb{Z}^+$ denotes the set of positive integers. The sequence is commonly written using the subscript notation x_n. That is, the sequence associates a real number with every positive integer. The generalization of the real-valued sequence to complex values and n-tuples is obvious in the ideas of convergence to follow, so long as an appropriate "distance," such as the Euclidean metric, is used.

A sequence x_n is said to **converge** to x (denoted $x_n \longrightarrow x$ or $\lim_{n\to\infty} x_n = x$) if for every $\epsilon > 0$, there is an $N \in \mathbb{Z}^+$ such that if $n > N$, then $|x_n - x| < \epsilon$. In this case, x is the **limit** of the sequence x_n. If the sequence does not converge to a limit, then the sequence is said to **diverge**, in which case there are still a couple of possibilities. If for any $x \in \mathbb{R}$, there is an $N \in \mathbb{Z}^+$ such that if $n > N$, then $x_n > x$, then the sequence x_n is said to diverge to ∞. Similarly, if for the same conditions $x_n < x$, then the sequence x_n is said to diverge to $-\infty$. If the sequence neither converges nor diverges to $\pm\infty$, it is said to **oscillate**.

As a quick example, let's prove formally that the sequence $x_n = 1/n$ converges to zero. We start by letting $\epsilon > 0$. Choose N to be the smallest integer larger than $1/\epsilon$ so that $1/N < \epsilon$. If $n > N$, then $|x_n - 0| = x_n = 1/n < 1/N < \epsilon$. Thus $x_n \longrightarrow 0$.

24.1.2 *O* and *o*

Now supposing that we have a sequence, we can address the question of how the sequence behaves asymptotically by comparing the two sequences. Thus, let x_n and y_n be two sequences, not necessarily convergent. Then we write

$$y_n = O(x_n), \tag{24.1}$$

read as "y_n is **big Oh** of x_n" or "y_n is **of the order of** x_n," if there is a $K > 0$ and an $N \in \mathbb{Z}^+$, such that if $n > N$, then $|y_n/x_n| < K$. On the other hand, if the sequence $(y_n/x_n) \longrightarrow 0$, then y_n evidently converges more quickly than x_n, and we write

$$y_n = o(x_n). \tag{24.2}$$

This statement is read as "y_n is **little oh** of x_n," or "y_n is **dominated by** x_n." Note that the notation here can be deceptive, because the statement $x_n = O(y_n)$ does not imply that $y_n = O(x_n)$.

24.1.2.1 Example

To illustrate these definitions, consider the sequence $x_n = 3 + 3n + 5n^2$.

1. Then $x_n = O(n^2)$, because we can pick $N = 1$ and $K = 11$, and then $x_n/n^2 = (3/n^2 + 3/n + 5) < 3 + 3 + 5 = 11 = K$ for any $n > 1 = N$.
2. Similarly, $x_n = o(n^3)$, because $x_n/n^3 = (3/n^3 + 3/n^2 + 5/n) \longrightarrow 0$ as $n \longrightarrow \infty$.

24.1.3 Convergence and Scaling of Functions

These same notions of convergence also apply to limits of functions. In particular, when we say that

$$\lim_{x \to x_0} f(x) = y, \tag{24.3}$$

we mean that for every sequence $x_n \longrightarrow x_0$, the corresponding sequence $f(x_n) \longrightarrow y$. If it turns out that $y = f(x_0)$, then we say the function f is **continuous** at x_0. If we consider only sequences of numbers larger than the limit, $x_n - x = |x_n - x|$, then the same definition gives the **limit from above**

$$\lim_{x \to x_0^+} f(x) = y^+, \tag{24.4}$$

while if we consider only sequences of numbers smaller than the limit, $x - x_n = |x - x_n|$, then the same definition gives the **limit from below**

$$\lim_{x \to x_0^-} f(x) = y^-, \tag{24.5}$$

assuming these limits exist. The function is then continuous if the limits agree, $y^+ = y^-$.

The concepts of O and o can similarly be extended to functions of a continuum. In particular, we can write

$$f(x) = O[g(x)] \quad \text{as} \quad x \longrightarrow x_0 \tag{24.6}$$

if and only if for every convergent sequence $x_n \longrightarrow x_0$, $f(x_n) = O[g(x_n)]$. In particular, this is equivalent to the condition that each sequence $|f(x_n)/g(x_n)|$ is eventually bounded,

$$\lim_{n \to \infty} \left| \frac{f(x_n)}{g(x_n)} \right| < \infty. \tag{24.7}$$

and thus we may equivalently write

$$\lim_{x \to x_0} \left| \frac{f(x)}{g(x)} \right| < \infty. \tag{24.8}$$

For the little o, we can similarly write

$$f(x) = o[g(x)] \tag{24.9}$$

if and only if for every divergent sequence $x_n \longrightarrow \infty$, $f(x_n) = o[g(x_n)]$. From the above definitions of convergence, this latter statement is equivalent to the statement that $\lim_{x \to \infty} f(x)/g(x) = 0$. Note that there is only a sense of ''little oh'' dominance of functions as the argument increases without bound.

24.1.3.1 Truncation Error

The O notation is particularly useful for representing the error in a truncated series expansion, or the **truncation error**. That is, suppose we write the series expansion of the exponential function near $x = 0$ as

$$e^{-x} = 1 - x + \frac{x^2}{2!} + O(x^3). \tag{24.10}$$

Since the remaining terms in the expansion are $-x^3/3!$, $x^4/4!$, and so on, we can compactly represent them by simply noting that all the unwritten terms are $O(x^3)$. This gives a rough idea of how the truncation error scales with x. The same idea obviously applies to asymptotic expansions about the point at ∞.

24.1.3.2 Example

Again, as a simple example, consider $f(x) = 2 + 5x + 7x^2$.

1. As $x \longrightarrow 0$, $f(x) = O(x^0)$, because $f(x)/x^0 = 2 + 5x + 7x^2 \longrightarrow 2$ as $x \longrightarrow 0$.
2. As $x \longrightarrow \infty$, $f(x) = O(x^2)$, because $f(x)/x^2 = 2/x^2 + 5/x + 7 \longrightarrow 7$ as $x \longrightarrow \infty$.
3. Also, $f(x) = o(x^3)$ because $f(x)/x^3 = 2/x^3 + 5/x^2 + 7/x \longrightarrow 0$ as $x \longrightarrow \infty$.

24.2 Euler Methods

Consider an ordinary differential equation of the form

$$\dot{y}(t) = f(y(t), t) \tag{24.11}$$

that we wish to solve. The **Euler method** is the simplest way to generate the solution $y(t)$ to the ODE. The idea is to consider the Taylor expansion of the evolved solution

$$\begin{aligned} y(t + \Delta t) &= y(t) + \Delta t\, \dot{y}(t) + O(\Delta t^2) \\ &= y(t) + \Delta t\, f(y(t), t) + O(\Delta t^2). \end{aligned} \tag{24.12}$$

Thus, the Euler method consists of making the approximation

$$y(t + \Delta t) \approx y(t) + \Delta t\, f(y(t), t), \tag{24.13}$$
(Euler method)

to obtain the advanced solution $y(t + \Delta t)$, since the function f and the present solution $y(t)$ are known. This process is iterated to generate the further advanced solutions $y(t + 2\Delta t)$, $y(t + 3\Delta t)$, and so on. Pretty simple, and pretty easy to implement on a computer. This is the simplest example of a **finite-difference method**, since finite time steps are taken to approximate the continuous solution.

In a slightly more compact notation, we may write the Euler method as the recurrence equation

$$y_{n+1} = y_n + \Delta t\, f(y_n, t_n). \tag{24.14}$$
(Euler method)

Here, we have defined $t_n := n\Delta t$ and $y_n := y(t_n)$.

Of course, this same method can be applied to system of coupled ODEs. If we have a set of ODEs of the form

$$\dot{y}_\alpha(t) = f_\alpha(\mathbf{y}(t), t), \tag{24.15}$$

then the finite-step update equations are

$$y_\alpha(t + \Delta t) \approx y_\alpha(t) + \Delta t\, f_\alpha(\mathbf{y}(t), t), \tag{24.16}$$

which is a pretty obvious generalization of the scalar case. In this sense, the Euler method may be applied generally to any problem, since any ODE can be decomposed into a system of first-order ODEs (an nth-order ODE decomposes into n coupled, first-order ODEs). However, typically a somewhat fancier method is usually appropriate.

24.2.1 Local and Global Truncation Error

Evidently, the error in taking a single step in the Euler method is $O(\Delta t^2)$, as we see from Eq. (24.12). Thus, we say that the **local trunction error** of the Euler method is $O(\Delta t^2)$.

The **global truncation error** refers to the error in generating the solution over a fixed interval, say from 0 to t in time. This takes $N = t/\Delta t$ steps, and in the worst case when the local truncation errors add,

the accumulated error in taking the N steps is $O(N\,\Delta t^2) = O(\Delta t)$. This is a reasonable assumption, since the errors are not random, but are predictably related to the form of $f(y(t), t)$ and the solution $y(t)$. In any case, in principle the solution converges to the true one as $\Delta t \longrightarrow 0$. The Euler method is said to be a **first-order** method, meaning that the local truncation is *correct* to first order in Δt, or that the global truncation error is first order in Δt.

24.2.2 Implicit Euler Method and Stiff Equations

An alternative to the expansion in Eq. (24.12) is to expand $y(t)$ about $t + \Delta t$:

$$\begin{aligned} y(t) &= y(t + \Delta t) - \Delta t\, \dot{y}(t + \Delta t) + O(\Delta t^2) \\ &= y(t + \Delta t) - \Delta t\, f(y(t + \Delta t), t + \Delta t) + O(\Delta t^2). \end{aligned} \tag{24.17}$$

Thus, an alternative stepping scheme is the **implicit Euler method**, given by rearranging the expansion as

$$y(t + \Delta t) \approx y(t) + \Delta t\, f(y(t + \Delta t), t + \Delta t). \tag{24.18}$$
(implicit Euler method)

In the compact notation mentioned above, we can also write this as

$$y_{n+1} = y_n + \Delta t\, f(y_{n+1}, t_{n+1}). \tag{24.19}$$
(implicit Euler method)

This is the same as the Euler method in Eq. (24.13), except that the derivative function is evaluated at the advanced time, $f(y(t + \Delta t), t + \Delta t)$, instead of at the old time, $f(y(t), t)$. The problem with this is that evolving $y(t)$ to $y(t + \Delta t)$ requires *knowing* $y(t + \Delta t)$ already, which is of course why the method is implicit. There are a few ways to solve this equation at each time step. For example, the most straightforward is to use $y(t)$ as a guess for $y(t + \Delta t)$, and plug it into $f(y(t + \Delta t), t + \Delta t)$. Then generate the next guess for $y(t + \Delta t)$. Plug it back in, and keep guessing until the process converges to a solution, which is the one you want. This procedure is called **fixed-point iteration**, since the iteration converges to the desired steady state, or fixed point.

Obviously, the implicit Euler method is a lot more work than the explicit counterpart, so why bother? What we gain in exchange for the extra effort is improved *stability* properties. Consider the simple model problem

$$\dot{y} = -\alpha\, y, \tag{24.20}$$

where α is some large, positive constant. Obviously the solution is simply

$$y(t) = y_0\, e^{-\alpha t}. \tag{24.21}$$

The (explicit) Euler method gives the update method

$$y_{n+1} = y_n - \Delta t\, \alpha\, y_n = (1 - \alpha \Delta t)\, y_n. \tag{24.22}$$

Clearly, this method has rather serious problems if $\alpha \Delta t > 1$, since the solution y_1 will be negative if y_0 is positive, which definitely shouldn't happen: the the analytic solution says the initially positive solution should stay that way. It's pretty clear that in fact the solution will oscillate about zero, changing sign on each iteration. Things get even worse if $\alpha \Delta t > 2$, since now the coefficient of y_n has a modulus of more than unity, and thus $|y_n|$ diverges to ∞. If α is large, then it may take very small step sizes Δt to obtain a stable recurrence. Even in more complicated systems of equations, the interesting dynamics may happen on relatively long time scales, but the step size may be limited to a short time interval by a fast decay (as in a set of rate equations with vastly different decay rates) to obtain a stable solution. Such a system of equations is called a **stiff system**.

The implicit Euler method helps here, since the recursion now becomes

$$y_{n+1} = y_n - \Delta t\, \alpha\, y_{n+1}, \tag{24.23}$$

which when solved for y_{n+1} becomes

$$y_{n+1} = \frac{y_n}{1 + \alpha \Delta t}. \tag{24.24}$$

This recurrence is clearly stable (and not oscillatory) for any $\Delta t > 0$, a distinct improvement over the explicit case. Of course, for a large step the solution still wouldn't be very *accurate*. Often it's the case that with an explicit method, a time step small enough to get an accurate solution also is small enough to guarantee stability. However, it sometimes helps to use an implicit method if a larger time step than would be explicitly stable gives an adequately accurate solution.

The implicit Euler method is also called the **backward Euler method**, while the explicit Euler method is called the **forward Euler method**. Again, the improved stability properties inherited by the implicit method here also generally apply to more complicated implicit methods. Note that when going to the effort of using an implicit method, it is probably worth using a better method than the simple implicit Euler method.

24.3 Runge–Kutta Methods

Now we will seek finite-difference methods for solving ODEs that are higher order, or that is to say methods that have global truncation errors that are $O(\Delta t^n)$ with $n > 1$. Of course, these methods will be more complicated, so again, why bother? Obviously, with a higher-order method, the step size Δt needed to generate a solution with a particular accuracy will be larger than for a lower-order method. Often it is the case that the number of steps required for a high-order method is drastically smaller than for a low-order method, and there is correspondingly a large overall savings in computer time to obtain a particular level of accuracy. Furthermore, with a low-order method, it may not even be *possible* to achieve decent accuracy with a low-order method. In principle, by making the step size very small, you could achieve any desired accuracy with the Euler method. However, when the finite differences start becoming comparable to machine epsilon, machine rounding errors will dominate any truncation errors, providing an effective accuracy bound for a given numerical precision. A high-order method helps here because the step size to achieve a good accuracy will be *much* larger than for the low-order method. The accumulation of many small differences is replaced by a more complicated sequence of larger finite differences, which are less susceptible to roundoff errors.

24.3.1 Second-Order Methods

To construct a second-order method, consider the following forward and backward Taylor expansions, expanded to higher order:

$$\begin{aligned} y(t + \Delta t) &= y(t) + \Delta t\, \dot{y}(t) + \frac{\Delta t^2}{2!}\, \ddot{y}(t) + \frac{\Delta t^3}{3!}\, \dddot{y}(t) + \frac{\Delta t^4}{4!}\, y^{(4)}(t) + O(\Delta t^5) \\ y(t - \Delta t) &= y(t) - \Delta t\, \dot{y}(t) + \frac{\Delta t^2}{2!}\, \ddot{y}(t) - \frac{\Delta t^3}{3!}\, \dddot{y}(t) + \frac{\Delta t^4}{4!}\, y^{(4)}(t) + O(\Delta t^5) \end{aligned} \tag{24.25}$$

Subtracting these, we find

$$y(t + \Delta t) - y(t - \Delta t) = 2\Delta t\, \dot{y}(t) + 2\frac{\Delta t^3}{3!}\, \dddot{y}(t) + O(\Delta t^5), \tag{24.26}$$

and now letting $\Delta t \longrightarrow t + \Delta t$ and then $\Delta t \longrightarrow \Delta t/2$,

$$y(t + \Delta t) = y(t) + \Delta t\, \dot{y}(t + \Delta t/2) + \frac{\Delta t^3}{4 \cdot 3!}\, \dddot{y}(t) + O(\Delta t^5). \tag{24.27}$$

Thus, the finite-difference method

$$\begin{aligned} y(t + \Delta t) &\approx y(t) + \Delta t\, f(y(t + \Delta t/2), t + \Delta t/2) \\ y_{n+1} &\approx y_n + \Delta t\, f(y_{n+1/2}, t_{n+1/2}) \end{aligned} \tag{24.28}$$

is $O(\Delta t^3)$ (i.e., locally second-order accurate). It also has what turns out to be the useful property that *all even powers in the expansion of the error term vanish*. This is because of the symmetry of the expression about the middle time $t_{n+1/2}$. However, it requires that we know the intermediate solution $y_{n+1/2}$, which of course we don't know, and we wouldn't even necessarily want to know it if we are trying to generate the solution samples y_n.

To deal with the intermediate value, we need to come up with an approximation for it. There are a number of ways to do this, each leading to different methods. One possibility is to take *half* of an Euler step to compute $y_{n+1/2}$ and then use this result in the formula (24.28). Thus noting the two formulae

$$\begin{aligned} y(t + \Delta t/2) &= y(t) + (\Delta t/2)\, f(y(t), t) + O(\Delta t^2) \\ y(t + \Delta t) &= y(t) + \Delta t\, f(y(t + \Delta t/2), t + \Delta t/2) + O(\Delta t^3), \end{aligned} \tag{24.29}$$

we can write

$$\begin{aligned} d_n &= y_n + (\Delta t/2)\, f(y_n, t_n) = y_{n+1/2} + O(\Delta t^2) \\ y_{n+1} &= y_n + \Delta t\, f(d_n, t_{n+1/2}) + O(\Delta t^3), \end{aligned} \tag{24.30}$$

since the error in using d_2 in place of $y_{n+1/2}$ is $O(\Delta t^2)$, and with the extra factor of Δt leads to an overall $O(\Delta t^3)$ error. This leads to the **midpoint method**, an example of a **second-order Runge–Kutta method**:

$$\begin{aligned} d_n &= y_n + \frac{\Delta t}{2}\, f(y_n, t_n) \\ y_{n+1} &= y_n + \Delta t\, f(d_n, t_{n+1/2}). \end{aligned} \tag{24.31}$$

(Runge–Kutta method, order 2)

Again, the local truncation error is $O(\Delta t^3)$, while the global truncation error is $O(\Delta t^2)$. There are other Runge–Kutta methods that achieve second-order accuracy, which involve different choices for intermediate steps.

24.3.1.1 Variations

The above choice of the second-order Runge–Kutta method is by no means unique.[1] We can parameterize second-order Runge–Kutta methods more generally as

$$\begin{aligned} d_n &= y_n + (a\, \Delta t)\, f(y_n, t_n) \\ y_{n+1} &= y_n + b_1\, \Delta t\, f(y_n, t_n) + b_2\, \Delta t\, f(d_n, t_n + c\Delta t). \end{aligned} \tag{24.32}$$

(general Runge–Kutta method, order 2)

That is d_n is the result y_{n+a} of taking an Euler step of $a\, \Delta t$. The final solution y_{n+1} is a linear combination of taking a normal Euler step from (y_n, t_n) and an Euler step from (y_{n+a}, t_{n+c}). To see how we can obtain more second-order methods, we can write out the Taylor expansions of these evolution equations as

$$\begin{aligned} d_n &= y_n + a\, \Delta t\, f(y_n, t_n) \\ y_{n+1} &= y_n + b_1\, \Delta t\, f(y_n, t_n) \\ &\quad + b_2\, \Delta t\, f(y_n, t_n) + b_2\, c\, \Delta t^2\, \partial_t f(y_n, t_n) + b_2\, a\, \Delta t^2\, f(y_n, t_n)\, \partial_y f(y_n, t_n) + O(\Delta t^3). \end{aligned} \tag{24.33}$$

Comparison to the *exact* Taylor expansion of y_{n+1}

$$\begin{aligned} y_{n+1} &= y_n + \Delta t\, \dot{y}_n + \frac{\Delta t^2}{2} \ddot{y}_n + O(\Delta t^3) \\ &= y_n + \Delta t\, f(y_n, t_n) + \frac{\Delta t^2}{2} \partial_t f(y_n, t_n) + \frac{\Delta t^2}{2} f(y_n, t_n)\, \partial_y f(y_n, t_n) + O(\Delta t^3), \end{aligned} \tag{24.34}$$

[1] For a detailed treatment, see Arieh Iserles, *A First Course in the Numerical Analysis of Differential Equations*, (Cambridge, 1996).

and matching terms up to $O(\Delta t^2)$ leads to the conditions

$$b_1 + b_2 = 1, \qquad b_2\, c = \frac{1}{2}, \qquad a = c.$$

(order 2 Runge–Kutta conditions) (24.35)

Any method of the form (24.32) with coefficients satisfying these three conditions is second order, which clearly leaves much wiggle room. Thus, we may regard the general order-two method as parameterized by the single parameter a:

$$\begin{aligned} d_n &= y_n + (a\,\Delta t)\, f(y_n, t_n) \\ y_{n+1} &= y_n + \left(1 - \frac{1}{2a}\right) \Delta t\, f(y_n, t_n) + \frac{1}{2a} \Delta t\, f(d_n, t_n + a\Delta t). \end{aligned}$$

(general Runge–Kutta method, order 2) (24.36)

The midpoint method (24.31) corresponds to $a = 1/2$. The choice that minimizes the coefficient of the $O(\Delta t^3)$ error turns out to be[2] $a = 2/3$.

24.3.2 Fourth-Order and General Schemes

To construct higher-order general schemes, recall that to evolve y_{n+1} from y_n, we are performing the integral from t_n to t_{n+1}:

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} dt'\, f(y(t'), t'). \tag{24.37}$$

In constructing the second-order method above, we chose

$$\int_{t_n}^{t_{n+1}} dt'\, f(y(t'), t') = f(y_{n+1/2}, t_{n+1/2})\, \Delta t + O(\Delta t^3), \tag{24.38}$$

as we showed in Eq. (24.27). The midpoint approximation is one of the simplest choices to represent the integral. We could of course sample the interval (t_n, t_{n+1}) with additional points, hoping to approximate the integral more exactly using a higher-resolution discrete sum, thus ending up with higher order methods. Thus, we can for example write

$$y_{n+1} \approx y_n + \Delta t \sum_\alpha b_\alpha\, f(y(t_n + c_\alpha\, \Delta t), t_n + c_\alpha\, \Delta t), \tag{24.39}$$

where the coefficients b_α and c_α are to be determined to give the best possible approximation to the integral. Of course, more terms in the sum should lead to higher-order methods. And, of course, the problem with this scheme are that the solution at the intermediate times $t_n + c_\alpha\, \Delta t$ is not known, and thus must be estimated numerically.

Following the idea from the second-order case, we write down approximations for each of the intermediate solutions, and then compute a linear combination of the intermediate solutions to obtain an accurate estimate for y_{n+1}. Choosing $c_1 = 0$ to include the (known) initial point in the linear combination, we may

[2] Anthony Ralston, "Runge–Kutta Methods with Minimum Error Bounds," *Mathematics of Computation* **16**, 431 (1962).

write down the general form for the Runge–Kutta methods as

$$\begin{aligned}
d_1 &= y_n \\
d_2 &= y_n + a_{21}\,\Delta t\, f(y_n, t_n) \\
d_3 &= y_n + a_{31}\,\Delta t\, f(y_n, t_n) + a_{32}\,\Delta t\, f(d_2, t_n + c_2\,\Delta t) \\
d_4 &= y_n + a_{41}\,\Delta t\, f(y_n, t_n) + a_{42}\,\Delta t\, f(d_2, t_n + c_2\,\Delta t) + a_{43}\,\Delta t\, f(d_3, t_n + c_3\,\Delta t) \\
&\vdots \\
d_m &= y_n + \Delta t \sum_{\alpha=1}^{m-1} a_{m\alpha}\, f(d_\alpha, t_n + c_\alpha\,\Delta t) \\
y_{n+1} &= y_n + \Delta t \sum_{\alpha=1}^{m} b_\alpha\, f(d_\alpha, t_n + c_\alpha\,\Delta t).
\end{aligned} \tag{24.40}$$

This set of formulae is commonly rewritten in the form

$$\begin{aligned}
k_1 &= \Delta t\, f(y_n, t_n) \\
k_2 &= \Delta t\, f(y_n + a_{21}k_1, t_n + c_2\,\Delta t) \\
k_3 &= \Delta t\, f(y_n + a_{31}k_1 + a_{32}k_2, t_n + c_3\,\Delta t) \\
k_4 &= \Delta t\, f(y_n + a_{41}k_1 + a_{42}k_2 + a_{43}k_3, t_n + c_4\,\Delta t) \\
&\vdots \\
k_m &= \Delta t\, f\left(y_n + \sum_{\alpha=1}^{m-1} a_{m\alpha}, t_n + c_m\,\Delta t\right) \\
y_{n+1} &= y_n + \sum_{\alpha=1}^{m} b_\alpha k_\alpha,
\end{aligned}$$

(general form for Runge–Kutta methods) (24.41)

which is somewhat more efficient in terms of coding.

The coefficients for fairly low-order systems can be chosen by matching the Taylor expansions of these formulae with the exact Taylor expansion, just as in the second-order case. This is cumbersome, and more elegant methods are available for constructing high-order methods.[3] Here we will just quote some results at fourth order, where there is a good compromise between complexity of the method and accuracy. One very popular method is the ''classical'' fourth-order Runge–Kutta method, which takes four intermediate steps and can be written explicitly as

$$\begin{aligned}
k_1 &= \Delta t\, f(y_n, t_n) \\
k_2 &= \Delta t\, f\left(y_n + \frac{k_1}{2}, t_n + \frac{\Delta t}{2}\right) \\
k_3 &= \Delta t\, f\left(y_n + \frac{k_2}{2}, t_n + \frac{\Delta t}{2}\right) \\
k_4 &= \Delta t\, f(y_n + k_3, t_n + \Delta t) \\
y_{n+1} &= y_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6}.
\end{aligned}$$

(classical fourth-order Runge–Kutta method) (24.42)

Again, it is possible to choose the coefficients to not only cancel errors up to fourth order, but also to

[3] Arieh Iserles, *op. cit.*

minimize the coefficient of the $O(\Delta t^5)$ error term.[4]

$$\begin{aligned} k_1 &= \Delta t\, f(y_n, t_n) \\ k_2 &= \Delta t\, f(y_n + a_{21}k_1, t_n + c_2\,\Delta t) \\ k_3 &= \Delta t\, f(y_n + a_{31}k_1 + a_{32}k_2, t_n + c_3\,\Delta t) \\ k_4 &= \Delta t\, f(y_n + a_{41}k_1 + a_{42}k_2 + a_{43}k_3, t_n + \Delta t) \\ y_{n+1} &= y_n + b_1k_1 + b_2k_2 + b_3k_3 + b_4k_4. \end{aligned}$$

(fourth-order Runge–Kutta) (24.43)

The coefficients for the time increments are the free parameters, and are given by

$$\begin{aligned} c_2 &= \frac{2}{5} &&\approx 0.4000\ 0000\ 0000\ 0000 \\ c_3 &= \frac{7}{8} - \frac{3\sqrt{5}}{16} &&\approx 0.4557\ 3725\ 4218\ 7894, \end{aligned}$$

(fourth-order Runge–Kutta coefficients) (24.44)

while the coefficients for the linear combination are

$$\begin{aligned} b_1 &= \frac{263 + 24\sqrt{5}}{1812} &&\approx 0.1747\ 6028\ 2262\ 6904 \\ b_2 &= \frac{125(1 - 8\sqrt{5})}{3828} &&\approx -0.5514\ 8066\ 2878\ 7329 \\ b_3 &= \frac{1024}{4869\sqrt{5} - 10038} &&\approx 1.2055\ 3559\ 9396\ 5235 \\ b_4 &= \frac{1}{9/2 + 3/\sqrt{5}} &&\approx 0.1711\ 8478\ 1219\ 5190, \end{aligned}$$

(fourth-order Runge–Kutta coefficients) (24.45)

and the coefficients for intermediate linear combinations of solutions are

$$\begin{aligned} a_{21} &= \frac{2}{5} &&\approx 0.4000\ 0000\ 0000\ 0000 \\ a_{31} &= \frac{3(476\sqrt{5} - 963)}{1024} &&\approx 0.2969\ 7760\ 9247\ 7536 \\ a_{32} &= \frac{5(757 - 324\sqrt{5})}{1024} &&\approx 0.1587\ 5964\ 4971\ 0358 \\ a_{41} &= \frac{2094\sqrt{5} - 3365}{6040} &&\approx 0.2181\ 0038\ 8225\ 9205 \\ a_{42} &= -\frac{975 + 3046\sqrt{5}}{2552} &&\approx -3.0509\ 6514\ 8692\ 9308 \\ a_{43} &= \frac{32(14595 + 6374\sqrt{5})}{240845} &&\approx 3.8328\ 6476\ 0467\ 0103. \end{aligned}$$

(fourth-order Runge–Kutta coefficients) (24.46)

This method produces an error bound about half that of the classical formula (24.42), though it takes somewhat more effort to program and a few extra arithmetic operations.

24.3.3 Implicit Runge–Kutta Methods

The Runge–Kutta methods we have considered so far are explicit. Looking for example at Eq. (24.40), the mth intermediate solution d_m only depends on the previous d_j where $j < m$. Of course this constraint can

[4] Anthony Ralson, *op. cit.*

be relaxed, and each d_m can depend on function evaluations at *every* other d_m. This generalization leads to **implicit Runge–Kutta** (Gauss–Legendre) methods. For example, the two-stage scheme

$$\begin{aligned} d_1 &= y_n + a_{11}\,\Delta t\, f(y_n, t_n + c_1\,\Delta t) + a_{12}\,\Delta t\, f(d_2, t_n + c_2\,\Delta t) \\ d_2 &= y_n + a_{21}\,\Delta t\, f(y_n, t_n + c_1\,\Delta t) + a_{22}\,\Delta t\, f(d_2, t_n + c_2\,\Delta t) \\ y_{n+1} &= y_n + b_1\,\Delta t\, f(d_1, t_n + c_1\,\Delta t) + b_2\,\Delta t\, f(d_2, t_n + c_2\,\Delta t). \end{aligned}$$

(two-stage implicit Runge–Kutta method) (24.47)

with coefficients

$$\begin{aligned} a_{11} &= \frac{1}{4} & a_{12} &= \frac{1}{4} - \frac{\sqrt{3}}{6} \\ a_{21} &= \frac{1}{4} + \frac{\sqrt{3}}{6} & a_{22} &= \frac{1}{4} \\ b_1 &= \frac{1}{2} & b_2 &= \frac{1}{2} \\ c_1 &= \frac{1}{2} - \frac{\sqrt{3}}{6} & c_2 &= \frac{1}{2} + \frac{\sqrt{3}}{6} \end{aligned}$$

(fourth-order implicit Runge–Kutta coefficients) (24.48)

turns out to have fourth-order accuracy.[5] This seems more compact than the four-stage explicit Runge–Kutta methods above, but of course there is extra complexity associated with functional iteration or whatever is used to deal with the implicit equations. However, the implicit method should be more stable and robust than the explicit counterpart, and better able to deal with stiff problems.

24.3.4 Adaptive Stepping

An important concept in numerical integration is *adaptive time stepping*. The idea is that, along with estimating the solution, we can also estimate the error in the solution. Then, assuming we have some goal to achieve for the local accuracy (truncation error), or **tolerance**, we can perform a sort of feedback to the step size to just meet the specified tolerance. In other words, if the error is too large, then the step is rejected, and a smaller step is taken. If the error estimate is below tolerance, then the next time step is chosen to be larger, so that the desired accuracy is achieved with the minimum level of work.

This strategy has two main advantages. First, it removes the problem of choosing a sensible step size from the user, since the algorithm handles this automatically. This reduces potential for user error and wasted cpu cycles. More importantly, it may be that in some solutions, the nature of the solution may be continually in flux, with long steps appropriate during slowly varying intervals, and short steps necessary during periods of intense action. If the quiet intervals are large, the time savings gained by using an adaptive method can be substantial.

In the case of Runge–Kutta methods, the methods we have outlined don't already have an error estimate built in. Thus, we must do extra work. For example, after computing y_{n+1} from y_n by one Runge–Kutta step, we could recompute it using *two* half-steps, using the same algorithm. This is the strategy used, for example, in Richardson extrapolation (Section 26.3). However, the more common method used with Runge–Kutta integrators is to use two different methods of different order. A very popular choice is called the (explicit) **Runge–Kutta–Fehlberg** method,[6] where a fourth-order and a fifth-order method are computed simultaneously on each time step, and the difference between the two acts as an estimate for the $O(\Delta t^5)$ truncation error of the fourth-order method. The Runge–Kutta–Fehlberg method cleverly arranges both methods to make use of the same set of function evaluations, to minimize the extra effort beyond the basic fourth-order method. Such methods are called **embedded**, because for example the fourth-order method is "embedded" within the fifth-order method. Because we know how the error scales, it is reasonably straightforward to decide how to rescale the step size to achieve the desired error tolerance.

[5] Arieh Iserles, *op. cit.*

[6] Erwin Fehlberg, "Low-Order Classical Runge-Kutta Formulas with Stepsize Control and their Application to some Heat Transfer Problems," NASA Technical Report R-315 (1969), Table III. See also Arieh Iserles, *op. cit.*, p. 84.

Adjusting the time step is fairly straightforward with explicit Runge–Kutta methods, because the solution y_{n+1} only depends on y_n. This property does *not* hold for the predictor-corrector methods below, and they are concomitantly much more complicated to code as adaptive-step methods.

The fixed-step Runge–Kutta methods are fairly straightforward to code, but the extra overhead involved with the adaptive methods makes them slightly more complicated to code. Fortunately, there are free, high-quality implementations available, such as RKSUITE on Netlib.[7]

Now the adaptive-step method will obviously generate a solution that is not uniformly sampled in time, whereas often a uniform sampling is desirable. There are two main ways to deal with this. One is to go ahead and generate the solution with nonuniform samples, and then use numerical interpolation (e.g., cubic splines) to transfer the solution to the desired set of sample times. Another popular method is to decide on the desired sample times in advance, and then call the integrator to evolve the solution forward from one sample time to the next. The integrator steps adaptively to the next time, and terminates at the desired output time by adjusting its final step to exactly hit the output time. The integrator subroutine is thus called once for each output time step (whereas in the former method, the integrator subroutine is called only once for the whole solution). The latter method is especially common in compiled languages (e.g., Fortran or C), as opposed to ''canned'' languages like Octave.

24.4 Multistep and Predictor–Corrector Methods

Another important class of higher-order numerical methods goes by the name of **multistep methods**. The idea is that, as in the Runge–Kutta methods, to achieve higher order you need to estimate the solution at multiple points in time. Runge–Kutta methods do this by estimating the solution at several intermediate times *between* t and $t+\Delta t$. Multistep methods, however, use the additional points in the *past*, say at $t-\Delta t$, $t-2\Delta t$, and so on, to achieve the higher order. We won't go into the derivation of these methods here,[8] but for example, the second-order **Adams–Bashforth method** is

$$y_{n+1} = y_n + \Delta t \left[\frac{3}{2} f(y_n, t_n) - \frac{1}{2} f(y_{n-1}, t_{n-2}) \right], \tag{24.49}$$

while the third-order method is

$$y_{n+1} = y_n + \Delta t \left[\frac{23}{12} f(y_n, t_n) - \frac{4}{3} f(y_{n-1}, t_{n-2}) + \frac{5}{12} f(y_{n-2}, t_{n-2}) \right]. \tag{24.50}$$

Common implicit versions of these methods are **Adams–Moulton methods** and **backward differentiation formulae**, which are better when improved stability is needed or when solving stiff problems.

Implementations of these methods can be complicated, especially in implementing adaptive-step versions. Since the next step may rely on several steps in the past, changing future time steps require either interpolating the past solution onto the new time grid or using much more complicated formulae that can handle nonuniform time steps. Further, imagine: how would you start a high-order multistep method, given only initial data $y(t_0)$? You could, for example step *backwards* using a low-order method (e.g., Euler) to generate the past steps. Sophisticated implementations start with the Euler method, and increase the order as the solution proceeds, adaptively choosing both the order of the method and the stepsize needed to maximize efficiency at the desired error tolerance.

Multistep methods find their biggest utility in *predictor–corrector methods*. The basic idea is to use two methods, one explicit and one implicit, such as an Adams–Bashforth and an Adams–Moulton method of the same order, and run them together. The explicit method (the predictor), generates the updated solution at the future time, which is then used as a starting point for iterating the implicit method (the corrector) to convergence. The difference between the two, or perhaps whether or not the iteration converges quickly enough, is used to determine whether or not the current step size is adequate. There are a number of different

[7] `http://www.netlib.org/ode/rksuite/`

[8] See Arieh Iserles, *A First Course in the Numerical Analysis of Differential Equations*, (Cambridge, 1996), Chapter 2.

strategies here, but these methods are again quite complicated to implement. Fortunately, high-quality, open-source integrators are available, such as the venerable ODEPACK[9] and others[10] at Netlib. These methods have the advantage of efficiently producing solutions of very high accuracy. However, they perform best on very smooth solutions. In cases where certain errors are involved in computing the derivative function (as when lookup tables are used), these methods may not perform so well. They may also not perform well in cases where the integrator must be restarted often, as when the solution has discontinuities (as in quantum-jump trajectories), since the restarting overhead can be substantial. In these cases it better to fall back to the less elegant but more robust Runge–Kutta methods.

24.5 Exercises

Problem 24.1

An alternate definition of the limit of a function is as follows. We say

$$y = \lim_{x \to x_0} f(x) \tag{24.51}$$

if for every $\epsilon > 0$, there is a $\delta > 0$ such that if $|x - x_0| < \delta$ then $|f(x) - f(x_n)| < \epsilon$. Essentially, this is saying that a sufficiently small perturbation to x can make an arbitrarily small perturbation to $f(x)$. Show that this "ϵ-δ" definition of the limit is equivalent to the definition in terms of limits of sequences.

[9] http://www.netlib.org/odepack/
[10] http://www.netlib.org/ode/

Chapter 25

Fourier Transforms

Fourier transforms are the basis for a number of powerful *analytical* tools for analyzing linear (and sometimes nonlinear) systems. But they are also important *numerical* tools, in part because they can be performed accurately and *very* efficiently. Knowing how to use them both analytically and numerically is a good way to get deep intuition into physical systems, in particular dynamical systems, and allows for the construction for sophisticated numerical analysis techniques.

There are numerous conventions for the Fourier transform, and we will discuss a few of them. For the sake of concreteness, we will consider the usual time-frequency convention for the Fourier transform $\mathscr{F}$ of a function $f(t)$

$$\mathscr{F}\left[f\right](\omega) \equiv \tilde{f}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{i\omega t} dt \qquad \text{(Fourier-transform definition)} \tag{25.1}$$

and the inverse Fourier transform

$$\mathscr{F}^{-1}\left[\tilde{f}\right](t) \equiv f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(\omega)\, e^{-i\omega t} d\omega. \qquad \text{(inverse-Fourier-transform definition)} \tag{25.2}$$

We will then discuss adaptations to other normalization conventions.

25.1 Sampling Theorem

A critical aspect of the numerical computation of a Fourier transform is adapting the integral transform to a finite sample of the data. The **sampling theorem** provides the basis for doing this, especially for understanding the errors involved in doing so. To understand the sampling theorem, suppose that the spectrum of $f(t)$ has compact support—that is, suppose that $\tilde{f}(\omega)$ vanishes for $|\omega| > \omega_{\rm c}$, where $\omega_{\rm c}$ is some ''cut-off frequency.'' Then defining the unit-rectangular-pulse function by

$$\mathrm{rect}(t) := \begin{cases} 1 & \text{if} \quad |t| < 1/2 \\ 1/2 & \text{if} \quad |t| = 1/2 \\ 0 & \text{if} \quad |t| > 1/2, \end{cases} \tag{25.3}$$

we can write the compact-support condition for the spectrum as

$$\tilde{f}(\omega) = \tilde{f}(\omega)\,\mathrm{rect}\left(\frac{\omega}{2\omega_{\rm c}}\right). \tag{25.4}$$

Recall (Section 17.1.2) that the convolution theorem for functions $f(t)$ and $g(t)$ reads

$$\mathscr{F}[f * g] = \mathscr{F}[f]\mathscr{F}[g], \tag{25.5}$$

where "$*$" denotes the convolution operation

$$(f * g)(t) := \int_{-\infty}^{\infty} f(t')g(t-t')\,dt'. \tag{25.6}$$

We will also later need the frequency-domain version, which reads

$$\mathscr{F}^{-1}\left[\tilde{f} * \tilde{g}\right] = 2\pi\mathscr{F}^{-1}\left[\tilde{f}\right]\mathscr{F}^{-1}\left[\tilde{g}\right] \tag{25.7}$$

for frequency-domain functions $\tilde{f}(\omega)$ and $\tilde{g}(\omega)$, where the extra factor of 2π is due to the same factor in the inverse transform (25.2). Applying the original form (25.5) of the convolution theorem to the compact-support condition (25.4), we may write

$$f(t) = \frac{\omega_c}{\pi}\operatorname{sinc}\omega_c t * f(t), \tag{25.8}$$
(prelude to sampling theorem)

where $\operatorname{sinc} x := \sin x/x$, and we have used the inverse Fourier transform

$$\mathscr{F}^{-1}[\mathrm{rect}(\omega)] = \frac{1}{2\pi}\operatorname{sinc}(t/2). \tag{25.9}$$

Thus, we see that $f(t)$ is an invariant under convolution with the sinc function.

25.1.1 Critical Sampling

Now suppose we discretely sample the function $f(t)$ at uniform time intervals Δt. That is, we represent $f(t)$ by the countable set of values

$$f_j := f(j\Delta t) = f(t_j), \tag{25.10}$$

where the sample times are $t_j = j\Delta t$. In particular, suppose that we represent the function $f(t)$ by the weighted comb function

$$f^{(\Delta t)}(t) := \sum_{j=-\infty}^{\infty} f_j\,\Delta t\,\delta(t - j\Delta t). \tag{25.11}$$

This is a function that (1) is determined only by the samples f_j, and (2) has the same coarse-grained area as $f(t)$, at least to $O(\Delta t^2)$. To compute the Fourier transform of this function, note that we may rewrite it as the product of $f(t)$ and the usual comb function:

$$f^{(\Delta t)}(t) = f(t)\,\Delta t\sum_{j=-\infty}^{\infty}\delta(t - j\Delta t). \tag{25.12}$$

Now using the fact that the Fourier transform of a comb function is a comb function,

$$\mathscr{F}\left[\sum_{j=-\infty}^{\infty}\delta(t-j\Delta t)\right] = \frac{2\pi}{\Delta t}\sum_{j=-\infty}^{\infty}\delta\left(\omega - \frac{2\pi j}{\Delta t}\right), \tag{25.13}$$

we can use the convolution theorem in the form (25.7) to write the Fourier transform of $f^{(\Delta t)}(t)$ as the *convolution* of $\tilde{f}(\omega)$ with a comb function:

$$\tilde{f}^{(\Delta t)}(\omega) = \tilde{f}(\omega) * \left[\sum_{j=-\infty}^{\infty}\delta\left(\omega - \frac{2\pi j}{\Delta t}\right)\right]. \tag{25.14}$$

This spectrum is periodic in frequency due to convolution with the comb function. Now suppose we choose the sampling interval Δt such that it is determined by the cut-off frequency by

$$\omega_c = \frac{\pi}{\Delta t}. \tag{25.15}$$

Then

$$\tilde{f}^{(\Delta t)}(\omega) = \tilde{f}(\omega) * \left[\sum_{j=-\infty}^{\infty} \delta\left(\omega - 2\omega_{\mathrm{c}} j\right) \right], \tag{25.16}$$

and we see that now the spacing between the ''teeth'' of the comb function is $2\omega_{\mathrm{c}}$, the same as the width of the (two-sided) spectrum $\tilde{f}(\omega)$. This means that, provided we are interested in the range $|\omega| < \omega_{\mathrm{c}}$, only one of the delta functions effectively contributes (in particular, $j = 0$), and thus

$$\tilde{f}^{(\Delta t)}(\omega) = \left[\tilde{f} * \delta \right](\omega) = \tilde{f}(\omega) \qquad (|\omega| < \omega_{\mathrm{c}}). \tag{25.17}$$

Another way to state this is that for *any* frequency, ω, the spectra satisfy

$$\tilde{f}^{(\Delta t)}(\omega)\ \mathrm{rect}\left(\frac{\omega}{2\omega_{\mathrm{c}}}\right) = \tilde{f}(\omega)\ \mathrm{rect}\left(\frac{\omega}{2\omega_{\mathrm{c}}}\right) = \tilde{f}(\omega).$$

(equivalence of sample-reconstructed and original spectra) (25.18)

Thus, we see that the Fourier transforms of $f(t)$ and $f^{(\Delta t)}(t)$ are the *same*. In other words, the samples f_j, where the sampling interval satisfies the condition (25.15) for **critical sampling** of the signal, are sufficient to completely reconstruct the spectrum $\tilde{f}(\omega)$, and thus the original function $f(t)$. This is essentially the content of the sampling theorem, without going into the precise conditions for its validity (as is usual in a physicist's treatment of the subject).

Of course, what this means is that a function whose spectrum has compact support is in some sense a very special function. However, so long as the weight of an *arbitrary* spectrum is very small outside the range $(-\omega_{\mathrm{c}}, \omega_{\mathrm{c}})$, the error in the reconstruction of the spectrum is correspondingly very small. The other situation in which this works is if the spectrum $\tilde{f}(\omega)$ is *periodic* in ω, since we can see from Eq. (25.16) that the *reconstructed* spectrum is periodic with period $2\omega_{\mathrm{c}}$. This makes sense, as we know from Fourier theory that a function that is either defined on a bounded interval or is periodic can be represented by a Fourier *series*, rather than a Fourier transform. Thus, in writing down Eq. (25.11), we have essentially written down the Fourier series for $\tilde{f}(\omega)$, but in the language of a Fourier transform.

25.1.2 Reconstruction

The sampling theorem also provides a direct way to ''reconstruct'' the original function from its samples. We assume the spectrum has compact support—otherwise, a periodic spectrum implies that the modulated comb (25.11) is in fact the *true* form of $f(t)$—and then combine the compact-support condition (25.4) with the reconstructed-spectrum condition (25.18) to write

$$\tilde{f}(\omega) = \tilde{f}^{(\Delta t)}(\omega)\ \mathrm{rect}\left(\frac{\omega}{2\omega_{\mathrm{c}}}\right). \tag{25.19}$$

Using the convolution theorem (25.5), we can thus write the inverse transform of this equation as

$$f(t) = f^{(\Delta t)}(t) * \left[\frac{\omega_{\mathrm{c}}}{\pi} \mathrm{sinc}\left(\omega_{\mathrm{c}} t\right) \right], \tag{25.20}$$

where we have again used the inverse Fourier transform (25.9). Writing this out in terms of the samples,

$$f(t) = \sum_{j=-\infty}^{\infty} f_j \Delta t \frac{\omega_{\mathrm{c}}}{\pi} \delta(t - t_j) * \mathrm{sinc}\left(\omega_{\mathrm{c}} t\right). \tag{25.21}$$

Again using the critical-sampling condition (25.15) and carrying out the convolution, this relation becomes the **Whittaker–Shannon sampling interpolation formula**:[1]

$$f(t) = \sum_{j=-\infty}^{\infty} f_j \,\mathrm{sinc}\,[\omega_c(t-t_j)] = \sum_{j=-\infty}^{\infty} f_j \,\mathrm{sinc}\,[\pi(t-t_j)/\Delta t] \tag{25.22}$$

(Whittaker–Shannon sampling interpolation formula)

The sinc functions are used to interpolate the function between the samples, and again, if the signal is bandwidth-limited and critically sampled, this reconstruction formula is *exact*. Note that the function $\mathrm{sinc}\,[\omega_c(t-t_j)]$ is zero for any t_k where $j \neq k$, and is unity at t_j, so the construction is obviously correct at the sample times. What is less obvious is the exactness of the formula *between* the sample times.

25.2 Discrete Fourier Transform

Any real-world numerical calculation of a Fourier transform will obviously require operations on *finite* data sets. By sampling the temporal signal, we have reduced the information from a function $f(t)$ on an *uncountable* set to a function f_j on a *countable* (discrete) set, but we must further reduce the information to a finite set. Of course, since sampling implied a spectrum of finite width, we can obtain a finite set simply by also sampling the frequency spectrum to obtain samples $\tilde{f}_k = \tilde{f}(k\Delta\omega)$ of the frequence spectrum at uniform frequency intervals $\Delta\omega$. This is equivalent to a truncation of the time samples, so that the time samples only occur within some frequency interval. Again, this will amount to the assumption that the temporal signal is either a pulse with compact support, or that it is periodic.

When time and frequency are both sampled, we can use the arguments above to impose constraints on the sample intervals and ranges. For example, as above, when the signal is temporally sampled with N points with interval Δt, we may regard the signal as extending from $t = 0$ to $t = 2t_{\max}$, where

$$t_{\max} = \frac{N}{2}\Delta t. \tag{25.23}$$

Note the factor of 2 here, since the sampling in both time and frequency imply the assumption that the sampled function $f(t)$ is periodic. Thus, we can also regard the signal as extending from $t = -t_{\max}$ to $t = t_{\max}$, which is why we have set up our notation this way. The sampling interval Δt, from our arguments in the last section, implies a maximum frequency

$$\omega_{\max} = \frac{\pi}{\Delta t}, \tag{25.24}$$

which is called the **Nyquist frequency**, or the largest frequency that is critically sampled. Thus, the spectrum extends in frequency from $\omega = -\omega_{\max}$ to $\omega = \omega_{\max}$. To have the same information, there will also be N samples in frequency, $N/2$ of which correspond to the range from $\omega = 0$ to $\omega_{\max}$, so that the frequency-sampling interval is

$$\Delta\omega = \frac{2\omega_{\max}}{N} = \frac{2\pi}{N\Delta t} = \frac{\pi}{t_{\max}}. \tag{25.25}$$

Thus, given the two free parameters, the total time $2t_{\max}$ for the sample and the total number of samples N, the above three formulae give the rest of the discretization parameters Δt, $\Delta\omega$, and $\omega_{\max}$.

To avoid all these details of time and frequency scales, the **discrete Fourier transform** (DFT) is defined to have the simple dimensionless form

$$F_k = \sum_{j=0}^{N-1} f_j e^{2\pi i j k/N}, \tag{25.26}$$

(discrete Fourier transform)

[1] The history of the sampling theorem is somewhat involved and unclear (convoluted?). For a nice overview see the Wikipedia entry, "Nyquist–Shannon sampling theorem," `http://en.wikipedia.org/wiki/Nyquist-Shannon_sampling_theorem`.

and essentially amounts to multiplying the f_j vector by the matrix $\exp(2\pi ijk/N)$. The inverse of this transform will obviously have the opposite sign in the exponential:

$$\hat{f}_j = \sum_{k=0}^{N-1} F_k e^{-2\pi ijk/N}.$$

(discrete inverse Fourier transform, unscaled) (25.27)

However, this transformation does not exactly invert the DFT, because the inverted numbers are a factor of N larger than the originals,

$$\hat{f}_j = N f_j. \tag{25.28}$$

This follows from the identity

$$\sum_{k=0}^{N-1} e^{2\pi ijk/N} e^{-2\pi ij'k/N} = N\delta_{jj'}, \tag{25.29}$$

which is obvious for $j = j'$, and in the case $j \neq j'$ can be seen to vanish because the left-hand side amounts to a sum over all Nth roots of unity. Thus, the inverse DFT is often defined with this factor of N:

$$f_j = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{-2\pi ijk/N}.$$

(discrete inverse Fourier transform, scaled) (25.30)

Unfortunately, there is no standard convention over which inverse, (25.27) or (25.30) is implemented as the "inverse Fourier transform." The unnormalized form (25.27) is commonly implented in low-level languages such as Fortran and C (a sensible choice since the user usually needs to multiply by extra normalization factors anyway), while the "normalized" inverse transform (25.30) is implemented in MATLAB and Octave. For any particular implementation of the DFT, you should consult the documentation, or just test the DFT, followed by an inverse DFT, on an array of ones to see if you get back the array with an extra factor of N, or if you get back the original array.

To map our desired integral Fourier transforms onto the above DFTs, we start by writing down the truncated form for the approximation (25.12) for $f(t)$, corresponding to N samples:

$$f^{(\Delta t, N)}(t) = f(t)\,\Delta t \sum_{j=0}^{N-1} \delta(t - j\Delta t). \tag{25.31}$$

The sum could equally well run over positive and negative times, since the function is effectively assumed to be periodic anyway. Now putting this into the Fourier-transform formula (25.1), we find

$$\begin{aligned} \tilde{f}(\omega) &\approx \int_{-\infty}^{\infty} f^{(\Delta t, N)}(t) e^{i\omega t} dt \\ &= \int_{-\infty}^{\infty} f(t)\,\Delta t \sum_{j=0}^{N-1} \delta(t - t_j) e^{i\omega t} dt \\ &= \sum_{j=0}^{N-1} f(t_j) e^{i\omega t_j}\,\Delta t. \end{aligned} \tag{25.32}$$

The last expression is the usual discrete approximation for an integral, applied to the Fourier integral. The approximation formula usually has $O(\Delta t^2)$ error, but in view of the sampling theorem, the error is really due just to the truncation, and can thus be much smaller than $O(\Delta t^2)$, as we will discuss below. In particular, at the N frequencies $\omega_k = k\Delta\omega$,

$$\tilde{f}_k \equiv \tilde{f}(\omega_k) = \sum_{j=0}^{N-1} f_j e^{i\omega_k t_j}\,\Delta t. \tag{25.33}$$

Noting that from Eq. (25.25), $t_j\omega_k = jk\Delta t\Delta\omega = 2\pi jk/N$.

$$\tilde{f}_k \equiv \tilde{f}(\omega_k) = \Delta t \sum_{j=0}^{N-1} f_j e^{2\pi ijk/N}. \tag{25.34}$$

Thus, apart from the factor of Δt, we have exactly that the DFT formula (25.26) approximates the Fourier integral, in the sense that

$$\tilde{f}_k = F_k\,\Delta t. \tag{25.35}$$

(DFT as approximation to Fourier integral)

so that we must multiply by Δt after applying the DFT to obtain the approximation to the Fourier integral.

Similarly, if we adapt the inverse transform (without the factor of N) to the spectrum samples $\tilde{f}_k$, we find

$$f_j = \frac{\Delta\omega}{2\pi} \sum_{k=0}^{N-1} \tilde{f}_k e^{-2\pi ijk/N}. \tag{25.36}$$

That is, putting in the $\tilde{f}_k$ for the F_k in the inverse DFT formula (25.27), the actual inverse transform samples are given by multiplying the results of the inverse DFT by $\Delta\omega$:

$$f_j = \hat{f}_j \frac{\Delta\omega}{2\pi}.$$

(inverse DFT as approximation to Fourier integral) (25.37)

Of course, the factor of 2π here is the same factor in the inverse Fourier integral (25.2).

25.2.1 Periodicity and Transform Ordering

One thing that should be clear from the above expressions is that the zero-frequency and zero-time components ($\tilde{f}_0$ and f_0, respectively) occur at the *beginning* of their respective arrays. This may seem odd, as the frequency spectrum has both positive and negative frequencies. Of course, the frequency spectrum is assumed by the DFT to be periodic with period $2\omega_{\max}$, which corresponds to a shift by N in the index. Thus, rather than thinking of the ordering

$$\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N-1} \tag{25.38}$$

where again $\tilde{f}_j = \tilde{f}(\omega_j) = \tilde{f}(j\Delta\omega)$, we can recognize that the right half of the array is "wrapped," and thus we may regard the elements in the same order as

$$\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N/2-1}, \tilde{f}_{-N/2}, \tilde{f}_{-N/2+1}, \ldots, \tilde{f}_{-2}, \tilde{f}_{-1}. \tag{25.39}$$

Here, we have implicitly assumed that N is even, which is almost always the case in practice. It is often convenient to have the zero frequency actually in the *center* of the array, in which case we simply swap the left and right halves of the array to obtain

$$\tilde{f}_{-N/2}, \tilde{f}_{-N/2+1}, \ldots, \tilde{f}_{-2}, \tilde{f}_{-1}, \tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N/2-1}. \tag{25.40}$$

Note the asymmetry here, since $f_{-N/2}$ appears at the left boundary, while $f_{N/2-1}$ appears at the right. However, the N-periodic nature of the discrete spectrum means that $\tilde{f}_{N/2} = \tilde{f}_{-N/2}$, so in principle we can copy the left boundary onto the right boundary.

The same ordering comments apply to the temporal signal. A straight signal out of your digital sampling oscilloscope would have the form

$$f_0, f_1, \ldots, f_{N-1} \tag{25.41}$$

where again $f_j = \tilde{f}(t_j) = \tilde{f}(j\Delta t)$. Of course, you can just plug this into the DFT formula and then get a spectrum of the form (25.38). On the other hand, you might have something like a correlation function

that satisfies $g(-\tau) = g^*(\tau)$, in which case it would be nice to impose this constraint explicitly by including the negative-time values. This is also easy to do, since the temporal array is also N-periodic (with period $N\Delta t = 2t_{\max}$ in time), so that the same array can be interpreted as

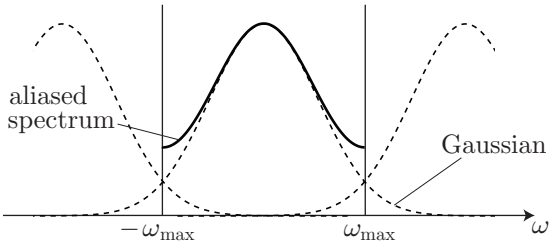
$$f_0, f_1, \ldots, f_{N/2-1}, f_{-N/2}, f_{-N/2+1}, \ldots, f_{-2}, f_{-1}. \tag{25.42}$$

For the correlation function, then we would impose the constraint $f_{-N} = f_N^*$. You can also swap the left and right halves to get a properly time-ordered array after all the DFT operations are done, and again it is still true that $f_{N/2} = f_{-N/2}$.

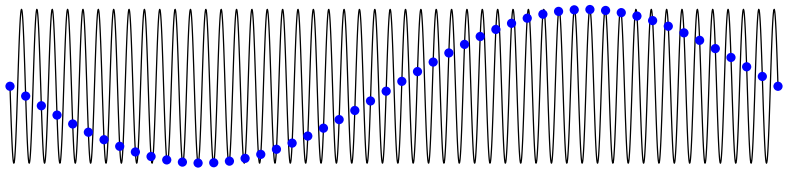
25.3 Aliasing

As we can see from the sampling theorem, the *only* error that we are introducing here is the fact that we are assuming $f(t)$ and $\tilde{f}(\omega)$ both have compact support (or equivalently, are periodic). If both had compact support, the DFT would be *exact* (at least up to rounding errors, if the DFT is performed on a computer). However, very few functions have compact support in both the time and frequency domain: remember that having compact support in one domain is equivalent to convolution with a sinc function in the other, and $\operatorname{sinc} x$ only falls off like $1/x$. Thus, the set of all functions for which the DFT can be exact is the set of all functions on a compact domain whose Fourier series are eventually zero (i.e., truncated). This set is obviously of zero measure in the space of all functions that we would want to transform.

So, then, what is the error associated with the DFT? For a general time signal $f(t)$ sampled with time interval Δt, in general the *true* spectrum $\tilde{f}(\omega)$ will have some contributions beyond the Nyquist frequency $\omega_{\max}$. Since the frequency spectrum is assumed by the DFT to be periodic with period $2\omega_{\max}$, the parts beyond $\omega_{\max}$ are spuriously "folded" into the computed spectrum. This effect is illustrated here for a Gaussian spectrum centered at $\omega = 0$.



Another way to see why this must be is that the DFT (25.26) is a unitary transformation on N points (except for the factor $N^{-1/2}$), and thus the total power of the sampled time signal must be equivalent to the total power of the sampled frequency spectrum—the discrete form of Parseval's theorem—whether or not the spectrum "fits" within the range $\pm\omega_{\max}$. Yet another good way to visualize this is to look at a harmonic wave that is *not* critically sampled. Recall that a frequency is critically sampled if it is sampled at least twice per period. However, if the sampling is below the critical rate, the samples will be indistinguishable from those from a lower-frequency harmonic wave (not to mention an infinity of yet higher-frequency harmonic waves).



This problem even occurs in the time domain in the laboratory on digital sampling oscilloscopes: if you're measuring a fast sin wave on such a scope, and you're seeing a wave with a frequency *way* lower than you expect, try cranking the "time/div" knob to see if you get a more appropriate (non-aliased) signal on a faster time scale. (Incidentally, this is the same effect that makes vibrating objects appear to be stationary or oscillate slowly with a strobe light, or that makes spinning wheels, propellers, and helicopter blades appear to precess slowly or even backwards on film or television.[2]) Of course, the same comments apply to starting

[2]For a good example, see `http://www.youtube.com/watch?v=eJ6vadFVjYg`.

off with a frequency spectrum, and then using the inverse DFT to obtain a time signal (say, a correlation function): aliasing can still occur in the time domain.

The point of all this is, to get accurate results with the DFT, say, when you are computing the spectrum of a pulse, you must pick your time grid such that the pulse and the spectrum fit well within $2t_{\max}$ and $2\omega_{\max}$, respectively. That is, near the outer boundaries of the ω and t grids, the signals in both domains should have fallen very close to zero, so you can neglect the power that falls outside the boundaries. Sometimes, you want to compute the DFT of a signal that *doesn't* fall off, such as a stationary, fluctuating signal. In this case you should still choose an appropriate sampling rate to avoid spectral aliasing, and realize that there may be some artifacts if the signal is not exactly periodic (as when sampling a sin wave, when the length of the sample does not match the period of the wave). The artifacts should be minor, however, so long as the sampling time is long compared to the correlation time of the signal.

25.4 Fast Fourier Transform

A main reason why the DFT is such an important computational tool is that it can be done so efficiently. From the DFT formulae (25.26) and (25.27), the DFT can be viewed as a multiplication of a matrix and a vector, and so for an array of length N, the computational effort (number of multiplications) should be $O(N^2)$. However, using a class of algorithms called **Fast Fourier Transforms** (FFTs) can do the same calculation in only $O(N \log N)$ operations. By far the most common algorithms are specific to the case where N is a power of 2, in which case the operation count is $O(N \log_2 N)$. Note that it is generally best to stick with these power-of-2 algorithms, and just live with this constraint on N: your data arrays can generally be interpolated or padded with zeros to get an appropriate array length. This savings in computational effort is huge: on the computer I'm typing on right now, an FFT of a half-million ($524288 = 2^{19}$), 64-bit data points takes about 0.3 s. Using the scalings above, a direct implementation of the DFT formula would take on the order of 2.4 hours for the same calculation! You might think that a couple of hours isn't so bad, but it is common for *many* FFTs to be needed in a single calculation. Also, because of the reduced operations count, round-off error does not accumulate nearly as much as in the straight DFT, and FFT algorithms are generally highly accurate. The FFT algorithms are so important that ''FFT'' is commonly used in place of ''DFT.''

The first FFT algorithm dates back to Gauss in 1805, but was not widely known until it was rediscovered by Cooley and Tukey[3] in 1965 (the Cooley-Tukey algorithm is a recursive method and remains a popular algorithm).[4] Many other algorithms beyond the Cooley-Tukey method are possible. We will not go into the details of various FFT algorithms,[5] since different implementations have different advantages, especially regarding accuracy and execution time on different architectures.[6] In this case, it is generally best to stick to algorithms written by specialists for reliability and good performance. Compiler vendors often provide highly optimized versions, but otherwise `FFTPACK`[7] and `FFTW`[8] are well-known libraries.

[3] James W. Cooley, John W. Tukey, ''An Algorithm for the Machine Calculation of Complex Fourier Series,'' *Mathematics of Computation* **19**, 297 (1965).

[4] For a nice overview of FFT algorithms, accuracy, and history, see the Wikipedia entry ''Fast Fourier transform,'' `http://en.wikipedia.org/wiki/Fast_Fourier_transform`. Another good source, particularly for history and algorithm details is the Wikipedia entry ''Cooley-Tukey FFT algorithm,'' `http://en.wikipedia.org/wiki/Cooley-Tukey_FFT_algorithm`.

[5] For the basic idea, see William H. Press, Brian P. Flannery, Saul A. Teukolsky, William T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. (Cambridge, 1992), Section 12.2, p. 498 (doi: 10.2307/2003354).

[6] For an example of how much a careful choice of method can make a big difference in execution time, see David H. Bailey, ''A High-Performance FFT Algorithm for Vector Supercomputers,'' *International Journal of Supercomputer Applications* **2**, 82 (1988) (doi: 10.1177/109434208800200106). An algorithm that performs reasonably well with large arrays on modern cache-based computers is the Stockham FFT (implemented in `FFTPACK`), detailed in Paul N. Swarztrauber, ''FFT algorithms for vector computers,'' *Parallel Computers* **1**, 45 (1984). For comparisons among many algorithms, see the benchFFT home page, `http://www.fftw.org/benchfft/`.

[7] `http://www.netlib.org/fftpack/`

[8] `http://www.fftw.org/`

25.5 Conventions

We have seen above how to use the general DFT formulae to construct numerical approximations to Fourier integral transforms. But keeping track of all the different sign, normalization, and ordering conventions can be a pain, so we go through a few examples here, considering bits of pseudocode appropriate for MATLAB/Octave and Fortran 90/95. For MATLAB/Octave we will consider the built-in functions `fft` and `ifft`, which implement the formulae (25.26) and (25.30), respectively. In Fortran, we will use a fictitious (but typical) subroutine `fft(isign, data)`, where `data` is the data array (both input and output), and `isign` specifies the sign of the transform exponent, so that `isign=1` specifies the "forward" transform (25.26), while `isign=-1` specifies the inverse transform (25.27). We will only consider one-dimensional transforms here, since the generalization to higher dimensions is reasonably obvious.

25.5.1 Temporal Signals

The simplest case we will consider is computing the energy spectrum $|\tilde{f}(\omega)|^2$ of a temporal signal $f(t)$, or the power spectrum $|\tilde{f}(\omega)|^2/T$, where T is the duration of the sample. The DFT formula (25.26) is easily adapted to this case for computing $\tilde{f}(\omega)$, since we simply need to multiply by Δt to obtain the correct scaling:

$$\tilde{f}_k = F_k \Delta t = \left[\sum_{j=0}^{N-1} f_j e^{2\pi i j k/N} \right] \Delta t. \tag{25.43}$$

The resulting spectrum array will have its zero frequency component first, with the negative frequencies in the last half of the array. The negative frequencies can be discarded in the power spectrum, since they will just be a mirror reflection of the positive frequencies. In the MATLAB/Octave code below, we simply compute the Fourier transform of the signal array `sig` of length `N`, and multiply by `dt`. The result is stored in the temporary array `spec` (also of length `N`), and after the negative frequencies are discarded, the result is squared and scaled, and is then stored in `energyspec` or `powerspec`, which is of length `N/2`. The scaling factor is `N*dt`, which is the total time T of the sample, and we cast the vector scaling as a multiplication to avoid many slower divisions. A frequency grid `w` is also generated for illustration, using Eqs. (25.23)-(25.25), so that the resulting array could be plotted with `plot(w, pwrspectrunc, '-')`.

```
% dt is the time sampling interval
spec = fft(sig) * dt;
energyspec = abs(spec(1:N/2)).^2;
powerspec  = abs(spec(1:N/2)).^2 * (1/(N*dt));
% w is the vector of frequencies
wmax = pi/dt;
dw = 2*wmax/N;
w = (0:dw:(wmax-dw))';
```

Below is the equivalent code in Fortran 90/95. Note that the variable declarations are not included, but some things to note: `wp` is a parameter declaring the floating-point precision (e.g., declare as

```
integer, parameter :: wp = selected_real_kind(p=14)
```

for double precision on IEEE machines); `sig` is a real (`kind=wp`) array of length `N`; `spec` is a complex (`kind=wp`) array of length `N`; `powerspec`, `energyspec`, and `w` are real (`kind=wp`) arrays of length `N/2`; and note the trick of using $\pi = 4\tan^{-1}(1)$.

```
! dt is the time sampling interval
spec(1:N) = sig(1:N) * dt
call fft( 1, spec)
energyspec(1:N/2) = abs(spec(1:N/2))**2
powerspec(1:N/2)  = abs(spec(1:N/2))**2 * (1/(N*dt))
! w is the vector of frequencies
pi = atan(1.0_wp)*4
```

```
wmax = pi/dt
dw = 2*wmax/N
w = (/ (real(j,kind=wp)*dw, j=0,N/2-1) /)
```

Typically then the values `w` and `pwrspectrunc` would then be dumped to a file or standard output for further processing.

In either case, to be accurate and avoid artifacts, the length of the sample of $f(t)$ should be well beyond the correlation time. This could be checked, for example, by computing the correlation function from the power spectrum (see below) and verifying that it decays to zero before the array ends.

25.5.2 Temporal Correlation Functions

As a more complicated example, suppose you have computed a correlation function $g(\tau)$ for $\tau \geq 0$, and now want to compute the corresponding power spectrum. Your data array is of length N. Before starting, to minimize artifacts it is useful to "enforce" the periodicity by pasting the complex-conjugated mirror image of the correlation function onto the end of the array, to obtain a "periodic" array of length $2N$. (There should be no discontinuity at the pasting point in the middle of the new array, because the samples should have decayed to zero by then.) Note one subtlety: we should paste all the original samples *except* the $t = 0$ sample, since we don't want to duplicate it, and this requires pasting in an extra zero. That's because the new array should be $2N$-periodic, and so the $t = 0$ (first) element should repeat in the $(2N + 1)$th place. Again, we can adapt the DFT formula (25.26) for this purpose,

$$\tilde{g}_k = G_k \Delta t = \left[\sum_{j=0}^{2N-1} f_j e^{2\pi i j k / 2N} \right] \Delta\tau, \tag{25.44}$$

so that again after computing the DFT we simply multiply by $\Delta\tau$. Note that we have $2N$ in place of the usual N, since we have doubled the array length before the transform. Also, to center the dc component in the middle of the array, we swap the two halves of the transform with the `fftshift` function. To recover the correlation function from the power spectrum samples $\tilde{g}_k$, we adapt the inverse DFT formula (25.30)

$$g_j = \left[\frac{1}{2N} \sum_{k=0}^{2N-1} \tilde{g}_k e^{-2\pi i j k / 2N} \right] \frac{2N\Delta\omega}{2\pi}. \tag{25.45}$$

The difference here is that we need to multiply by $\Delta\omega$ instead of $\Delta\tau$ for the frequency integral, undo the factor of $2N$ in the inverse DFT, and divide by 2π for the time-frequency normalization convention of Eq. (25.2). This is the proper approach in MATLAB/Octave, where the `ifft` function implements the bracketed transform in Eq. (25.45).

```
% dtau is the time sampling interval
gext = [g(1:N); 0; conj(g(N:-1:2))];
pwrspec = fftshift(fft(gext)) * dtau;
% w is the vector of frequencies
wmax = pi/dtau;
dw = 2*wmax/(2*N);
w = (-wmax:dw:(wmax-dw))';
% to recover g from pwrspec
gext = ifft(fftshift(pwrspec))*2*N*dw/(2*pi);
g = gext(1:N);
```

In Fortran 90/95, though, typical implementations leave out the factor of $2N$, so we are performing an inverse DFT as in Eq. (25.27):

$$g_j = \left[\sum_{k=0}^{2N-1} \tilde{g}_k e^{-2\pi i j k / 2N} \right] \frac{\Delta\omega}{2\pi}. \tag{25.46}$$

Thus, we need not multiply by $2N$ after the DFT. Also, note that swapping the two halves of the frequency array is conveniently performed with the Fortran intrinsic `cshift` (cyclic array shift, specifically for half the array length).

```
! dtau is the time sampling interval
pwrspec = 0
pwrspec(1:N) = g(1:N)
pwrspec(N+2:2*N) = conjg(g(N:2:-1))
call fft( 1, pwrspec)
pwrspec = cshift(pwrspec, N) * dtau
! w is the vector of frequencies
pi = atan(1.0_wp)*4
wmax = pi/dtau
dw = 2*wmax/(2*N)
w = (/ (real(j-N,kind=wp)*dw, j=0,2*N-1) /)
! to recover g from pwrspec (destroys pwrspec in the process)
pwrspec = cshift(pwrspec, N)
call fft(-1, pwrspec)
g(1:N) = pwrspec(1:N) * dw / (2*pi)
```

In the above code, `pwrspec` and `g` are complex (`kind=wp`) arrays of length `2N`, unless you are using an FFT routine especially adapted for real inputs and outputs.

25.5.3 Standard Frequency

Often in DFT applications, we will want to deal with the standard frequency ν instead of the angular frequency $\omega = 2\pi\nu$, in which case the alternate Fourier transform becomes

$$\bar{f}(\nu) = \int_{-\infty}^{\infty} f(t) e^{i2\pi\nu t} dt, \tag{25.47}$$

and the inverse Fourier transform becomes

$$f(t) = \int_{-\infty}^{\infty} \bar{f}(\nu) e^{-i2\pi\nu t} dt. \tag{25.48}$$

That is, there is no longer the factor of $1/2\pi$ scaling the inverse transform, and now there are explicit factors of 2π in the exponents. Everything is the same as in the ω–t convention, except the Nyquist frequency is now

$$\nu_{\max} = \frac{\omega_{\max}}{2\pi} = \frac{1}{2\Delta t}, \tag{25.49}$$

and the frequency-sampling interval is

$$\Delta\nu = \frac{\Delta\omega}{2\pi} = \frac{\omega_{\max}}{\pi N} = \frac{1}{N\Delta t} = \frac{1}{2t_{\max}}. \tag{25.50}$$

Other than these changes in the frequency grid, the only other difference in the above code snippets is that the division by 2π after the inverse DFT should be omitted.

25.5.4 Wave Functions

To transform a wave function between position and momentum space—as is useful, for example, in implmenting split-operator methods for evolving the Schrödinger equation, as in Chapter 26—the conventions are a bit different than for time and frequency. To compute the Fourier transform of $\psi(x)$ to obtain the momentum-space version $\phi(p)$, the integral is

$$\phi(p) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \psi(x) e^{-ipx/\hbar} dx, \tag{25.51}$$

while the inverse Fourier transform is

$$\psi(x) = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} \phi(p) e^{ipx/\hbar} dp. \tag{25.52}$$

Note the symmetric normalization and the presence of the extra $\hbar$. Furthermore, the $x = 0$ and $p = 0$ components are generally at the center of the wave-function arrays, so the array halves must be swapped both before and after the transform. We can adapt the *inverse* DFT (25.30) for the Fourier transform here, because of the opposite sign convention of the phase factors,

$$\phi_k = \left[\frac{1}{N} \sum_{j=0}^{N-1} \psi_j e^{-2\pi ijk/N} \right] \frac{N\Delta x}{\sqrt{2\pi\hbar}}, \tag{25.53}$$

as well as the DFT (25.26) for the *inverse* Fourier transform,

$$\psi_j = \left[\sum_{k=0}^{N-1} \phi_k e^{2\pi ijk/N} \right] \frac{\Delta p}{\sqrt{2\pi\hbar}}. \tag{25.54}$$

Also, since $p/\hbar$ plays the role of frequency, the relations between the increments are as follows, if we take the number N of grid points and the extent of the grid from $-x_{\max}$ to $x_{\max}$ to be fixed:

$$\begin{aligned} \Delta x &= \frac{2x_{\max}}{N} \\ p_{\max} &= \frac{\hbar\pi}{\Delta x} \\ \Delta p &= \frac{2p_{\max}}{N} = \frac{2\pi\hbar}{N\Delta x}. \end{aligned} \tag{25.55}$$

Thus, the MATLAB/Octave code would read as follows:

```
% specify N and xmax for the x grid
% x is the x grid
dx = 2*xmax/N;
x = (-xmax:dx:(xmax-dx))';
phi = fftshift(ifft(fftshift(psi))) * N * dx / sqrt(2*pi*hbar);
% p is the vector of momenta
pmax = hbar*pi/dx;
dp = 2*pmax/N;
p = (-pmax:dp:(pmax-dp))';
% to recover psi from phi
psi = fftshift(fft(fftshift(phi))) * dp / sqrt(2*pi*hbar);
```

In Fortran 90/95, we again simply specify the sign of the exponent for the transform, and forget about extra factors of N.

```
! specify N and xmax for the x grid
! x is the x grid
dx = 2*xmax/N;
x = (/ (real(j-N/2,kind=wp)*dx, j=0,N-1) /)
phi = cshift(psi, N/2)
call fft(-1, phi)
phi = cshift(phi, N/2) * dx / sqrt(2*pi*hbar)
! p is the vector of momenta
pi = atan(1.0_wp)*4
pmax = hbar*pi/dx;
dp = 2*pmax/N;
```

```
p = (/ (real(j-N/2,kind=wp)*dp, j=0,N-1) /)
! to recover psi from phi
phi = cshift(psi, N/2)
call fft( 1, psi)
psi = cshift(psi, N/2) * dp / sqrt(2*pi*hbar)
```

In this code, `psi` and `phi` are obviously complex arrays of length `N`.

25.6 Discrete Wigner Transform

Recall from Section 4.3 that the Wigner transform of a wave function $\psi(x)$ is

$$W(x,p) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dx' e^{-ipx'/\hbar}\psi(x+x'/2)\psi^*(x-x'/2). \tag{25.56}$$

We can simply regard $W(x,p)$ as a (quantum) Fourier transform of the form

$$W(x,p) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{\infty} d\xi\, e^{-ip\xi/\hbar} f(\xi), \tag{25.57}$$

where the function to be transformed in

$$f(\xi) = \frac{1}{\sqrt{2\pi\hbar}}\psi(x+\xi/2)\psi^*(x-\xi/2). \tag{25.58}$$

However, because of the appearance of $x'/2$ in the argument of ψ, if $\psi(x)$ is sampled with N samples with an interval of Δx, the appropriate sampling interval to use for the Fourier transform is $\Delta\xi = n\Delta x$, where n is even and $n \geq 2$. Since $\psi(x)$ is assumed to fall off to zero at the ends of the sampling range, we can always pad ψ with zeros such that $f(\xi)$ is always defined from $-nx_{\max}$ to $nx_{\max}$, and thus that $f(\xi)$ is still represented by N samples. Then we have the following modified parameters for the grid for $W(x,p)$:

$$\begin{aligned} \Delta x &= \frac{2x_{\max}}{N} \\ p_{\max} &= \frac{\hbar\pi}{\Delta\xi} = \frac{\hbar\pi}{n\Delta x} \\ \Delta p &= \frac{2p_{\max}}{N} = \frac{2\pi\hbar}{N\Delta\xi} = \frac{2\pi\hbar}{Nn\Delta x}. \end{aligned} \tag{25.59}$$

The Fourier transform must be repeated for each point in the x grid from $-x_{\max}$ to $x_{\max} - \Delta x$ in steps of Δx. Of course, you can skip some of these position values, as when making a three-dimensional plot, it is generally best to keep the density of points the same in both the position and momentum directions. In any case, we can write the explicit formula for the discrete Wigner transform as

$$W(x_j,p_k) = \frac{\Delta\xi}{2\pi\hbar}\sum_{l=-N/2}^{N/2-1} e^{-2\pi ikl/N}\psi(x_j+l\Delta\xi/2)\psi^*(x_j-l\Delta\xi/2), \tag{25.60}$$

where $x_j = j\Delta x$ and $p_k = k\Delta p$ (with j and k running from $-N/2$ to $N/2+1$), and again $\Delta\xi = n\Delta x$ with the even integer $n \geq 2$. Note that with this ordering of l, the zero-momentum component is in the center of the array, so array-swapping operations as in Section 25.5.4 are necessary to use the DFT/FFT formulae to evaluate the summations here.

For the discrete Wigner transform, the MATLAB/Octave code would read as follows (`NN` corresponds to N above):

```
n = 4; % must be even and >= 2; controls aspect ratio of Wigner transform
NN = length(psi);
```

```
NN2 = 2^floor(log2(NN)+0.49);
if (NN ~= NN2), error('input length not a power of 2'); end

Wout = zeros(NN,NN);
for j=1:NN,
  % order of indices is p, x for faster access
  extent = floor(min(j-1,NN-j)*2/n);
  lbd = j - extent*n/2;
  ubd = j + extent*n/2;
  lbdp = NN/2 - extent + 1;
  ubdp = NN/2 + extent + 1;
  Wout(lbdp:ubdp,j) = psi(ubd:(-n/2):lbd) .* conj(psi(lbd:(n/2):ubd));
end %for j
Wout = fftshift(ifft(fftshift(Wout)));

% transpose to x,p order, if desired, and scale
Wout = real(Wout)'*(dx*n*NN/(2*pi*hbar));
```

In Fortran 90/95, we again refer to the fictitious routine `fft(isign, data)`. The wave-function input is in the array `psi` of length `NN`, and the output array `W` is a `real` array of `dimension(NN,NN)`. The intermediate-storage array `Wtmp` is `complex` and of `dimension(NN,NN)`. The variables `lbd`, `ubd`, `lbdp`, `ubdp`, `rstep`, and `fstep` are all of type `integer`, and `pi` is of type `real`.

```
NN = size(psi)
pi = 4*atan(-1.0_wp)
if ( iand(NN, NN-1) .ne. 0 ) then
  write(0,*) "Error: array length not power of 2"
  stop
end if
if ( size(W,1) .ne. NN .or. size(W,2) .ne. NN ) then
  write(0,*) "Error: input and output array sizes do not match"
  stop
end if
if ( iand(n, 1) .ne. 0 .or. n .lt. 2 ) then
  write(0,*) "Error: n not even and positive"
  stop
end if

Wtmp = 0;
do j = 1, NN
  ! order of indices is p, x for faster access
  ! do shifted copy
  extent = floor(min(j-1,NN-j)*2/n*(1+epsilon(1.0_wp)))
  lbd = j - extent*n/2
  ubd = j + extent*n/2
  lbdp = NN/2 - extent + 1
  ubdp = NN/2 + extent + 1
  rstep = -n/2
  fstep = n/2
  Wtmp(lbdp:ubdp,j) = psi(ubd:lbd:rstep) * conjg(psi(lbd:ubd:fstep))

  ! do FT
  Wtmp(:,j) = cshift(Wtmp(:,j), NN/2)
  call fft(-1, Wtmp(:,j))
  Wtmp(:,j) = cshift(Wtmp(:,j), NN/2)
end do

! transpose to x,p order and scale
```

```
W = transpose(real(Wtmp)) * (n*dx/(2*pi*hbar))
```

Chapter 26

Split-Operator Methods

The method of **operator splitting** is a general technique for making the evolution of a system numerically tractable.[1] In general, the evolution (e.g., the equations of motion) of a system can be represented in terms of an operator (e.g., a matrix for a linear system). In some situations there may exist approximate decompositions of the operator into parts that are particularly simple to compute, and this is the essence of operator splitting. The unitary evolution of a quantum system according to the Schrödinger equation is particularly suited for operator-splitting methods, since the unitary time-evolution operator can often be decomposed into products of simple operators.

26.1 Splitting the Unitary Evolution Operator

Consider the Schrödinger equation in one dimension for a particle in a potential $V(x)$:

$$i\hbar\partial_t\psi(x,t) = H\psi(x,t) = \left[\frac{p^2}{2m} + V(x)\right]\psi(x,t), \tag{26.1}$$

where, as usual,

$$p = -i\hbar\partial_x. \tag{26.2}$$

We will now develop the **split-operator exponential method** as the basis for generating the time-dependent solutions $\psi(x,t)$ for this Schrödinger equation. The general idea is to note that for the time-independent Hamiltonian, we can write the evolution of the wave function over a time Δt as

$$\psi(x,t+\Delta t) = e^{-iH\Delta t/\hbar}\psi(x,t). \tag{26.3}$$

In general, the matrix exponential of H is difficult to calculate, since it is not in general diagonal in either the position or the momentum representation. However, often—as is the case for the particle Hamiltonian—the Hamiltonian is of the general form

$$H(x,p) = T(p) + V(x), \tag{26.4}$$

so that it splits into two parts, each of which is diagonal in either the position or the momentum basis. From the Baker–Campbell–Hausdorff expansion[2] for arbitrary operators A and B,

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[[A,B],B] + \ldots\right), \tag{26.5}$$

we can see that

$$e^{-iH\Delta t/\hbar} = e^{-iT(p)\Delta t/\hbar}e^{-iV(x)\Delta t/\hbar} + O(\Delta t^2). \tag{26.6}$$

[1] Hans De Raedt, "Product Formula Algorithms for Solving the Time Dependent Schrödinger Equation," *Computer Physics Reports* **7**, 1 (1987) (doi: 10.1016/0167-7977(87)90002-5).

[2] R. M. Wilcox, "Exponential Operators and Parameter Differentiation in Quantum Physics," *Journal of Mathematical Physics* **8**, 962 (1967) (doi: 10.1063/1.1705306).

In fact, a *symmetrized* splitting is even more accurate:

$$e^{-iH\Delta t/\hbar} = e^{-iV(x)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x)\Delta t/2\hbar} + O(\Delta t^3). \tag{26.7}$$

Note that the errors here are regarded as asymptotic, since there is no guarantee that the BCH expansion converges. These factorizations (the second, more accurate one being our preference) imply the following numerical method to evolve the wave function $\psi(x,t)$ to $\psi(x,t+\Delta t)$:[3]

1. Compute the half-step for the spatial part:

$$\psi(x) \longrightarrow e^{-iV(x)\Delta t/2\hbar}\psi(x). \tag{26.8}$$

   The exponential operator is diagonal in the position representation, so this is easy.

2. Compute the Fourier transform $\phi(p)$ of the new wave function:

$$\psi(x) \longrightarrow \phi(p) = \mathscr{F}[\psi(x)]. \tag{26.9}$$

   In view of fast-Fourier-transform (FFT) numerical algorithms, this operation can be done efficiently.

3. Now apply the ''drift'' part of the evolution operator:

$$\phi(p) \longrightarrow e^{-iT(p)\Delta t/\hbar}\phi(p). \tag{26.10}$$

   The exponential operator is diagonal in the momentum representation, so this is also easy.

4. Compute the inverse Fourier transform $\psi(x)$ of the updated wave function:

$$\phi(p) \longrightarrow \psi(x) = \mathscr{F}^{-1}[\phi(p)]. \tag{26.11}$$

5. Finally, compute the last half-step for the spatial part, now that we are back in the representation where it is diagonal:

$$\psi(x) \longrightarrow e^{-iV(x)\Delta t/2\hbar}\psi(x). \tag{26.12}$$

This procedure is then repeated for many time steps Δt until the desired final time is reached. The *local* error in each step is $O(\Delta t^3)$, but because the number of steps until the final time is $O(\Delta t^{-1})$, the *global* error of the final solution is $O(\Delta t^2)$.

Note that this type of symmetrized splitting works for more general splittings. For example, the decomposition

$$H = H_1 + H_2 + H_3 \tag{26.13}$$

leads to the operator splitting

$$e^{-iH\Delta t/\hbar} = e^{-iH_1\Delta t/2\hbar} e^{-iH_2\Delta t/2\hbar} e^{-iH_3\Delta t/\hbar} e^{-iH_2\Delta t/2\hbar} e^{-iH_1\Delta t/2\hbar} + O(\Delta t^3). \tag{26.14}$$

It can be shown that this type of symmetric operator splitting has $O(\Delta t^3)$ error for decompositions of H into *any* finite set of terms H_j, and the overall accuracy is still preserved even if we can only approximately compute each H_j with an error of $O(\Delta t^3)$.

[3] This type of split-operator Fourier transform algorithm was first proposed by J. A. Fleck, Jr., J. R. Morris, and M. D. Feit, ''Time-Dependent Propagation of High Energy Laser Beams through the Atmosphere,'' *Applied Physics* **10**, 129 (1976); M. D. Feit, J. A. Fleck, Jr., and A. Steiger, ''Solution of the Schrödinger Equation by a Spectral Method,'' *Journal of Computational Physics* **47**, 412 (1982) (doi: 10.1016/0021-9991(82)90091-2).

26.2 Time-Dependent Potentials

To handle a *time-dependent* potential $V(x,t)$, the time-evolution operator has the same form except for the replacement

$$V(x)\Delta t \longrightarrow \int_t^{t+\Delta t} dt'\, V(x,t'), \tag{26.15}$$

since we recall from Section 4.2.1 that the unitary time-evolution operator (from time t to $t+\Delta t$) now has the form

$$U(t+\Delta t,t) = \exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right], \tag{26.16}$$

if we assume the Hamiltonian commutes with itself at different times. The simplest approximation is to take

$$\int_t^{t+\Delta t} dt'\, V(x,t') = V(x,\tau)\Delta t + O(\Delta t^2), \tag{26.17}$$

where we may take τ to be any time in the interval $[t,t+\Delta t]$ without changing the order of the approximation (that is, with the exception of $\tau = \Delta t/2$, which we return to below). However, the accuracy here suffers, as then the factorized unitary evolution operator becomes

$$\exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right] = e^{-iV(x,\tau)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x,\tau)\Delta t/2\hbar} + O(\Delta t^2). \tag{26.18}$$

It is better to use an approximation for the integral of $V(x,t)$ accurate to *second* order, as in the **trapezoidal rule**:

$$\int_t^{t+\Delta t} dt'\, V(x,t') = [V(x,t) + V(x,t+\Delta t)]\,\frac{\Delta t}{2} + O(\Delta t^3). \tag{26.19}$$

This expression then gives the more accurate splitting

$$\exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right] = e^{-i[V(x,t)+V(x,t+\Delta t)]\Delta t/4\hbar} e^{-iT(p)\Delta t/\hbar} e^{-i[V(x,t)+V(x,t+\Delta t)]\Delta t/4\hbar} + O(\Delta t^3). \tag{26.20}$$

However, it turns out to have the same order of accuracy if we instead use the simpler splitting

$$\exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right] = e^{-iV(x,t+\Delta t)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x,t)\Delta t/2\hbar} + O(\Delta t^3), \tag{26.21}$$

where we evaluate the potential at the beginning and end points of the time interval in the two spatial half-steps. This splitting follows from the BCH expansion, which implies that

$$e^{hA_1}e^{hB}e^{hA_2} = \exp\left(h(A_1+A_2+B) + \frac{h^2}{2}[A_1-A_2,B] + \frac{h^2}{2}[A_1,A_2] + O(h^3)\right) = e^{h(A+B)} + O(h^3), \tag{26.22}$$

where the last equality follows if $A_1 + A_2 = A + O(h^2)$, $A_1 - A_2 = O(h)$, and $[A_1, A_2] = 0$. Note that taking $\tau = \Delta t/2$ in the above first-order method above corresponds to the **midpoint rule** of integration:

$$\int_t^{t+\Delta t} dt'\, V(x,t') = V(x,t+\Delta t/2)\Delta t + O(\Delta t^3). \tag{26.23}$$

Thus, we may also take

$$\exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right] = e^{-iV(x,t+\Delta t/2)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x,t+\Delta t/2)\Delta t/2\hbar} + O(\Delta t^3), \tag{26.24}$$

or

$$\exp\left[-\frac{i}{\hbar}\int_t^{t+\Delta t} dt'\, H(t')\right] = e^{-iT(p)\Delta t/2\hbar} e^{-iV(x,t+\Delta t/2)\Delta t/\hbar} e^{-iT(p)\Delta t/2\hbar} + O(\Delta t^3), \tag{26.25}$$

as symmetric, second-order splittings.

26.3 Richardson Extrapolation

Now we will introduce the generally applicable method of **Richardson extrapolation**[4] to develop yet higher-order approximations. First, notice again that the *global* error associated with the splitting (26.21) is $O(\Delta t^2)$, one order lower than that of the *local* truncation error. We then take Richardson's *ansatz* that the global error of the numerical solution takes the form

$$\tilde{\psi}_{\Delta t}(x,t) - \psi(x,t) = e_2(x,t)\Delta t^2 + e_4(x,t)\Delta t^4 + e_6(x,t)\Delta t^6 + \ldots, \tag{26.26}$$

where $\psi(x,t)$ is the exact solution, $\tilde{\psi}_{\Delta t}(x,t)$ is the numerical approximation generated with time steps of Δt, and the $e_n(x,t)$ are functions independent of the time step Δt. Note that only the *even*-order terms appear here; we leave it as an exercise to give an argument for this, but it is important to note that this expansion is valid for evolution via the second-order, *symmetric* splitting above. Thus, by computing the solution at time t with *multiple* step sizes (say, Δt, $\Delta t/2$, $\Delta t/3$), we can devise the proper linear combinations of these solutions that will cancel the error terms, and thus generate methods of higher global order. For example, a fourth-order method follows from using step sizes of Δt and $\Delta t/2$,

$$\frac{4}{3}\tilde{\psi}_{\Delta t/2}(x,t) - \frac{1}{3}\tilde{\psi}_{\Delta t}(x,t) = \psi(x,t) + O(\Delta t^4). \tag{26.27}$$

The error term e_2 (which has been cancelled in the method here) can also be estimated by

$$\frac{4}{3}\left[\tilde{\psi}_{\Delta t}(x,t) - \tilde{\psi}_{\Delta t/2}(x,t)\right] = \Delta t^2\, e_2(x,t) + O(\Delta t^4). \tag{26.28}$$

A sixth-order method follows from using step sizes of Δt, $\Delta t/2$, and $\Delta t/3$:

$$\frac{1}{24}\tilde{\psi}_{\Delta t}(x,t) - \frac{16}{15}\tilde{\psi}_{\Delta t/2}(x,t) + \frac{81}{40}\tilde{\psi}_{\Delta t/3}(x,t) = \psi(x,t) + O(\Delta t^6), \tag{26.29}$$

in which case the smallest cancelled error term e_4 is estimated by

$$-\frac{13}{24}\tilde{\psi}_{\Delta t}(x,t) + \frac{32}{3}\tilde{\psi}_{\Delta t/2}(x,t) - \frac{81}{8}\tilde{\psi}_{\Delta t/3}(x,t) = \Delta t^4\, e_4(x,t) + O(\Delta t^6). \tag{26.30}$$

These methods follow from writing out the Richardson expansions for $\psi(t+\Delta t)$ in powers of Δt (for all the step sizes), setting up the equations as a linear system, and then solving them for the ''true'' solution and the error coefficients. Higher order methods are, of course, possible, but at some point become too computationally expensive to be worthwhile.

Thus, we can construct a fourth-order [i.e., global error of $O(\Delta t^4)$] method as follows:

1. Start with the simulated state $\tilde{\psi}(x,t)$ at time t.
2. Compute $\tilde{\psi}(x,t+\Delta t)$ using a *single* step of Δt, using the $O(\Delta t^3)$ operator splitting (26.21). Call this $\tilde{\psi}_{\Delta t}(x,t+\Delta t)$.
3. Compute $\tilde{\psi}(x,t+\Delta t)$ using a *two* steps of $\Delta t/2$, using the same operator splitting. Call this $\tilde{\psi}_{\Delta t/2}(x,t+\Delta t)$.
4. Combine the two results as in Eq. (26.27) to obtain the updated state:

$$\tilde{\psi}(x,t+\Delta t) = \frac{4}{3}\tilde{\psi}_{\Delta t/2}(x,t+\Delta t) - \frac{1}{3}\tilde{\psi}_{\Delta t}(x,t+\Delta t) + O(\Delta t^4). \tag{26.31}$$

Then iterate this method until the desired final time. The global error of the solution will then be $O(\Delta t^4)$. The sixth-order method is a straightforward generalization of this procedure (the same is true of higher-order methods). The advantage of the higher-order methods, of course, is the possibility for improved accuracy for a given step size, and thus a larger step size (and, ideally, less work) for a given accuracy goal. On the other hand, the second-order method is unitary (i.e., it explicitly preserves the norm of ψ), while the fourth- and sixth-order methods are not.

[4] L. F. Richardson, ''On the Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam,'' *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **83**, 335 (1910).

26.3.1 Numerical Test

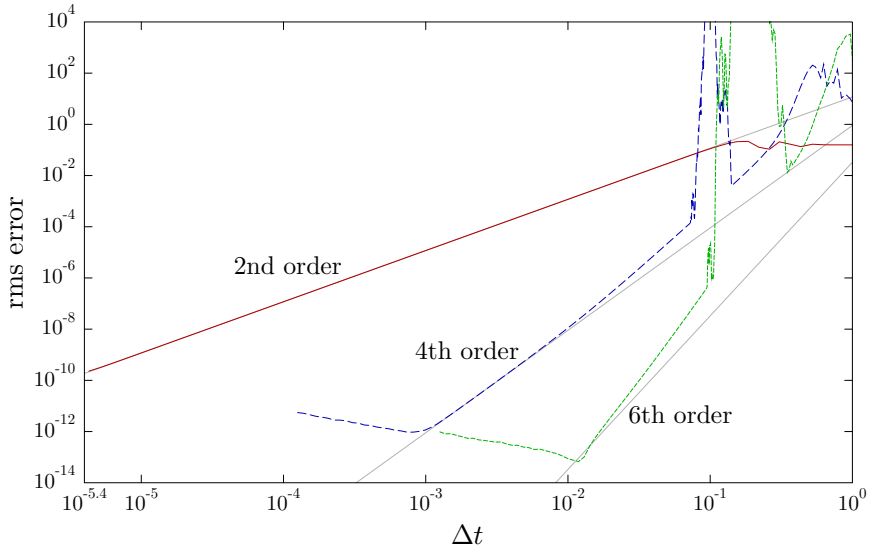
To illustrate these methods, we will try them out on the harmonic oscillator, with Hamiltonian

$$H = \frac{p^2}{2} + \frac{x^2}{2}, \tag{26.32}$$

using a Gaussian wave packet with initially $\sigma_x = \sigma_p = 1/\sqrt{2}$ ($\hbar = 1$ in these units), and the center of the wave packet is initially $x_0 = 0$ and $p_0 = 20$. The grid consisted of 1024 points, with a range in momentum of $\pm p_{\max} = \pm 16\sqrt{2\pi}$ and a grid spacing $\Delta p = \sqrt{2\pi}/32$ (with the same range and spacing for the position grid). The evolved wave packet at time $t = 4\pi$ (i.e., two complete oscillations) is compared with the initial wave packet, and according to the exact solution these should be exactly equal. Below is plotted the rms error as a function of the step size Δt, defined as

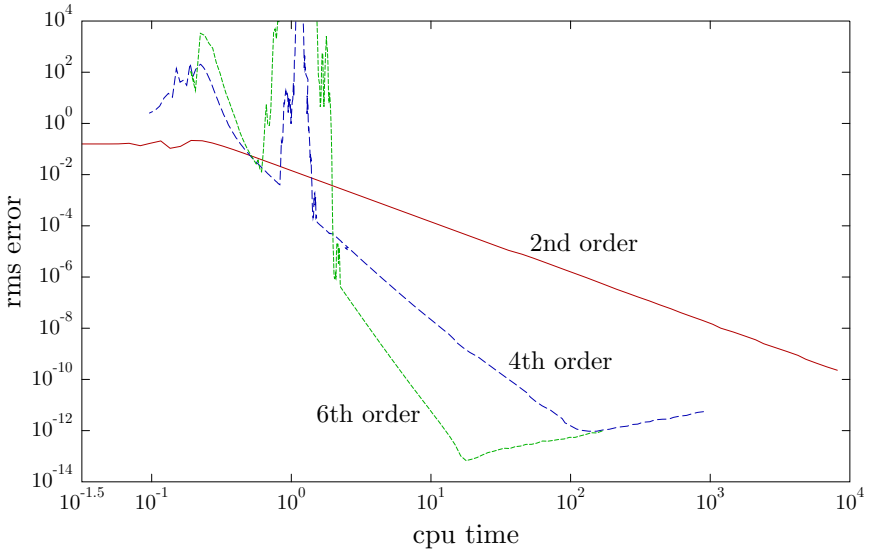
$$\delta\psi_{\text{rms}} = \frac{1}{N}\sqrt{\sum_{j=1}^{N} \left|\hat{\phi}_j - \phi_j\right|^2}, \tag{26.33}$$

where $\phi(p)$ is the exact momentum-space wave function, $\phi_j \equiv \phi(p_j)$ is the exact result on the grid, where p_j is the momentum of the jth grid point, $\hat{\phi}_j$ is the numerical approximation to ϕ_j, and N is the number of grid points.



The results for methods of orders 2, 4, and 6 are plotted along with lines to indicate scaling according to Δt^2, Δt^4, and Δt^6. From the plot, it is easy to see that the second-order method scales nicely as Δt^2 until the error is quite large. The fourth-order method also scales as Δt^4 once the error is sufficiently small. However, there is a strong deviation for large step sizes, where it is no longer possible to ignore the higher-order terms in the Taylor expansions, and the Richardson steps are far from unitary. For sufficiently small steps, the fourth-order scaling also breaks down as rounding errors dominate the rms error (note that the rounding error increases as the *number* of steps increases). For the same step, the fourth-order method also shows *far* better accuracy than the second-order method. The sixth-order method is qualitatively similar to the fourth-order method, with the same instability and roundoff behaviors, but there seems to be no regime in which the scaling is Δt^6. Strangely, the scaling seems to be somewhat *faster*. Note that the sixth-order method is correspondingly more accurate in the regime of "sensible" step sizes, and achieves a much better accuracy at the "sweet spot" of best accuracy compared to fourth-order.

Thus we see that for a given *step size*, a higher-order method performs much better than a lower-order counterpart, provided the step size is small enough. However, a higher-order step costs more than a lower-order step in terms of computer time, simply because it involves more substeps. Plotted below is the same performance data, but now the error is plotted as a function of the *execution time* on a computer.



Obviously, the execution time will vary widely depending on problem-specific optimizations, language and compilers, and other details. For this test, the methods were implemented in the Octave language, with no special effort to optimize the methods (i.e., substeps that could have been combined into larger steps to save time were not combined, and so on), to give the highest-order methods the least "advantage." The cpu times shown are nominally in seconds, but again this would vary widely depending on the hardware, language, and compiler used. However, it is clear that for a given (high) accuracy, the savings in cpu time in using a high-order method are enormous, potentially measured in orders of magnitude.

26.4 Unitary Evolvers

It is possible to construct higher-order, explicitly *unitary* split-operator methods. Let's start with the case of a time-independent Hamiltonian. Then we can write the second-order approximation for the unitary evolution operator $U(t, t+\Delta t) \equiv U(\Delta t)$ as

$$\tilde{U}(\Delta t) \equiv \tilde{U}^{(2)}(\Delta t) := e^{-iV(x)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x)\Delta t/2\hbar} = U(\Delta t) + O(\Delta t^3). \tag{26.34}$$

Then the idea is to look for ordered products that give order m approximations to the unitary evolution operator:

$$\tilde{U}^{(m)}(\Delta t) = \prod_j \tilde{U}(s_j\,\Delta t) = U(\Delta t) + O(\Delta t^{m+1}). \tag{26.35}$$

That is, to search for products that give better approximations ($m > 2$) to the evolution operator. Clearly, these approximations are still unitary.

26.4.1 Hierarchical Construction

There are multiple strategies for doing this, but one particularly useful strategy is the following hierarchical construction.[5] The construction is that given the $(m-1)$th approximant $\tilde{U}^{(m-1)}(\Delta t)$, we can construct the mth approximant by choosing

$$\tilde{U}^{(m)}(\Delta t) = \prod_{j=1}^{r} \tilde{U}^{(m-1)}(s_j\,\Delta t), \tag{26.36}$$

[5] Masuo Suzuki, "Fractal Decomposition of Exponential Operators with Applications to Many-Body Theories and Monte Carlo Simulations," *Physics Letters A* **146**, 319 (1990) (doi: 10.1016/0375-9601(90)90962-N); G. Dattoli, L. Giannessi, M. Quattromini, and A. Torre, "Symmetric decomposition of exponential operators and evolution problems," *Physica D* **111**, 129 (1998) (doi: 10.1016/S0167-2789(97)80008-5).

provided the step factors s_j satisfy the two conditions

$$\sum_{j=1}^{r} s_j = 1, \qquad \sum_{j=1}^{r} s_j^m = 0. \tag{26.37}$$

The first condition assures that the total effect of the approximant is to evolve the system by a time step Δt, since the approximant is the composition of r total substeps. The second condition ensures the cancellation of the order m error terms, which we can see by noting the form

$$U(\Delta t) = e^{-iH\Delta t/\hbar} = \prod_{j=1}^{r} \exp\left(-\frac{iH}{\hbar} s_j\, \Delta t\right) = \tilde{U}^{(m)}(\Delta t) + O(\Delta t^{m+1}) \tag{26.38}$$

of the evolution operator, and then substituting the $(m-1)$th approximant

$$\exp\left(-\frac{iH}{\hbar} s_j\, \Delta t\right) = \tilde{U}^{(m-1)}(s_j\, \Delta t) + g_m(H/\hbar)\,(s_j\, \Delta t)^m + O(\Delta t^{m+1}), \tag{26.39}$$

where g_m is an undetermined function representing the uncontrolled error at the mth order. We thus conclude that

$$U(\Delta t) = \prod_{j=1}^{r} \tilde{U}^{(m-1)}(s_j\, \Delta t) + g_m(H/\hbar) \sum_{j=1}^{r} (s_j\, \Delta t)^m + O(\Delta t^{m+1}) = \tilde{U}^{(m)}(\Delta t) + O(\Delta t^{m+1}), \tag{26.40}$$

in which case the order m error term vanishes if we require the conditions (26.37) to be satisfied.

In fact, recalling from the above argument that the global error expansion involves only *even* powers of Δt, it similarly follows that if $U(\Delta t)$ and the $\tilde{U}^{(m)}(\Delta t)$ are unitary, and the splittings are time-reversal symmetric,

$$\tilde{U}^{(m)}(\Delta t)\, \tilde{U}^{(m)}(-\Delta t) = 1, \tag{26.41}$$

then the *local* truncation error of $\tilde{U}^{(m)}(\Delta t)$ involves only *odd* powers of Δt. Thus, the recursive method steps *two* orders at a time, so long as the decompositions are symmetric:

$$\tilde{U}^{(m+1)}(\Delta t) = \prod_{j=1}^{r} \tilde{U}^{(m-1)}(s_j\, \Delta t). \tag{26.42}$$

This is true providing that (26.37) are satisfied as well as the symmetry condition

$$s_j = s_{j-r+1} \tag{26.43}$$

for all j.

To construct a specific method we can use the splitting

$$\tilde{U}^{(2m)}(\Delta t) = \tilde{U}^{(2m-2)}(s\, \Delta t)\, \tilde{U}^{(2m-2)}[(1-2s)\, \Delta t]\, \tilde{U}^{(2m-2)}(s\, \Delta t), \tag{26.44}$$

where s satisfies

$$2s^{2m-1} + (1-2s)^{2m-1} = 0. \tag{26.45}$$

Requiring s to be a real root gives

$$s = \frac{1}{2 - \sqrt[2m-1]{2}}. \tag{26.46}$$

Thus, a fourth-order method is[6]

$$\tilde{U}^{(4)}(\Delta t) = \tilde{U}^{(2)}(s_4\, \Delta t)\, \tilde{U}^{(2)}[(1-2s_4)\, \Delta t]\, \tilde{U}^{(2)}(s_4\, \Delta t), \tag{26.47}$$

[6]Michael Creutz and Andreas Gocksch, "Higher-order hybrid Monte Carlo algorithms," *Physical Review Letters* **63**, 9 (1989) (doi: 10.1103/PhysRevLett.63.9). Haruo Yoshida, "Construction of higher order symplectic integrators," *Physics Letters A* **150**, 262 (1990) (doi: 10.1016/0375-9601(90)90092-3); Etienne Forest and Ronald D. Ruth, "Fourth-Order Symplectic Integration," *Physica D* **43**, 105 (1990); J. Candy and W. Rozmus, "A symplectic integration algorithm for separable Hamiltonian functions," *Journal of Computational Physics* **92**, 230 (1991) (doi: 10.1016/0021-9991(91)90299-Z); André Bandrauk and Hai Shen, "Improved exponential split operator method for solving the time-dependent Schrödinger equation," *Chemical Physics Letters* **176**, 428 (1991) (doi: 10.1016/0009-2614(91)90232-X).

where s_4 and $1-2s_4$ are given by setting $m=2$ in (26.46)

$$\begin{aligned} s_4 &= \frac{1}{2-\sqrt[3]{2}} \approx 1.3512\ 0719\ 1959\ 6576\ 3405 \\ 1-2s_4 &= \frac{-\sqrt[3]{2}}{2-\sqrt[3]{2}} \approx -1.7024\ 1438\ 3919\ 3152\ 6810. \end{aligned} \tag{26.48}$$

Similarly, iterating this procedure produces the sixth-order method

$$\tilde{U}^{(6)}(\Delta t) = \tilde{U}^{(4)}(s_6\,\Delta t)\,\tilde{U}^{(4)}[(1-2s_6)\,\Delta t]\,\tilde{U}^{(4)}(s_6\,\Delta t), \tag{26.49}$$

where s_6 and $1-2s_6$ are given by setting $m=3$ in (26.46)

$$\begin{aligned} s_6 &= \frac{1}{2-\sqrt[5]{2}} \approx 1.1746\ 7175\ 8089\ 3633\ 8450 \\ 1-2s_6 &= \frac{-\sqrt[5]{2}}{2-\sqrt[5]{2}} \approx -1.3493\ 4351\ 6178\ 7267\ 6899. \end{aligned} \tag{26.50}$$

Of course, it is straightforward to compute the coefficients for the higher-order methods.

Let's compare these methods to the Richardson-extrapolation methods. The fourth-order Richardson method requires three applications of $\tilde{U}^{(2)}(\Delta t)$ for various step sizes for either the Richardson or unitary method. For the particular decomposition (26.4) of the Hamiltonian, note that some of the potential operators can be combined in both cases, but the number of Fourier transforms is the same. The sixth-order Richardson method, on the other hand, requires six applications of $\tilde{U}^{(2)}(\Delta t)$ for the Richardson method, compared to nine applications for the unitary method. Thus, at higher orders, unitarity comes at the expense of extra computational effort. Unitarity may well be a desireable property of an integration method, since it preserves a known invariant. However, one may take the *opposite* view, and advocate a method that does *not* preserve unitarity, since the computed norm of the wave function can act as a diagnostic for the overall accuracy of the solution. On the other hand, at high orders, the methods may become quite unstable for large step sizes, since the assumption that the error terms are small breaks down.

26.4.1.1 High-Order Methods with Reduced Substep Intervals

One potential concern of the methods based on the unitary three-term recursion (26.44) is that the step-size factors s_{2m} and $1-2s_{2m}$, by examination of Eq. (26.46), can be seen to always be "oversteps,"

$$\begin{aligned} s_{2m} &> 1 \\ 1-2s_{2m} &< -1, \end{aligned} \tag{26.51}$$

so that the middle "backwards" step is necessary to compensate for the forward steps. The cancellation required by these large steps may cause problems with stability and accuracy of the solutions. It is possible to find other hierarchical constructions that have smaller steps, but at the expense of additional computational effort to achieve the same order. For example, the five-factor symmetric decomposition[7]

$$\tilde{U}^{(2m)}(\Delta t) = \tilde{U}^{(2m-2)}(s\,\Delta t)\,\tilde{U}^{(2m-2)}(s\,\Delta t)\,\tilde{U}^{(2m-2)}[(1-4s)\,\Delta t]\,\tilde{U}^{(2m-2)}(s\,\Delta t)\,\tilde{U}^{(2m-2)}(s\,\Delta t), \tag{26.52}$$

is of the same form as Eq. (26.40), and thus s satisfies

$$4s^{2m-1} + (1-4s)^{2m-1} = 0, \tag{26.53}$$

or explicitly,

$$s = \frac{1}{4-\sqrt[2m-1]{4}}. \tag{26.54}$$

[7] G. Dattoli, *op. cit.*

so that the step factors s and $1-4s$ satisfy the bounds

$$\begin{aligned} \frac{1}{3} < s_{2m} < \frac{1}{2} \\ |1-4s_{2m}| < \frac{2}{3}, \end{aligned} \tag{26.55}$$

Specifically, the fourth-order values are

$$\begin{aligned} s_4 &= \frac{1}{4-\sqrt[3]{4}} \approx 0.4144\ 9077\ 1794\ 3757\ 3714 \\ 1-4s_4 &= \frac{-\sqrt[3]{4}}{4-\sqrt[3]{4}} \approx -0.6579\ 6308\ 7177\ 5029\ 4856, \end{aligned} \tag{26.56}$$

and the sixth-order values are

$$\begin{aligned} s_6 &= \frac{1}{4-\sqrt[5]{4}} \approx 0.3730\ 6582\ 7733\ 2728\ 2478 \\ 1-4s_6 &= \frac{-\sqrt[5]{4}}{4-\sqrt[5]{4}} \approx -0.4922\ 6331\ 0933\ 0912\ 9910. \end{aligned} \tag{26.57}$$

Again, at fourth order, the method here requires five applications of $\tilde{U}^{(2)}(\Delta t)$, compared to three for either the Richardson or the three-factor unitary method. At sixth-order, the method here requires 25 unitary factors, compared to nine for the three-factor construction and six for the Richardson method. Going to higher orders by this route is computationally rather expensive, so the benefits must be carefully weighed against the extra cost in terms of time.[8]

26.4.1.2 High-Order Minimal-Product Methods

In defining the sixth-order unitary methods above, we saw that their computational effort went far beyond that of the Richardson method. It is also possible to skip any hierarchical construction and search for high-order methods directly in terms of products of $\tilde{U}^{(2)}(\Delta t)$, and thereby generate methods that use products of fewer operators. This is possible because there are fewer constraints on the relative time steps within each factor. This is done by directly computing the error terms in the BCH expansion. For example, seven-factor, sixth-order symmetric splittings of the form

$$\tilde{U}^{(6)}(\Delta t) = \tilde{U}^{(2)}(w_3\,\Delta t)\,\tilde{U}^{(2)}(w_2\,\Delta t)\,\tilde{U}^{(2)}(w_1\,\Delta t)\,\tilde{U}^{(2)}(w_0\,\Delta t)\,\tilde{U}^{(2)}(w_1\,\Delta t)\,\tilde{U}^{(2)}(w_2\,\Delta t)\,\tilde{U}^{(2)}(w_3\,\Delta t), \tag{26.58}$$

can be constructed, but in general the equations to be solved to determine the w_j factors are complicated, and the solutions must be found numerically. One such method has[9]

$$\begin{aligned} w_1 &\approx -1.1776\ 7998\ 4178\ 87 \\ w_2 &\approx 0.2355\ 7321\ 3359\ 357 \\ w_3 &\approx 0.7845\ 1361\ 0477\ 560, \end{aligned} \tag{26.59}$$

with $w_0 = 1-2(w_1+w_2+w_3)$. Again, this has 7 unitary factors, as compared to the 6 required by Richardson extrapolation of the same order; evidently, the preservation of unitarity still demands some additional work.

Additionally, 15-factor, eighth-order symmetric splitting of the form

$$\tilde{U}^{(8)}(\Delta t) = \tilde{U}^{(2)}(w_7\,\Delta t)\,\ldots\tilde{U}^{(2)}(w_1\,\Delta t)\,\tilde{U}^{(2)}(w_0\,\Delta t)\,\tilde{U}^{(2)}(w_1\,\Delta t)\,\ldots\tilde{U}^{(2)}(w_7\,\Delta t), \tag{26.60}$$

[8] There may also be advantages to further requiring that the steps all be positive. See Siu A. Chin and C. R. Chen, "Fourth order gradient symplectic integrator methods for solving the time-dependent Schödinger equation," *Journal of Chemical Physics* **114**, 7338 (2001) (doi: 10.1063/1.1362288).

[9] Haruo Yoshida, *op. cit.*; Yoshida gives two other sixth-order methods in this reference.

can be constructed, but in general the equations to be solved to determine the w_j factors are complicated, and the solutions must be found numerically. An example is[10]

$$\begin{aligned} w_1 &\approx 0.3117\ 9081\ 2418\ 427 \\ w_2 &\approx -1.5594\ 6803\ 8214\ 47 \\ w_3 &\approx -1.6789\ 6928\ 2596\ 40 \\ w_4 &\approx 1.6633\ 5809\ 9633\ 15 \\ w_5 &\approx -1.0645\ 8714\ 7891\ 83 \\ w_6 &\approx 1.3693\ 4946\ 4168\ 71 \\ w_7 &\approx 0.6290\ 3065\ 0210\ 433, \end{aligned} \tag{26.61}$$

with $w_0 = 1 - 2(w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7)$. Again, this splitting has 15 unitary factors, as compared to the 10 required by Richardson extrapolation of the same order.

26.4.1.3 High-Order Treatment of Time-Dependent Potentials

The same strategies above for unitary, high-order product methods apply to cases of time-dependent potentials, so long as we again make the generalization

$$V(x)\Delta t \longrightarrow \int_t^{t+\Delta t} dt'\, V(x,t'), \tag{26.62}$$

in each factor

$$\tilde{U}^{(2)}(\Delta t) \longrightarrow \tilde{U}^{(2)}(t+\Delta t, t) \tag{26.63}$$

that composes the higher order methods, so long as we define the time interval for each factor appropriately.[11] In fact, we can use the splitting (26.21) as our fundamental second-order factor to handle the time-dependent potentials, since it has the right order of accuracy and possesses time-reversal symmetry:

$$\tilde{U}^{(2)}(t+\Delta t, t) = e^{-iV(x,t+\Delta t)\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV(x,t)\Delta t/2\hbar} = U(t+\Delta t, t) + O(\Delta t^3). \tag{26.64}$$

Thus, for example, the fourth-order symplectic method (26.47) becomes

$$\tilde{U}^{(4)}(t+\Delta t, t) = \tilde{U}^{(2)}(t + s_4\,\Delta t, t)\, \tilde{U}^{(2)}[t + (1-s_4)\,\Delta t, t + s_4\,\Delta t]\, \tilde{U}^{(2)}(t+\Delta t, t + (1-s_4)\,\Delta t), \tag{26.65}$$

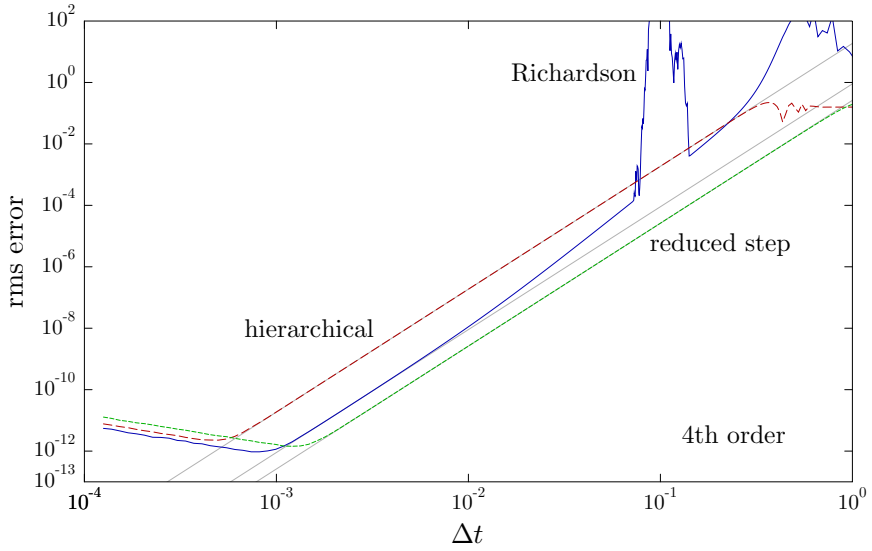
so that the fourth-order operator is the composition of the three second-order operators, and the time intervals covered by each second-order operator is the same as before. The other methods above generalize in the same way.

26.4.1.4 Numerical Test: Fourth Order

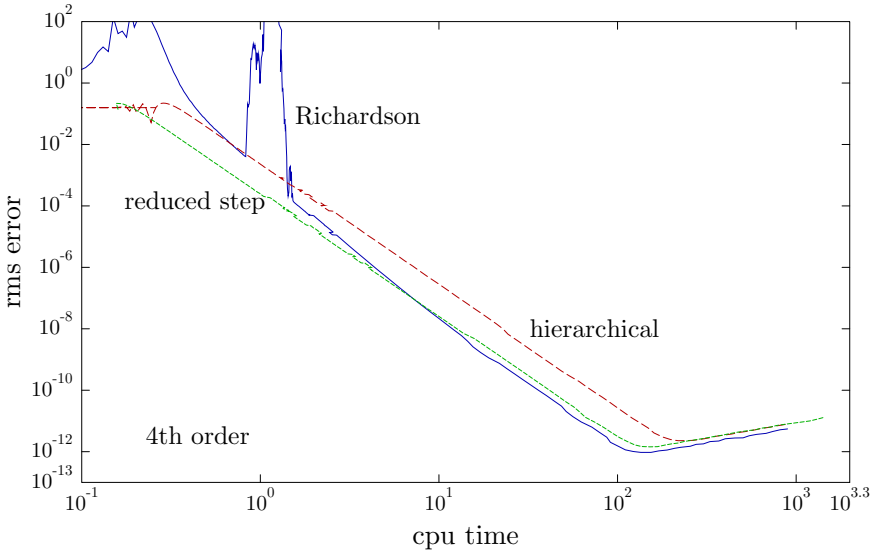
Using the same test problem as for the Richardson-extrapolation test, we can compare different fourth-order methods above: the Richardson-extrapolation method, the unitary hierarchical method of Section 26.4.1, and the unitary reduced-step method of Section 26.4.1.1.

[10] Haruo Yoshida, *op. cit.*; Yoshida gives four other eighth-order methods in this reference.

[11] André Bandrauk and Hai Shen, "Exponential split operator methods for solving coupled time-dependent Schrödinger equations," *Journal of Chemical Physics* **99**, 1185 (1993) (doi: 10.1063/1.465362).



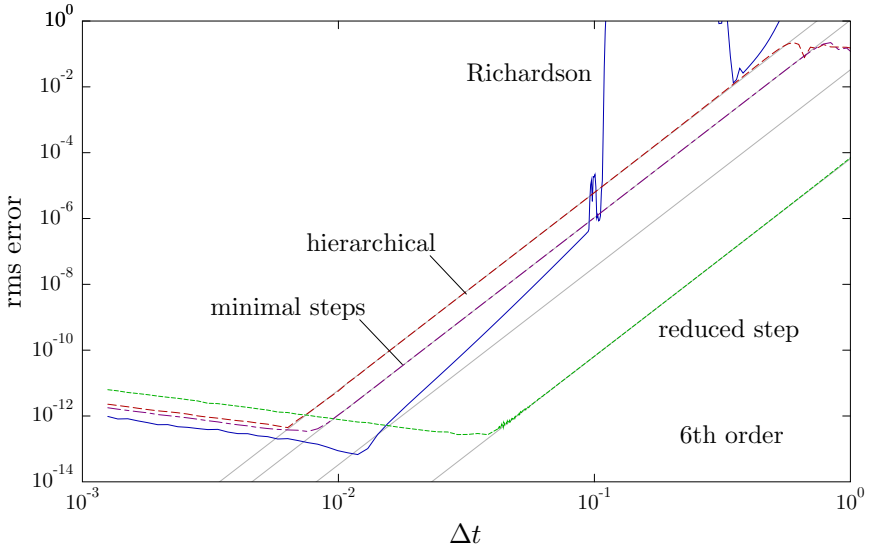
As a function of step size, the Richardson and reduced-step methods are clearly superior to the hierarchical method in this test, with the reduced-step unitary method having the highest accuracy. Of course, both unitary methods do not suffer from the instability for large step sizes, but this is in the regime of poor accuracy anyway. The Richardson method is the best of the methods in terms of roundoff error. However, there is a wide range of complexity among the different methods, so it is useful to consider the accuracy as a function of cpu time.



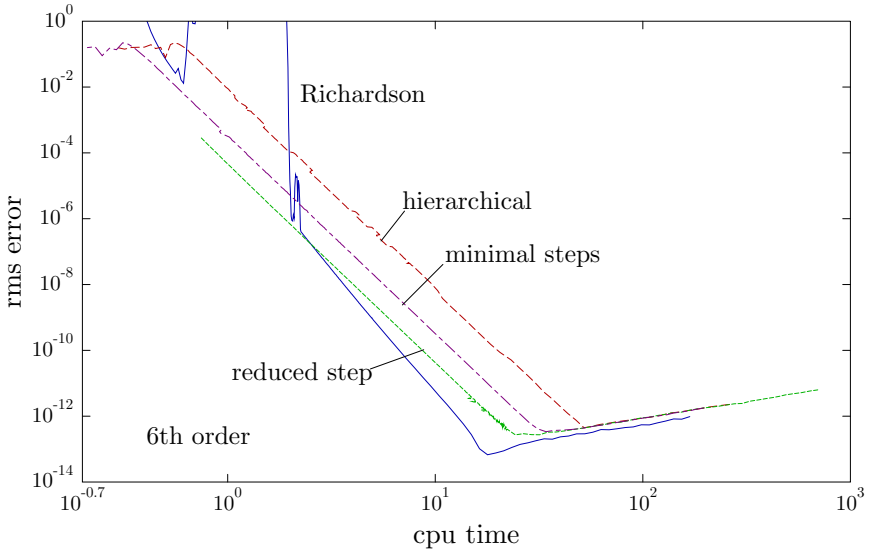
Perhaps counterintuitively, despite the extra complexity, the reduced-step method does quite well, with the Richardson method narrowly being the most efficient in terms of cpu time for high accuracy.

26.4.1.5 Numerical Test: Sixth Order

Similarly, we can compare the various sixth-order algorithms: Richardson-extrapolation, the unitary hierarchical method of Section 26.4.1, the unitary reduced-step-size method of Section 26.4.1.1, and the unitary minimal-product method of Section 26.4.1.2.



The Richardson method is notable here as, oddly, the only one that does not scale at sixth order, but also achieves the best overall accuracy at the optimal step size. The reduced-step method clearly achieves the largest step size for a given accuracy of all the methods. However, recall that with 25 second-order substeps, it is also by far the most computationally expensive on each step, so it is important to consider cpu time for any target accuracy.



However, we see that the reduced-step method compares quite favorably with the other unitary methods. Despite its complexity, it seems to be a reasonable choice. However, for high accuracies it still seems to be edged out by the Richardson method. It could be that in more complicated problems, the stability properties of the reduced-step unitary method are beneficial enough to make it the method of choice, although Richardson extrapolation certainly seems to be a generally useful and simple method.

26.4.2 Nonlinear Schrödinger Equations

One can also consider *nonlinear* Schrödinger equations with potentials of the form

$$V(x, \psi, t) = V(x, t) + g|\psi(x, t)|^2, \tag{26.66}$$

which lead to cubically nonlinear terms in the Schrödinger equation. The **Gross–Pitaevskii** equation is precisely of this form, and models the evolution of a Bose–Einstein condensate: the nonlinear term models

energy shifts due to two-body collisions, so that there is an extra spatial potential proportional to the condensate number density. Clearly, we can adapt the above methods if we take the order-two operator to be

$$\tilde{U}^{(2)}(t+\Delta t,t) = e^{-iV[x,\psi(x,t+\Delta t),t+\Delta t]\Delta t/2\hbar} e^{-iT(p)\Delta t/\hbar} e^{-iV[x,\psi(x,t),t]\Delta t/2\hbar} = U(t+\Delta t,t) + O(\Delta t^3). \quad (26.67)$$

The problem here is that we need to know $\psi(x,t+\Delta t)$ to construct the leftmost factor, but that's what we're trying to calculate. Thus, we have an implicit method. The simplest way to handle this is to use functional iteration (or perturbation theory, if you want to think of it that way): simply compute $\psi(x,t+\Delta t)$, but using as a guess the partially evolved wave function

$$\psi_{\text{guess}}(x,t) = e^{-iT(p)\Delta t/\hbar} e^{-iV[x,\psi(x,t),t]\Delta t/2\hbar} \psi(x,t). \quad (26.68)$$

Then use the result as a guess to recompute $\psi(x,t+\Delta t)$, and so on until the procedure converges. This strategy is certainly "safe" in adapting the above procedures, since the nonlinear potential amounts to an ordinary time-dependent potential, so long as we can calculate the potential at any time.

However, things turn out to be a bit easier than this, at least at low orders.[12] Observe that

$$|\psi(x,t+\Delta t)|^2 = |U(t+\Delta t,t)\,\psi(x,t)|^2 = |\psi(x,t)|^2 + O(\Delta t^2). \quad (26.69)$$

Since in the methods above, $|\psi(x,t)|^2$ is always multiplied by Δt^2, so with an error of $O(\Delta t^3)$, we can evaluate $\psi(x,t)$ in the nonlinear potential at any other time up to Δt away. What this means, is that even with a single iteration—that is, using $\psi_{\text{guess}}(x,t)$ in place of $\psi(x,t+\Delta t)$ in (26.67)—the order of the method is not changed. (Recall that this was not true of the ordinary potential $V(x,t)$, which had to be evaluated at the proper points.) Thus, we have an explicit, second-order method for evolving the GPE.

However, the explicit method here seems to lack time-reversal symmetry, and so it is not obvious how useful it is to compose higher-order methods. An intermediate strategy between the explicit and implicit methods is to iterate the implicit scheme once, for a Runge–Kutta style approximation to a symmetric splitting. (Note that iterating an implicit method a fixed number of times is in fact an *explicit* scheme, but may approximate the implicit scheme well.) However, it has been shown[13] that the *simple* explicit scheme above works to construct a fourth-order method. It is not clear if the simple explicit strategy works at higher orders, and if higher-order methods are to be used, it is worth checking the sensitivity of the solution to the number of iterations.

26.4.3 Symplectic Integration

The high-order unitary methods above are also important in *classical* mechanics, when simulating trajectories of Hamiltonian systems. Rather than conserving the norm of a wave function, the methods applied in this context conserve *phase-space volumes*, and in the case of autonomous Hamiltonian systems, they do a better job of conserving total energy.[14] For example, such methods are important in celestial mechanics, where it is desireable to obtain highly accurate, long-term solutions to Newton's equations without spurious damping effects.

Recall from Section 4.4.3.2 that Hamilton's equations can be written

$$\partial_t z_\alpha = \{z_\alpha, H\}_{\text{P}} = \Omega_{\alpha\beta} \frac{\partial H}{\partial z_\beta}, \quad (26.70)$$

where $\{f,g\}_{\text{P}}$ is the Poisson bracket, the $2N$ canonically conjugate coordinates are combined as

$$z_\alpha := (x_1,\ldots,x_N,p_1,\ldots,p_N), \quad (26.71)$$

[12] A. D. Bandrauk and Hai Shen, "High-order split-step exponential methods for solving coupled nonlinear Schrödinger equations," *Journal of Physics A: Mathematical and General* **27**, 7147 (1994) (doi: 10.1088/0305-4470/27/21/030); Juha Javanainen and Janne Ruostekoski, "Symbolic calculation in development of algorithms: split-step methods for the Gross-Pitaevskii equation," *Journal of Physics A: Mathematical and General* **39**, L179 (2006) (doi: 10.1088/0305-4470/39/12/L02).

[13] Juha Javanainen and Janne Ruostekoski, *op. cit.*

[14] For further reading and the source of some of the notation here, see Denis Donnelly and Edwin Rogers, "Symplectic integrators: An introduction," *American Journal of Physics* **73**, 938 (2005) (doi: 10.1119/1.2034523).

and $\Omega_{\alpha\beta}$ is the canonical cosymplectic form,

$$(\Omega_{\alpha\beta}) := \begin{pmatrix} 0_n & \mathcal{I}_n \\ -\mathcal{I}_n & 0_n \end{pmatrix}, \tag{26.72}$$

with $\mathcal{I}_n$ denoting the $n \times n$ identity matrix and 0_n the $n \times n$ null matrix.

Suppose we write the Poisson bracket as a differential operator,

$$\partial_t z_\alpha = D_H z_\alpha, \tag{26.73}$$

where for any phase-space function $f(z)$,

$$D_H f(z) = \{f, H\}_{\text{P}} = \frac{\partial f}{\partial z_\alpha} \Omega_{\alpha\beta} \frac{\partial H}{\partial z_\beta}. \tag{26.74}$$

Noting that D_H acts individually on each component of z_α, we can simply write the vector relation

$$\partial_t z = D_H z, \tag{26.75}$$

The formal solution of Eq. (26.75) is in terms of the exponentiated operator:

$$z(t) = e^{D_H t} z(0). \tag{26.76}$$

Now we will specialize to one degree of freedom, where the generalization to higher dimensions is simple. If the Hamiltonian splits into the form

$$H(x, p) = T(p) + V(x) \tag{26.77}$$

as before, then the Poisson-bracket operator takes the explicit form

$$D_H = T'(p)\partial_x - V'(x)\partial_p =: D_T + D_V \tag{26.78}$$

of the sum of two noncommuting operators, since $[D_V, D_T] = T'(p)V''(x) - V'(x)T''(p)$ does not vanish in general. We can thus use exactly the above theory to split the exponential operator $\exp(D_H\,\Delta t)$ for evolution over a short time Δt into simpler parts of the form $\exp(D_T\,\Delta t)$ and $\exp(D_V\,\Delta t)$, where the effects of each component operator can be written explicitly. Thus, the partial kinetic solution is

$$e^{D_T \Delta t} z(t) = e^{\Delta t\, T'(p)\partial_x} \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} x(t) + T'[p(t)]\,\Delta t \\ p(t) \end{bmatrix}, \tag{26.79}$$

while the partial potential solution is

$$e^{D_V \Delta t} z(t) = e^{-\Delta t\, V'(x)\partial_p} \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} x(t) \\ p(t) - V'[x(t)]\,\Delta t \end{bmatrix}. \tag{26.80}$$

Since we can write out the explicit effect of each operator on $z(t)$, there is no need to implement Fourier transforms between operators, as in the quantum-mechanical case. However, the splittings above still provide numerical methods of the same orders as before.

26.4.3.1 Euler–Cromer Method

The simplest splitting we can use is the first-order splitting (26.6), which here bcomes

$$e^{D_H \Delta t} = e^{D_T \Delta t} e^{D_V \Delta t} + O(\Delta t^2). \tag{26.81}$$

Writing this out explicitly,

$$e^{D_H \Delta t} z(t) = e^{D_T \Delta t} e^{D_V \Delta t} \begin{bmatrix} x(t) \\ p(t) \end{bmatrix} + O(\Delta t^2) = \begin{bmatrix} x(t) + T'[p(t + \Delta t)]\,\Delta t \\ p(t) - V'[x(t)]\,\Delta t \end{bmatrix} + O(\Delta t^2) \tag{26.82}$$

Written out as a mapping, this method becomes

$$\begin{aligned} x_{j+1} &= x_j + T'(p_{j+1})\,\Delta t \\ p_{j+1} &= p_j - V'(x_j)\,\Delta t, \end{aligned} \tag{26.83}$$

where $x_j := x(j\Delta t)$ and $p_j := p(j\Delta t)$. This first-order method is the **Euler–Cromer method**. We can see that this method is first-order accurate and half-implicit, but it easy to iterate since you just compute the new p_{j+1} before you compute the new x_{j+1}. This mapping is a canonical transformation, since the transformation $z_j \longrightarrow z_{j+1}$ preserves the cosymplectic form:

$$\Omega_{\alpha\beta} = \frac{\partial(z_{j+1})_\alpha}{\partial(z_j)_\mu}\Omega_{\mu\nu}\frac{\partial(z_{j+1})_\beta}{\partial(z_j)_\nu}. \tag{26.84}$$

This follows from writing out the Jacobian matrix

$$\frac{\partial(z_{j+1})}{\partial(z_j)} = \begin{bmatrix} 1 - T''(p_{j+1})V''(x_j)\,\Delta t^2 & T''(p_{j+1})\,\Delta t \\ -V''(x_j)\,\Delta t & 1 \end{bmatrix} \tag{26.85}$$

and multiplying out the matrix product to see that $\Omega_{\alpha\beta}$ is unchanged:

$$\begin{bmatrix} 1 - T''(p_{j+1})V''(x_j)\,\Delta t^2 & T''(p_{j+1})\,\Delta t \\ -V''(x_j)\,\Delta t & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 - T''(p_{j+1})V''(x_j)\,\Delta t^2 & -V''(x_j)\,\Delta t \\ T''(p_{j+1})\,\Delta t & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{26.86}$$

Also, directly from the fact that the Jacobian determinant is unity,

$$\det\frac{\partial(z_{j+1})}{\partial(z_j)} = \det\begin{bmatrix} 1 - T''(p_{j+1})V''(x_j)\,\Delta t^2 & T''(p_{j+1})\,\Delta t \\ -V''(x_j)\,\Delta t & 1 \end{bmatrix} = 1, \tag{26.87}$$

we see explicitly the preservation of phase-space areas. Since these maps correspond to canonical transformations, they are **symplectic map** and thus this approximation scheme is a **symplectic integrator**. Compared to the Euler method, for approximately harmonic oscillation, the Euler–Cromer method produces bounded energy errors, while the Euler method produces asymptotically growing errors.

We can obviously interchange the order of the operators in (26.82) to obtain another first-order symplectic algorithm. The mapping for this method analogous to Eqs. (26.83) is

$$\begin{aligned} x_{j+1} &= x_j + T'(p_j)\,\Delta t \\ p_{j+1} &= p_j - V'(x_{j+1})\,\Delta t, \end{aligned} \tag{26.88}$$

and it follows in the same way as the Euler–Cromer method that this mapping is symplectic.

26.4.3.2 Verlet Method

It is thus straightforward to construct a second-order symplectic method by using the symmetric splitting (26.7), which adapted to symplectic integration becomes

$$e^{D_H\Delta t} = e^{D_V\Delta t/2}e^{D_T\Delta t}e^{D_V\Delta t/2} + O(\Delta t^3). \tag{26.89}$$

The mapping for this splitting can be rewritten as

$$e^{D_H\Delta t} = \left[e^{D_V\Delta t/2}e^{D_T\Delta t/2}\right]\left[e^{D_T\Delta t/2}e^{D_V\Delta t/2}\right] + O(\Delta t^3), \tag{26.90}$$

and thus corresponds to a composition of an Euler–Cromer step as in Eqs. (26.83) with a ''reverse'' Euler–Cromer step as in Eqs. (26.88), both of step size $\Delta t/2$. The composition of symplectic maps is still symplectic, and thus we have a second-order symplectic method.

To write out the mapping explicitly, first we write out the composition of the two maps (26.83) and (26.88), each for a half step:

$$\begin{aligned} x_{j+1/2} &= x_j + T'(p_{j+1/2})\,(\Delta t/2) \\ p_{j+1/2} &= p_j - V'(x_j)\,(\Delta t/2) \\ x_{j+1} &= x_{j+1/2} + T'(p_{j+1/2})\,(\Delta t/2) \\ p_{j+1} &= p_{j+1/2} - V'(x_{j+1})\,(\Delta t/2). \end{aligned} \tag{26.91}$$

Eliminating the intermediate step, we find

$$\begin{aligned} x_{j+1} &= x_j + T'[p_j - V'(x_j)\,\Delta t/2]\,\Delta t \\ p_{j+1} &= p_j - \frac{V'(x_j) + V'(x_{j+1})}{2}\,\Delta t. \end{aligned} \tag{26.92}$$

For the important ''particle'' case of $T(p) = p^2/2$, the mapping becomes

$$\begin{aligned} x_{j+1} &= x_j + p_j\,\Delta t - V'(x_j)\,\Delta t \\ p_{j+1} &= p_j - \frac{V'(x_j) + V'(x_{j+1})}{2}\,\Delta t. \end{aligned} \tag{26.93}$$

This is the velocity form of the **Verlet method**, or the **velocity Verlet method**. The original form for the Verlet method[15] follows from using the other possible form of the symmetric splitting, where the ''drift'' operator is applied first and last. This amounts to the replacements $x \longrightarrow p$, $p \longrightarrow x$, $T \longrightarrow -V$, and $V \longrightarrow -T$ in Eqs. (26.92):

$$\begin{aligned} x_{j+1} &= x_j + \frac{T'(p_j) + T'(p_{j+1})}{2}\,\Delta t \\ p_{j+1} &= p_j - V'[x_j + T'(p_j)\,\Delta t/2]\,\Delta t. \end{aligned} \tag{26.94}$$

Again, if $T(p) = p^2/2$, the mapping becomes

$$\begin{aligned} x_{j+1} &= x_j + \frac{p_j + p_{j+1}}{2}\,\Delta t \\ p_{j+1} &= p_j - V'[x_j + p_j\,\Delta t/2]\,\Delta t. \end{aligned} \tag{26.95}$$

Notice that in both cases, each variable is evolved according to a method symmetric about the middle time $\Delta t/2$, either by computing the average function value $(f_j + f_{j+1})/2$, which approximates the middle value $f_{j+1/2}$, or by evaluating a function after taking half of an Euler-type step. The second-order nature of these approximations follow from the two second-order integral approximations

$$\begin{aligned} \int_t^{t+\Delta t} f(t')\,dt' &= f(t + \Delta t/2)\Delta t + O(\Delta t^3) \\ &= \frac{f(t) + f(t+\Delta t)}{2}\Delta t + O(\Delta t^3), \end{aligned} \tag{26.96}$$

which both appear in the Verlet methods.

26.4.3.3 Higher Order Methods

The nice thing about symplectic methods here is that we have already worked out the theory to extend the second-order methods to higher order. Because of the time-reversal symmetry noted above for the Verlet method, the *local* truncation error only involves *odd* powers of Δt (or the *global* error involves only *even*

[15] Loup Verlet, ''Computer 'Experiments' on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules,'' *Physical Review* **159**, 98 (1967) (doi: 10.1103/PhysRev.159.98).

powers of Δt). All the techniques above for the split-operator Fourier-transform method thus work to extend the Verlet method, if we replace each application of $\tilde{U}^{(2)}(s\Delta t)$ by a Verlet step (in either form, as long as we stick to the same form) of $s\Delta t$. It is not useful to write out the explicit algorithms here, since in general one would simply code a subroutine to take a Verlet step and then call it multiple times on each time step of Δt—say, three times for the fourth-order method with step sizes $s_4\Delta t$, $(1-2s_4)\Delta t$, and $s_4\Delta t$ corresponding to the three-step method of Eqs. (26.47).

26.4.3.4 Time-Dependent Potentials

In principle, the above assumption of a time-independent Hamiltonian is not restrictive, since any Hamiltonian system of N degrees of freedom with explicit time dependence is formally equivalent to a Hamiltonian system of $N+1$ degrees of freedom with a time-*independent* Hamiltonian. (Thus, an explicitly time-dependent system of N degrees of freedom is often said to have "$N\frac{1}{2}$ degrees of freedom.") However, it is useful to generalize the above methods for time-dependent Hamiltonians anyway as a more direct method for handling explicit time dependence. To handle explicitly the case of time-dependent potentials, we make the same adjustments as for the unitary methods to evaluate the potential at the proper time. Thus, the Euler–Cromer method (26.83) is not sensitive to the precise time of evaluation, and for example becomes

$$\begin{aligned} x_{j+1} &= x_j + T'(p_{j+1})\,\Delta t \\ p_{j+1} &= p_j - V'(x_j, t_j)\,\Delta t. \end{aligned} \tag{26.97}$$

For the Verlet methods, we can take in analogy with Eq. (26.21) the splitting

$$e^{D_H \Delta t} = \left[e^{D_V(t+\Delta t)\Delta t/2} e^{D_T \Delta t/2}\right]\left[e^{D_T \Delta t/2} e^{D_V(t)\Delta t/2}\right] + O(\Delta t^3), \tag{26.98}$$

where $D_V(t)$ has the potential evaluated at time t, so that the velocity Verlet method (26.92) becomes

$$\begin{aligned} x_{j+1} &= x_j + T'[p_j - V'(x_j, t_j)\,\Delta t/2]\,\Delta t \\ p_{j+1} &= p_j - \frac{V'(x_j, t_j) + V'(x_{j+1}, t_{j+1})}{2}\,\Delta t, \end{aligned} \tag{26.99}$$

and the usual Verlet method becomes

$$\begin{aligned} x_{j+1} &= x_j + \frac{T'(p_j) + T'(p_{j+1})}{2}\,\Delta t \\ p_{j+1} &= p_j - V'[x_j + T'(p_j)\,\Delta t/2, t_{j+1/2}]\,\Delta t, \end{aligned} \tag{26.100}$$

preserving the second-order accuracy of the methods in the explicitly time-dependent case.

26.5 Exercises

Problem 26.1

Verify the factorizations of Eqs. (26.6) and (26.7) using the BCH formula.

Problem 26.2

Justify the following statement: despite the different *local* errors, the error in the evolution of the quantum state according to the two splittings, Eqs. (26.6) and (26.7), after *many* steps is approximately the *same*.

Problem 26.3

Why do only the even-order error terms appear in the Richardson expansion (26.26)?

Problem 26.4

Verify that the Richardson-extrapolation formulae (26.27), (26.28), (26.29), and (26.30) are correct and have the advertised accuracy.

Chapter 27

Stochastic Differential Equations

Now we will consider the numerical solution to *stochastic* differential equations (SDEs) of the form

$$dy = \alpha(y,t)\,dt + \beta(y,t)\,dW, \tag{27.1}$$

where $dW(t)$ is the Wiener process as usual (see Chapter 17). We will stick to considering only Itō SDEs for simplicity. Because of the unusual and singular nature of $dW(t)$, the methods that apply to ODEs (Chapter 24) do not work well here, and we will have to develop new methods. We will also see that accurately solving SDEs is much more difficult than ODEs. However, we will then have the formalism to put stochastic Schrödinger and master equations (Chapters 18 and 19) on the computer.[1]

27.1 Stochastic Euler Method

The simplest numerical method we considered in solving ODEs was the Euler method (24.2). This applies to Eq. (27.1) when $\beta = 0$, in which case the update equation is

$$y_{n+1} = y_n + \alpha(y_n, t_n)\,\Delta t + O(\Delta t^2), \tag{27.2}$$

where again the solution is evolved in finite time steps of Δt, and we are using the condensed notation $y_n := y(n\Delta t)$ and $t_n := n\Delta t$. We can try extending this method to the SDE (27.1) by taking the same linear approximation to the stochastic term, to arrive at the **stochastic Euler method** (often called the **Euler–Maruyama method**):[2]

$$y_{n+1} = y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\,\Delta W_n. \tag{27.3}$$
(stochastic Euler method)

We are defining the time increment as before, and the Wiener increment ΔW_n is defined in an analogous way:

$$\begin{aligned} \Delta t_n &:= \int_{t_n}^{t_{n+1}} dt' \\ \Delta W_n &:= \int_{t_n}^{t_{n+1}} dW(t'). \end{aligned} \tag{27.4}$$
(time increments)

Of course, $\Delta t_n \equiv \Delta t$ is independent of n, so we need not indicate explicit time dependence. The key, however, is that the approximation works when α and β vary slowly over the interval $[t_n, t_{n+1}]$, and thus

[1] One of the best and most complete references on this subject is Peter E. Kloeden and Eckhard Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd ed. (Springer, 2000).

[2] Gisirō Maruyama, "Continuous Markov Processes and Stochastic Equations," *Rendiconti del Circolo Matematico di Palermo* **4**, 48 (1955).

when we try to compute the solution

$$y_{n+1} = y_n + \int_{t_n}^{t_{n+1}} dt'\, \alpha(y(t'), t') + \int_{t_n}^{t_{n+1}} dW(t')\, \beta(y(t'), t'), \tag{27.5}$$

we can, to first approximation, treat α and β as constants and pull them out of their respective integrals, which precisely yields Eq. (27.3).

27.1.1 Truncation Error

Recall the Itō chain rule (17.193) for a function $f(y)$, where $y(t)$ satisfies the SDE (27.1):

$$df(y) = \left[f'(y)\, \alpha(y,t) + \frac{1}{2} f''(y)\, \beta^2(y,t) \right] dt + f'(y)\, \beta(y,t)\, dW. \tag{27.6}$$

Integrating this expression from $t' = t_0$ to t,

$$f(y(t)) = f(y(t_0)) + \int_{t_0}^{t} \left[f'(y)\, \alpha(y,t') + \frac{1}{2} f''(y)\, \beta^2(y,t') \right] dt' + \int_{t_0}^{t} f'(y)\, \beta(y,t')\, dW(t'). \tag{27.7}$$

Now we can let $f(y) \longrightarrow \alpha(y,t)$, $f(y) \longrightarrow \beta(y,t)$, and $t_0 \longrightarrow t_n$, and then put the two resulting expressions in Eq. (27.5) to obtain

$$\begin{aligned} y_{n+1} = y_n &+ \alpha(y_n, t_n)\, \Delta t + \beta(y_n, t_n)\, \Delta W_n \\ &+ \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt'' \left[\alpha'(y,t'')\, \alpha(y,t'') + \frac{1}{2} \alpha''(y,t'')\, \beta^2(y,t'') \right] + \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t'')\, \alpha'(y,t'')\, \beta(y,t'') \\ &+ \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt'' \left[\beta'(y,t'')\, \alpha(y,t'') + \frac{1}{2} \beta''(y,t'')\, \beta^2(y,t'') \right] \\ &+ \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'')\, \beta'(y,t'')\, \beta(y,t''). \end{aligned} \tag{27.8}$$

The primes here are equivalent to the partial derivatives ∂_y. Again, we can pretend that the α and β functions are constant over the short time interval Δt. Then we have simple integrals of the form

$$\int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt''; \qquad \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t''); \qquad \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt''; \qquad \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t''), \tag{27.9}$$

to deal with. The first integral is just $\Delta t^2/2$, or for our purposes, simply $O(\Delta t^2)$. Similarly, if we proceed with counting each ΔW on average as equivalent to $\Delta t^{1/2}$, the second and third integrals are $O(\Delta t^{3/2})$, and the last integral is $O(\Delta t^1)$. Clearly, any error in treating α and β as constants will result in higher-order errors, so the truncation error here is $O(\Delta t)$. Thus, we may again write the stochastic Euler method as

$$y_{n+1} = y_n + \alpha(y_n, t_n)\, \Delta t + \beta(y_n, t_n)\, \Delta W_n + O(\Delta t). \tag{27.10}$$
(stochastic Euler method)

The local truncation error here is, as we see, much worse than the deterministic Euler method (27.2). To estimate the *global* truncation error, again suppose we use this method to evolve the solution out to a fixed final time t in N steps of $\Delta t = t/N$. The local truncation error on each step is $O(\Delta t)$, due essentially to the stochastic term. The dominant errors will thus be random and uncorrelated, so we consider the error in a random walk of N steps of mean size of order t/N, which will scale as $\sqrt{N}(t/N) = t/\sqrt{N} = O(\Delta t^{1/2})$. Thus the global error of the stochastic Euler scheme converges very badly, as $O(\Delta t^{1/2})$. This is indicative of the fact that stochastic differential equations are much more difficult to solve than the deterministic counterparts. In any case, as in the ODE case, we call this method an $O(\Delta t^{1/2})$ method, because the local truncation is correct to this order, or equivalently, the global error is $O(\Delta t^{1/2})$.

27.2 Milstein Method

To construct the next, relatively simple, higher-order method, we note that the only term we neglected in Eq. (27.8) that was $O(\Delta t)$ was the last one, involving the double integral

$$\begin{aligned}
\int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'') &= \int_{t_n}^{t_{n+1}} dW(t') \left[W(t') - W(t_n) \right] \\
&= \frac{1}{2} \int_{t_n}^{t_{n+1}} \left\{ d[W(t')]^2 - dt \right\} - W(t_n) \int_{t_n}^{t_{n+1}} dW(t') \\
&= \frac{1}{2} \left[W^2(t_{n+1}) - W^2(t_n) \right] - \frac{\Delta t}{2} - W(t_n) \Delta W(t_n) \\
&= \frac{1}{2} \left[W(t_{n+1}) + W(t_n) \right] \Delta W(t_n) - \frac{\Delta t}{2} - W(t_n) \Delta W(t_n) \\
&= \frac{1}{2} \left[W(t_{n+1}) - W(t_n) \right] \Delta W(t_n) - \frac{\Delta t}{2} \\
&= \frac{1}{2} \left\{ [\Delta W(t_n)]^2 - \Delta t \right\},
\end{aligned} \tag{27.11}$$

where we used the Itō rule $d(W^2) = 2W\, dW + dW^2 = 2W\, dW + dt$ in the second step. The idea behind the **Milstein method**[3] is to keep this correction term, keeping in mind that the factor we just worked out is multiplied by $\beta(y, t_n)\, \beta'(y, t_n)$, which we are treating as constant to this order of approximation:

$$y_{n+1} = y_n + \alpha(y_n, t_n)\, \Delta t + \beta(y_n, t_n)\, \Delta W_n + \frac{1}{2} \beta(y_n, t_n)\, \beta'(y_n, t_n) \left(\Delta W_n^2 - \Delta t \right) + O(\Delta t^{3/2}). \tag{27.12}$$

(Milstein method)

Again, the prime refers to partial differentiation with respect to y_n. Clearly the local truncation error is $O(\Delta t^{3/2})$, since we have ignored the two integrals in Eqs. (27.9) of the same order. By the same argument as for the Euler method, the global error is a factor of $\Delta t^{1/2}$ worse, or $O(\Delta t)$, which is the same global error as the *deterministic* Euler method.

27.2.1 Multiplicative vs. Additive Noise

Note that the correction term in the Milstein method (27.12) is of the form $\beta\beta'$, and thus is only required for *multiplicative* noise (where β is a function of y). For *additive* noise (β independent of y), the correction term vanishes, and the stochastic Euler method has $O(\Delta t)$ global error. In general, we expect additive noise to be easier to solve numerically than multiplicative noise, and one strategy to improve the accuracy of numerical methods is to use the Lamperti transform (Section 17.7.4.2) to change a multiplicative-noise process to an additive process, assuming that the transform can be inverted to give the original solution.

27.3 Stochastic Taylor Expansion

What we have written down in Eq. (27.8) is something like a Tayler expansion for y_{n+1} in terms of y_n. Again, treating α and β as constants, essentially what we have is a Taylor expansion, neglecting $O(\Delta t^{3/2})$ terms. The full Taylor expansion is given by iterating the above procedure, for example using Eq. (27.7) to replace the functions α and β evaluated at time t''. This procedure obviously introduces triple stochastic integrals; the next iteration introduces quadruple integrals, and so forth.

[3] G. N. Mil'shtein, "Approximate Integration of Stochastic Differential Equations," *Theory of Probability and its Applications* **19**, 557 (1974).

27.3.1 Single and Double Integrals

The first thing that is clear is that we will need to employ a hierarchy of stochastic integrals. The ones we have considered so far are the single integrals

$$\begin{aligned}(I_0)_n &:= \int_{t_n}^{t_{n+1}} dt' = \Delta t \\ (I_1)_n &:= \int_{t_n}^{t_{n+1}} dW(t') = \Delta W_n,\end{aligned} \tag{27.13}$$

as well as the double integrals

$$\begin{aligned}(I_{00})_n &:= \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dt'' = \frac{\Delta t^2}{2} \\ (I_{10})_n &:= \int_{t_n}^{t_{n+1}} dt' \int_{t_n}^{t'} dW(t'') \\ (I_{01})_n &:= \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dt'' = \Delta t\, \Delta W_n - (I_{10})_n = (I_0)_n (I_1)_n - (I_{10})_n \\ (I_{11})_n &:= \int_{t_n}^{t_{n+1}} dW(t') \int_{t_n}^{t'} dW(t'') = \frac{\Delta W_n^2}{2} - \frac{\Delta t}{2}.\end{aligned} \tag{27.14}$$

The first and the last integral we worked out before, and the third integral follows from

$$\begin{aligned}(I_{01})_n &= \int_{t_n}^{t_{n+1}} dW(t')\,(t' - t_n) \\ &= \int_{t_n}^{t_{n+1}} dW(t')\,t' - t_n \Delta W_n \\ &= [t' W(t')]_{t_n}^{t_{n+1}} - \int_{t_n}^{t_{n+1}} dt'\, W(t') - t_n \Delta W_n \\ &= t_{n+1} W(t_{n+1}) - t_n W(t_n) - \int_{t_n}^{t_{n+1}} dt'\,[W(t') - W(t_n)] - \Delta t\, W(t_n) - t_n \Delta W_n \\ &= t_{n+1} W(t_{n+1}) - t_n W(t_{n+1}) - (I_{10})_n - \Delta t\, W(t_n) \\ &= \Delta t\, W(t_{n+1}) - (I_{10})_n - \Delta t\, W(t_n) \\ &= \Delta t\, \Delta W_n - (I_{10})_n,\end{aligned} \tag{27.15}$$

where we integrated by parts in the third step. Note that all the double integrals can be reduced to expressions in terms of single integrals and $(I_{10})_n$; however, this last double integral is irreducible in the sense that it cannot be written only in terms of single integrals. We can thus characterize it more completely. It is clearly Gaussian, as from its definition it is a sum over independent, Gaussian random variables. It has mean, variance, and covariance with ΔW_n given by

$$\begin{aligned}\langle\langle (I_{10})_n \rangle\rangle &= 0 \\ \left\langle\left\langle (I_{10})_n^2 \right\rangle\right\rangle &= \frac{\Delta t^3}{3} \\ \langle\langle (I_{10})_n\, \Delta W_n \rangle\rangle &= \frac{\Delta t^2}{2}.\end{aligned} \tag{27.16}$$

The mean is obvious, as $(I_{10})_n$ is again the sum over independent, Gaussian random variables of zero mean. The variance and covariance we leave as an exercise (Problem 27.1). For the purposes of simulation, given two independent, standard-normal random numbers z_1 and z_2 (i.e., variance 1 and mean 0), it is not hard to verify that the linear combinations

$$\begin{aligned} \Delta W_n &= \sqrt{\Delta t}\, z_1 \\ (I_{10})_n &= \frac{\Delta t^{3/2}}{2}\left(z_1 + \frac{z_2}{\sqrt{3}}\right) \end{aligned} \tag{27.17}$$

have the correct statistics for the two desired quantities. However, we will return to more useful strategies for computing these numbers below.

27.3.2 Multiple Itō Integrals

The notation here generalizes readily to further multiple integrals. In the case of I_0, the zero subscript indicates a simple integration of dt over the time interval; for I_1, the unit subscript indicates a simple integration of $dW(t)$ instead. For the double-integrals, the subscripts have the same meaning, but the result of integrating according to the first index is integrated again according to the second integral; thus, I_{10} means to integrate dW, and then integrate the result with a the differential dt. In the general case, with $I_{j_1 j_2 j_3 \ldots j_n}$, where $j_\alpha \in \{0, 1\}$, we again integrate dt or dW according to the value of j_1, then integrate the result over dt or dW according to the value of j_2, and so on.

Clearly, when we iterate the procedure leading to Eq. (27.8) to generate the stochastic Taylor expansion, we are generating on the very next iteration triple integrals, such as I_{110}. Roughly speaking, when counting order we should count each 1 as a 1/2 order, while a 0 is a full extra order. Thus, the Taylor expansion to order Δt involves I_0, I_1, and I_{11}. To order $\Delta t^{3/2}$, the expansion also involves I_{10}, I_{01}, and I_{111}; while to order Δt^2, the expansion also involves I_{00}, I_{110}, I_{101}, I_{011}, and I_{1111}. It is not hard to see that the stochastic Taylor expansion is much more complicated than the regular Taylor expansion: the regular Taylor expansion only involves I_0, I_{00}, I_{000}, and so on. In terms of actual calculations, there is the additional complication that at higher orders, new random quantities are introduced that cannot be expressed entirely in terms of lower-order quantities. Again, these integrals may be written in terms of other integrals, although this becomes complicated where many indices are involved; however, it can be shown that[4]

$$(\Delta W)^j I_{j_1 \ldots j_n} = \sum_{i=0}^{n} I_{j_1 \ldots j_{i-1} j j_{i+1} \ldots j_n} + \sum_{i=1}^{n} j\, \delta_{j j_i} \Delta W\, I_{j_1 \ldots j_{i-1} 0 j_{i+1} \ldots j_n}, \tag{27.18}$$

where recall that j and all j_i are either 0 or 1.

It is precisely the existence of extra terms that causes traditional numerical methods for ODEs to fail in general for SDEs. In certain special cases (such as additive noise), ordinary methods may provide better performance (since certain terms in the stochastic Taylor expansion will vanish). However, in general different methods must be developed to handle SDEs at "high" order.[5]

[4] Kloeden and Platen, *op. cit.*, Proposition 5.2.3, p. 170.

[5] see Kevin Burrage, Pamela Burrage, Desmond J. Higham, Peter E. Kloeden, and Eckhard Platen, "Comment on 'Numerical methods for stochastic differential equations,' " *Physical Review E* **74**, 068701 (2006) (doi: 10.1103/PhysRevE.74.068701), which is a comment on the use of standard Runge–Kutta methods for SDEs by Joshua Wilkie, "Numerical methods for stochastic differential equations," *Physical Review E* **70**, 017701 (2004) (doi: 10.1103/PhysRevE.70.017701).

27.3.3 Expression for the Taylor Expansion

As we have indicated thus far, the stochastic Taylor expansion is quite complicated. We will simply state the result here for the Itō–Taylor expansion[6] for the solution $y(t)$ to the scalar SDE (27.1), which is

$$y(t) = y(t_0) + \sum_{n=1}^{\infty} \sum_{j_1 \dots j_n = 0}^{1} I_{j_1 \dots j_n} f_{j_1 \dots j_n}, \tag{27.19}$$
(Itō–Taylor expansion)

where the multiple stochastic integrals $I_{j_1 \dots j_n}$ are defined over the time interval from t_0 to t, and the coefficient functions are defined recursively in terms of the functions with one index removed:

$$f_{j_1 \dots j_n} = L^{j_1} f_{j_2 \dots j_n}, \tag{27.20}$$

The lowest-order case of f (with all indices removed) is defined by $f = y$, and the operators L^j are defined by

$$\begin{aligned} L^0 &= \partial_t + \alpha \partial_y + \frac{1}{2} \beta^2 \partial_y^2 \\ L^1 &= \beta \partial_y. \end{aligned} \tag{27.21}$$

Thus, for example,

$$\begin{aligned} f_0 &= L^0 y = \alpha \\ f_1 &= L^1 y = \beta \\ f_{11} &= L^1 f_1 = \beta \partial_y \beta \\ f_{01} &= L^0 f_1 = \partial_t \beta + \alpha \partial_y \beta + \frac{1}{2} \beta^2 \partial_y^2 \beta. \end{aligned} \tag{27.22}$$

Clearly, these coefficient functions become much more complicated as the order increases. However, it should also be reasonably clear how this expansion comes about by the iterated application of the Itō integral forumula. This is one case where Stratonovich calculus is fairly nice: the corresponding Stratonovich formulae for the multiple integrals and coefficients are simplified, for example, by the absence of the final term in the expression for L^0.

27.3.4 Multiple Wiener Processes

As you can imagine, a *vector* SDE driven by multiple Wiener processes dW_k is *way* more complicated than the scalar case. In particular, the multiple integrals must be generalized to indices beyond 0 and 1, so that there are many more terms in the Taylor expansion, and the differential operators must be similarly generalized. In any case, some extensions of the methods we show below to vector Wiener processes are given by Kloeden and Platen[7]

27.4 Stochastic Runge–Kutta Methods

The Milstein method (27.12) is the simplest example of ‘‘Taylor’’ methods that explicitly cancel the higher-order terms in the Itō–Tayler expansion. The problem with these methods is that they require not only specifying the functions α and β in the SDE (27.1), but also their derivatives. In the Milstein case, we have to specify β', but for higher-order methods, more and more derivatives must be specified. This is not a problem in principle, but merely for convenience of the user.

Instead, we can try an approach following the ODE case of Runge–Kutta methods, where multiple function evaluations with different arguments are used to estimate the information that would otherwise

[6] Kloeden and Platen, *op. cit.*, Section 5.5, p. 181. The Itō–Taylor expansion given in this reference is much more rigorous than what we have written here; here, we have written what amounts to an infinite number of iterations of the Itō integral formula. Truncations of the expansion should in principe be done carefully.

[7] Kloeden and Platen, *op. cit.*

be supplied by the derivatives. Again, though, these are not the same Runge–Kutta methods as in the deterministic case, but different methods designed to cancel error terms in the Itō–Taylor expansion above. We will simply quote two Itō methods here.[8] An $O(\Delta t)$ method (i.e., a stochastic Runge–Kutta method of comparable accuracy to the Milstein method) is

$$\begin{aligned} d_n &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\sqrt{\Delta t} \\ y_{n+1} &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\,\Delta W_n + \frac{1}{2\sqrt{\Delta t}}[\beta(d_n, t_{n+1}) - \beta(y_n, t_n)][\Delta W_n^2 - \Delta t]. \end{aligned}$$

(stochastic Runge–Kutta method, order 1) (27.23)

An $O(\Delta t^{3/2})$ method is

$$\begin{aligned} d_{1\pm} &= y_n + \alpha(y_n, t_n)\,\Delta t \pm \beta(y_n, t_n)\sqrt{\Delta t} \\ d_{2\pm} &= d_{1+} \pm \beta(d_{1+}, t_{n+1})\sqrt{\Delta t} \\ y_{n+1} &= y_n + \beta(y_n, t_n)\,\Delta W_n + \frac{1}{2\sqrt{\Delta t}}\,[\alpha(d_{1+}, t_{n+1}) - \alpha(d_{1-}, t_{n-1})]\;(I_{10})_n \\ &\quad + \frac{1}{4}\,[\alpha(d_{1+}, t_{n+1}) + 2\alpha(y_n, t_n) + \alpha(d_{1-}, t_{n-1})]\;\Delta t \\ &\quad + \frac{1}{4\sqrt{\Delta t}}\,[\beta(d_{1+}, t_{n+1}) + \beta(d_{1-}, t_{n-1})]\;\left(\Delta W_n^2 - \Delta t\right) \\ &\quad + \frac{1}{2\Delta t}\,[\beta(d_{1+}, t_{n+1}) - 2\beta(y_n, t_n) + \beta(d_{1-}, t_{n-1})]\;[\Delta W_n\,\Delta t - (I_{10})_n] \\ &\quad + \frac{1}{4\Delta t}\,[\beta(d_{2+}, t_{n+1}) - \beta(d_{2-}, t_{n-1}) - \beta(d_{1+}, t_{n+1}) + \beta(d_{1-}, t_{n-1})]\;\left(\frac{\Delta W_n^2}{3} - \Delta t\right)\Delta W_n. \end{aligned}$$

(stochastic Runge–Kutta method, order 3/2) (27.24)

At the time of writing, there seems to have been no true $O(\Delta t^2)$ methods that have been reported for the general (multiplicative noise) case.

To see the relative performance of the schemes we have presented thus far, we test them on the sample problem[9]

$$dy = -(a + b^2 y)(1 - y^2)\,dt + b(1 - y^2)\,dW, \tag{27.25}$$

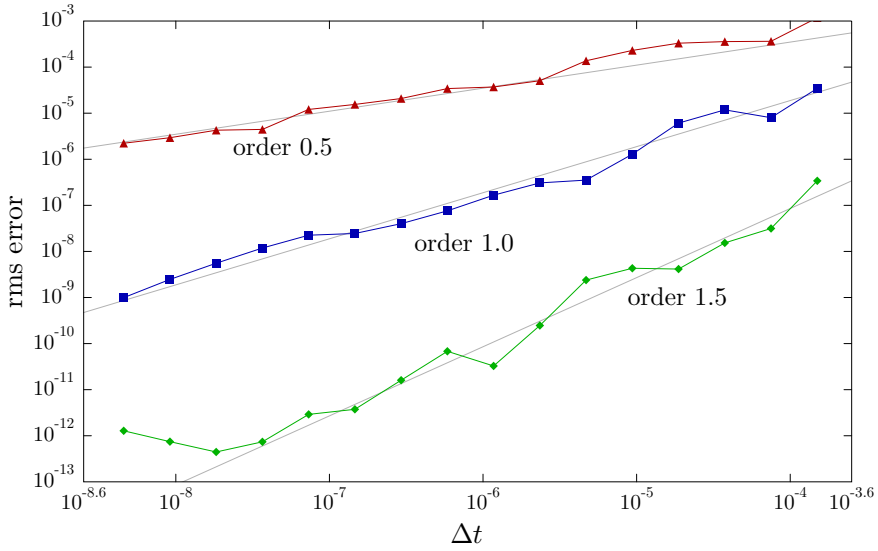
where a and b are constants. This is a multiplicative SDE with the analytic solution

$$y(t) = \frac{(1 + y_0)\exp[-2at + 2bW(t)] + y_0 - 1)}{(1 + y_0)\exp[-2at + 2bW(t)] + 1 - y_0)}. \tag{27.26}$$

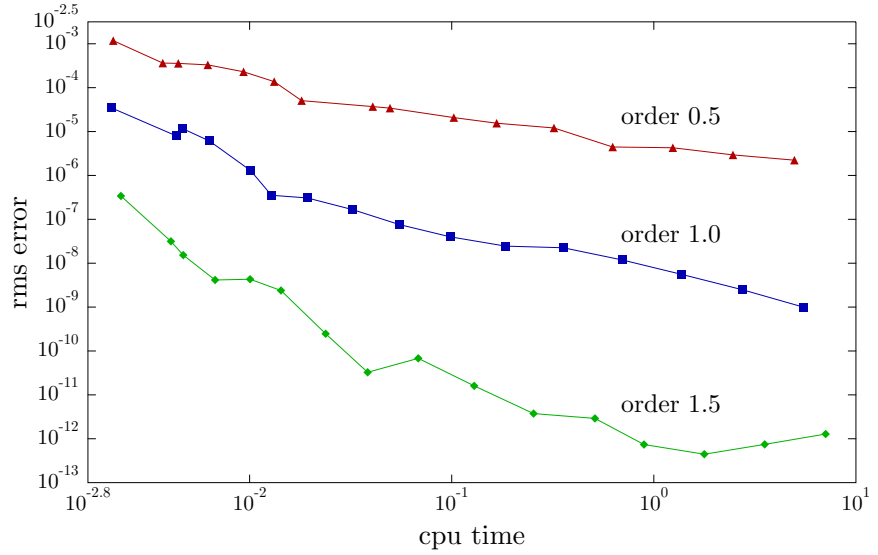
For this calculation we took $a = b = 1$ and $y_0 = 0$. The base step size was $\Delta t = 0.00015$, with the other runs having the same base step size reduced by powers of 2, down to $\Delta t = 0.00015/2^{15}$. The solutions were computed from $t = 0$ out to $t = 0.045$ (300 of the largest steps), and the rms error was computed for a single trajectory in each case. Every calculation was performed on the same Brownian path $W(t)$ (see below). The following plot compares the accuracy of the stochastic Euler method (27.3), the $O(\Delta t)$ Runge–Kutta method (27.23), and the $O(\Delta t^{3/2})$ Runge–Kutta method (27.24).

[8] Kloeden and Platen, *op. cit.*; P. E. Kloeden and E. Platen, "Higher-Order Implicit Strong Numerical Schemes for Stochastic Differential Equations," *Journal of Statistical Physics* **66**, 283 (1992) (doi: 10.1007/BF01060070).

[9] K. Burrage and P. M. Burrage, "High strong order explicit Runge–Kutta methods for stochastic ordinary differential equations," *Applied Numerical Mathematics* **22**, 81 (1996); Peter E. Kloeden and Eckhard Platen, *Numerical Solution of Stochastic Differential Equations*, 3rd ed. (Springer, 2000).



The scaling behavior in each case is clear, though again the convergence is slow compared to ODEs, and the step sizes required to achieve high accuracy are quite small. Note that in the $O(\Delta t^{3/2})$ case, rounding dominates the error at the smallest step sizes. In terms of cpu time, the higher order methods are again superior in this problem.



In this case, each method takes a similar amount of time for a given time step Δt. This is because the test SDE is very simple, and the cpu effort is dominated by random-number generation and construction of the Brownian path. For very complicated SDEs, there will obviously be more of a difference in run times between the different methods. However, it is clear that higher-order methods still provide substantial benefit despite the added complexity.

27.5 Implicit Schemes

As we discussed in Section 24.2.2, implicit methods have some advantages in certain classes of deterministic ODEs. It is therefore natural to try to construct similar implicit methods for SDEs. Unfortunately, we can quickly run into problems. Consider the simple SDE

$$dy = -ay\,dt - by\,dW \tag{27.27}$$

with constant coefficients a and b. The stochastic Euler method for this SDE reads

$$y_{n+1} = y_n - a y_n \,\Delta t - b y_n \,\Delta W_n. \tag{27.28}$$

The obvious generalization of the impicit Euler method for deterministic ODEs is to evaluate the functions on the right-hand side at time t_{n+1} instead of t:

$$y_{n+1} = y_n - a y_{n+1} \,\Delta t - b y_{n+1} \,\Delta W_n. \tag{27.29}$$

Note that ΔW_n is unchanged, because it must be chosen according to the convention for Itō SDEs. Solving this equation for y_{n+1},

$$y_{n+1} = \frac{y_n}{1 + a\,\Delta t + b\,\Delta W_n}. \tag{27.30}$$

Clearly there is a problem here: ΔW_n is a stochastic variable that can take on essentially any value with nonzero probability, and typically takes on values of the order of $\sqrt{\Delta t}$. For small Δt it may thus happen that the denominator can come very close to zero, when ΔW_n takes on values near $-(a/b)\Delta t$. Thus we lose the stability properties we gained in the deterministic case, due just to the stochastic nature of the Wiener process. The problem is apparently just with the *stochastic* part of the equation. There is no problem, for example, if we choose a hybrid scheme where the deterministic step is taken implicitly, but the stochastic parts are treated explicitly, as in

$$y_{n+1} = y_n - a y_{n+1} \,\Delta t - b y_n \,\Delta W_n \tag{27.31}$$

for the example problem, which becomes

$$y_{n+1} = \frac{y_n}{1 + a\,\Delta t} - b y_n \,\Delta W_n. \tag{27.32}$$

which has no special problems with small denominators.

Again, we will simply quote two implicit Itō methods here,[10] corresponding to the explicit $O(\Delta t)$ and $O(\Delta t^{3/2})$ methods above. An $O(\Delta t)$ method (i.e., a stochastic Runge–Kutta method of comparable accuracy to the Milstein method) is

$$\begin{aligned} d_n &= y_n + \alpha(y_n, t_n)\,\Delta t + \beta(y_n, t_n)\sqrt{\Delta t} \\ y_{n+1} &= y_n + \frac{1}{2}\left[\alpha(y_n, t_n) + \alpha(y_{n+1}, t_{n+1})\right]\,\Delta t + \beta(y_n, t_n)\,\Delta W_n \\ &\quad + \frac{1}{2\sqrt{\Delta t}}[\beta(d_n, t_{n+1}) - \beta(y_n, t_n)][\Delta W_n^2 - \Delta t]. \end{aligned}$$

(implicit stochastic Runge–Kutta method, order 1) (27.33)

Here, we have taken a simple average of $\alpha(y_n, t_n)$ and $\alpha(y_{n+1}, t_{n+1})$. Of course, any weighted average could

[10] Kloeden and Platen, *op. cit.*; P. E. Kloeden and E. Platen, "Higher-Order Implicit Strong Numerical Schemes for Stochastic Differential Equations," *Journal of Statistical Physics* **66**, 283 (1992) (doi: 10.1007/BF01060070).

be taken, interpolating between explicit and fully implicit. Similarly, an $O(\Delta t^{3/2})$ method is

$$
\begin{aligned}
d_{1\pm} &= y_n + \alpha(y_n,t_n)\,\Delta t \pm \beta(y_n,t_n)\sqrt{\Delta t} \\
d_{2\pm} &= d_{1+} \pm \beta(d_{1+},t_{n+1})\sqrt{\Delta t} \\
y_{n+1} &= y_n + \beta(y_n,t_n)\,\Delta W_n + \frac{1}{2}\left[\alpha(y_{n+1},t_{n+1}) + \alpha(y_n,t_n)\right]\Delta t \\
&\quad + \frac{1}{4\sqrt{\Delta t}}\left[\beta(d_{1+},t_{n+1}) + \beta(d_{1-},t_{n-1})\right]\left(\Delta W_n^2 - \Delta t\right) \\
&\quad + \frac{1}{2\Delta t}\left[\beta(d_{1+},t_{n+1}) - 2\beta(y_n,t_n) + \beta(d_{1-},t_{n-1})\right]\left[\Delta W_n\,\Delta t - (I_{10})_n\right] \\
&\quad + \frac{1}{2\sqrt{\Delta t}}\left[\alpha(d_{1+},t_{n+1}) - \alpha(d_{1-},t_{n-1})\right]\left[(I_{10})_n - \frac{\Delta W_n\,\Delta t}{2}\right] \\
&\quad + \frac{1}{4\Delta t}\left[\beta(d_{2+},t_{n+1}) - \beta(d_{2-},t_{n-1}) - \beta(d_{1+},t_{n+1}) + \beta(d_{1-},t_{n-1})\right]\left(\frac{\Delta W_n^2}{3} - \Delta t\right)\Delta W_n.
\end{aligned}
$$

(stochastic Runge–Kutta method, order 3/2) (27.34)

This form is specific to the choice of a half "degree of implicitness" (that is, an average of α at the present and advanced times). When comparing these methods to the corresponding explicit methods in the last section on the example problem (27.25), the performance is about the same (with a slight advantage in the order 1 case) for a given time step, and the cpu time is again about the same since the calculation is dominated by the random-number generation, not by the implementation of the finite-difference formulae or by the functional iteration in the implicit schemes.

Of course, similar tricks can be done with any of the schemes we have presented so far: the deterministic step can be taken with high-order deterministic methods (such as fourth-order Runge–Kutta), so long as the stochastic parts are treated according to SDE-adapted methods as presented here.

The formulae here and in the last section have been implemented in a publicly-available Fortran 90 module.[11] This module includes the facilities described below to generate consistent Brownian paths with different step sizes, and a sample code implementing the test problem above is included.

27.6 Strong and Weak Convergence

Up until now, we have been considering a particular kind of convergence of solutions of SDEs, **strong convergence**. This means we are considering *pathwise* convergence of solutions. If $\tilde{y}(t;\Delta t)$ is a finite-difference approximation to the true solution y(t) to the SDE (27.1), then the method for generating $\tilde{y}(t;\Delta t)$ is of **strong order** γ if at fixed t,

$$\langle\langle y(t) - \tilde{y}(t;\Delta t)\rangle\rangle = O(\Delta t^{\gamma}). \qquad (27.35)$$

(strong convergence condition)

The order here again refers to the scaling behavior of the *global* error. Another less-demanding convergence criterion refers to convergence of ensemble means. The same numerical method is said to have **weak order** δ if for every polynomial $g(y)$,

$$\langle\langle g(y(t))\rangle\rangle - \langle\langle g(\tilde{y}(t;\Delta t))\rangle\rangle = O(\Delta t^{\delta}) \qquad (27.36)$$

(weak convergence condition)

at fixed time t. Strong convergence at order γ implies weak convergence of at least the same order. For example, we argued that the stochastic Euler method has strong order 1/2. This is due to an error term proportional to I_{11}, which has zero mean; when considering expectation values, this error term is wiped out, and the stochastic Euler scheme is actually of weak order 1.[12] The Milstein method turns out to have

[11] `SDERK90`, available online at `http://atomoptics.uoregon.edu/~dsteck/computer.html`.

[12] G. N. Milshtein, "A Method of Second-Order Accuracy Integration of Stochastic Differential Equations," *Theory of Probability and its Applications* **23**, 396 (1978) (doi: 10.1137/1123045).

both strong and weak order 1.[13] Since we are normally thinking about solutions to SDEs as individually interesting objects (modeling individual realizations of a continuous measurement process, for example), we will generally be concerned with the more difficult case of strong convergence.

27.7 Consistent Brownian Paths

In considering the strong convergence of the solution to an SDE, it is in general easy to check the convergence where the exact solution is known, as in the test problem above. However, this is obviously more difficult in the generic case where the exact solution is *not* known. In deciding whether or not to accept a particular numerical solution, in the deterministic case you would usually just run the integration again, but with a different step size (say half the original step size). If the two solutions match to within some tolerance, then you can accept the solution. But this is trickier with SDEs. Suppose that you use a random number generator to generate a sequence of Wiener increments

$$\Delta W_0,\ \Delta W_1,\ \Delta W_2,\ \ldots, \tag{27.37}$$

where $\langle\langle \Delta W_n^2 \rangle\rangle = \Delta t$. To cut the step size in half and rerun things, first of all, the same set of random numbers must be used, otherwise the two runs will not in general be well correlated. But even if we use the same set of random numbers to generate the new Wiener increments,

$$\Delta W_0^{(1/2)},\ \Delta W_1^{(1/2)},\ \Delta W_2^{(1/2)},\ \Delta W_3^{(1/2)},\ \Delta W_4^{(1/2)},\ \Delta W_5^{(1/2)},\ \ldots, \tag{27.38}$$

where $\langle\langle (\Delta W_n^{(1/2)})^2 \rangle\rangle = \Delta t/2$, we will *still* have problems, because the increments don't line up: the first random number in the two cases generated ΔW_0 and $\Delta W_0^{(1/2)}$, while the second random number generated ΔW_1 and $\Delta W_1^{(1/2)}$. However, ΔW_1 and $\Delta W_1^{(1/2)}$ don't correspond to the same absolute time. In fact, for consistency, what we require is

$$\Delta W_0 = \Delta W_0^{(1/2)} + \Delta W_1^{(1/2)};\quad \Delta W_1 = \Delta W_2^{(1/2)} + \Delta W_3^{(1/2)};\quad \Delta W_3 = \Delta W_4^{(1/2)} + \Delta W_5^{(1/2)}, \tag{27.39}$$

and so on. These conditions allow both sequences to correspond to time integrals $(I_1)_n$ of the same ideal, particular realization of the Wiener process $W(t)$, which we will refer to as a particular **Brownian path**. Thus, what we require is a procedure for constructing sequences of Wiener increments with different step sizes, but corresponding to different paths.

One convenient method for doing this is to start with the sequence of Wiener increments ΔW_n on the coarser time step Δt, and then to *refine* it consistently onto the finer time grid of step $\Delta t/2$.[14] That is, given a Gaussian Wiener increment ΔW_n with variance Δt, we will use another Gaussian random variable z of unit variance (i.e., standard normal) to generate two new, uncorrelated Gaussian random numbers $\Delta W_n^{(1)}$ and $\Delta W_n^{(2)}$, each of variance $\Delta t/2$ and satisfying $\Delta W_n^{(1)} + \Delta W_n^{(2)} = \Delta W_n$. These conditions are clearly satisfied if we choose[15]

$$\begin{aligned} \Delta W_n^{(1)} &= \frac{1}{2}\Delta W_n + \frac{\sqrt{\Delta t}}{2}\, z \\ \Delta W_n^{(2)} &= \frac{1}{2}\Delta W_n - \frac{\sqrt{\Delta t}}{2}\, z. \end{aligned} \tag{27.40}$$

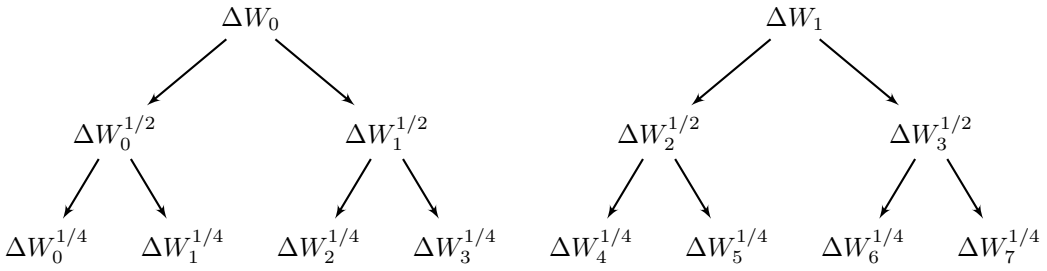
(refinement of Brownian path)

This procedure may obviously be iterated to obtain consistent Brownian paths with yet smaller time steps (by powers of two), creating a tree structure of Wiener increments, with different levels corresponding to different step sizes.

[13] G. N. Milshtein, *op. cit.*

[14] J. G. Gaines and T. J. Lyons, "Variable Step Size Control in the Numerical Solution of Stochastic Differential Equations," *SIAM Journal on Applied Mathematics* **57**, 1455 (1997) (doi: 10.1137/S0036139995286515).

[15] Paul Lévy, *Processus Stochastiques et Mouvement Brownien* (Gauthier–Villars, 1948).



Descending one level on this Brownian tree involves generating a new random number for each Wiener increment on the coarser level, and then combining them as in Eqs. (27.40) to generate the Wiener increments on the finer level. This procedure may be continued indefinitely to consistently realize Brownian paths with arbitrarily fine steps.

This procedure allows you, for example, to run one simulation with one time step Δt and another $\Delta t/2$ on the same Brownian path, and then compare the two runs to estimate the numerical error. Or you can run many different step sizes and examine the convergence behavior of the solution. Consistency of the Brownian path is also an essential first step in implementing adaptive time-step methods for SDEs, which are much more difficult than for ODEs. (The stochastic Euler method, for example, as a basis for an adaptive method can lead to convergence to a wrong solution![16])

27.7.1 Consistent Multiple Itō Integrals

One remaining point to address is that for the order 3/2 Runge–Kutta methods, we also need to refine $(I_{10})_n$ consistently onto finer time grids. One conceptually simple way to generate the $(I_{10})_n$ in the first place is to generate the Wiener increments on a *finer* scale than needed for the integration. That is, if we intend to generate a finite-difference solution to an SDE with time step Δt, Wiener increments ΔW_n, and multiple integrals $(I_{10})_n$, then we should start by generating *finer* Wiener increments δW_n for step size $\delta t < \Delta t$. A reasonable choice in practice would be, say, $\delta t = \Delta t/20$ (although this also works if $\delta t = \Delta t/2$). Then compute ΔW_n by the simple sum,

$$\Delta W_n = \sum_{j=1}^{N_\delta} \delta W_j, \tag{27.41}$$

where N_δ is the number of temporal substeps per integration step (20 or 2, as we mentioned above). This expression is exact, and hasn't gotten us anything because we could have just picked ΔW_n directly. The point is that we may also compute $(I_{10})_n$ via the deterministic Euler method

$$(I_{10})_n = \int_{t_n}^{t_{n+1}} dt'\, W(t') \approx \delta t \sum_{j=2}^{N_\delta} \left(\sum_{k=1}^{j} \delta W_k \right). \tag{27.42}$$

We would normally count the global error of this approximation as $O(\delta t) = O(\Delta t)$ if we keep N_δ fixed. However, $(I_{10})_n$ itself is $O(\Delta t^{3/2})$, and so accounting for the *coefficient* of the global error term, the error in the approximation (27.42) is $O(\Delta t^{3/2}\delta t) = O(\Delta t^{5/2})$. Thus, this approximation is easily adequate for the order 1.5 methods (27.24) and (27.34). Furthermore, from the last section we know how to refine the increments δW_n onto a finer time grid, and thus we can consistently generate the $(I_{10})_n$ also with finer time steps on the same Brownian path.

[16] Gaines and Lyons, *op. cit.*

27.7.1.1 Lévy Areas

A better approximation to the integral $(I_{10})_n$ is to compute the **Lévy area**[17]

$$(A_{ij})_n := \frac{1}{2}\left(\int_{t_n}^{t_{n+1}} dW^{(j)}(t') \int_t^{t'} dW^{(i)}(t'') - \int_{t_n}^{t_{n+1}} dW^{(i)}(t') \int_t^{t'} dW^{(j)}(t'')\right) = \frac{1}{2}\left[(I_{ij})_n - (I_{ji})_n\right], \quad \text{(Lévy area)} \tag{27.43}$$

where in the notation here, $dW^0 \equiv dt$ and $dW^1 \equiv dW$. We can also approximate this by the (stochastic) Euler scheme (which is equivalent to the Milstein scheme for these integrals),

$$(A_{ij})_n \approx \frac{1}{2}\left(\sum_{p=2}^{N_\delta} \delta W_p^{(j)} \sum_{q=1}^{p-1} \delta W_q^{(i)} - \sum_{p=2}^{N_\delta} \delta W_p^{(i)} \sum_{q=1}^{p-1} \delta W_q^{(j)}\right). \tag{27.44}$$

In particular, the Lévy area that we want to approximate is

$$\begin{aligned}(A_{10})_n &\approx \frac{1}{2}\left(\sum_{p=2}^{N_\delta} \delta t \sum_{q=1}^{p-1} \delta W_q - \sum_{p=2}^{N_\delta} \delta W_p \sum_{q=1}^{p-1} \delta t\right) \\ &= \frac{\delta t}{2}\left(\sum_{p=2}^{N_\delta} \sum_{q=1}^{p-1} \delta W_q - \sum_{p=2}^{N_\delta} \delta W_p (p-1)\right).\end{aligned} \tag{27.45}$$

Then since

$$(A_{10})_n = \frac{1}{2}\left[(I_{10})_n - (I_{01})_n\right] = (I_{10})_n - \frac{1}{2}\Delta t\, \Delta W_n, \tag{27.46}$$

where we used Eqs. (27.14), we can compute $(I_{10})_n$ based on this approximation to the Lévy area.

To see why the Lévy area is better despite the extra complication of the formula, we can compute the variance of the numerical approximant (27.42) to $(I_{10})_n$:

$$\begin{aligned}\left\langle\left\langle (\hat{I}_{10})_n^2 \right\rangle\right\rangle &= \delta t^2 \left[(N_\delta - 1)^2 \left\langle\left\langle \delta W_1^2 \right\rangle\right\rangle + (N_\delta - 1)^2 \left\langle\left\langle \delta W_2^2 \right\rangle\right\rangle + (N_\delta - 2)^2 \left\langle\left\langle \delta W_3^2 \right\rangle\right\rangle + \cdots + (1)^2 \left\langle\left\langle \delta W_{N_\delta}^2 \right\rangle\right\rangle\right] \\ &= \delta t^3 \left[(N_\delta - 1)^2 + \sum_{j=1}^{N_\delta - 1} j^2\right] \\ &= \delta t^3 \left[\frac{N_\delta^3}{3} + \frac{N_\delta^2}{2} - \frac{11 N_\delta}{6} + 1\right] \\ &= \frac{\Delta t^3}{3}\left[1 + \frac{3}{2}\frac{\delta t}{\Delta t} - \frac{11}{2}\left(\frac{\delta t}{\Delta t}\right)^2 + 3\left(\frac{\delta t}{\Delta t}\right)^3\right].\end{aligned} \tag{27.47}$$

Here, we used $N_\delta\, \delta t = \Delta t$. The variance *should* be $\Delta t^3/3$, as we wrote in Eq. (27.16). Thus, this numerical approximation gets variance of the generated $(I_{10})_n$'s wrong by an $O(\delta t/\Delta t)$ bias.

[17] Gaines and Lyons, *op. cit.*

For the Lévy area, we can also compute the variance of the numerical approximant:

$$\begin{aligned}\left\langle\left\langle (\hat{A}_{10})_n^2 \right\rangle\right\rangle &= \frac{\delta t^2}{4}\Big[(N_\delta - 1)^2 \left\langle\left\langle \delta W_1^2 \right\rangle\right\rangle + (N_\delta - 1 - 1)^2 \left\langle\left\langle \delta W_2^2 \right\rangle\right\rangle + (N_\delta - 2 - 2)^2 \left\langle\left\langle \delta W_3^2 \right\rangle\right\rangle \\ &\qquad + \cdots + (1 - (N_\delta - 1))^2 \left\langle\left\langle \delta W_{N_\delta}^2 \right\rangle\right\rangle\Big] \\ &= \frac{\delta t^3}{4}\left[(N_\delta - 1)^2 + \sum_{j=1}^{N_\delta - 1}[j - (N_\delta - j)]^2\right] \\ &= \frac{\delta t^3}{4}\left[\frac{N_\delta^3}{3} - \frac{4N_\delta}{3} + 1\right] \\ &= \frac{\Delta t^3}{12}\left[1 - 4\left(\frac{\delta t}{\Delta t}\right)^2 + 3\left(\frac{\delta t}{\Delta t}\right)^3\right].\end{aligned} \tag{27.48}$$

The true variance of the Lévy area is

$$\left\langle\left\langle (A_{10})_n^2 \right\rangle\right\rangle = \left\langle\left\langle (I_{10})_n^2 \right\rangle\right\rangle + \frac{\Delta t^2}{4}\left\langle\left\langle \Delta W_n^2 \right\rangle\right\rangle - 2\frac{\Delta t}{2}\left\langle\left\langle (I_{10})_n \Delta W_n \right\rangle\right\rangle = \frac{\Delta t^3}{3} + \frac{\Delta t^2}{4}\Delta t - 2\frac{\Delta t}{2}\frac{\Delta t^2}{2} = \frac{1}{12}\Delta t^3, \tag{27.49}$$

and thus the approximant biases the variance by an error of $O[(\delta t/\Delta t)^2]$. Since we can compute the integral $(I_{10})_n$ directly from the Lévy area by adding a quantity that can be computed exactly, the bias to the variance of $(I_{10})_n$ by this method is only $O[(\delta t/\Delta t)^2]$, which is one order better than the previous method.

27.7.1.2 Direct Refinement

Another method for directly refining ΔW_n and $(I_{10})_n$ onto a finer temporal grid, in the spirit of Eqs. (27.40), is to combine them directly with *two* independent, standard-normal random numbers z_1 and z_2, according to the linear transformation[18]

$$\begin{bmatrix} \Delta W_n^{(1)} \\ (I_{10})_n^{(1)} \\ \Delta W_n^{(2)} \\ (I_{10})_n^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{\sqrt{\Delta t}}{4} & -\frac{1}{4} & \frac{3}{2\Delta t} \\ \frac{\Delta t^{3/2}}{8\sqrt{3}} & 0 & -\frac{\Delta t}{8} & \frac{1}{2} \\ 0 & \frac{\sqrt{\Delta t}}{4} & \frac{5}{4} & -\frac{3}{2\Delta t} \\ -\frac{\Delta t^{3/2}}{8\sqrt{3}} & \frac{\Delta t^{3/2}}{8} & \frac{\Delta t}{4} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \Delta W_n \\ (I_{10})_n \end{bmatrix}. \tag{27.50}$$

(refinement of Brownian path)

Here, the $\Delta W_n^{(1)}$ and $\Delta W_n^{(2)}$ are the Wiener increments for the two subintervals of duration $\Delta t/2$, and $(I_{10})_n^{(1)}$ and $(I_{10})_n^{(2)}$ are the corresponding double integrals. This method obviously has the advantage of fewer extra random numbers generated and fewer arithmetic operations for each refinement, compared to the Lévy-area method above. This method also has no bias in terms of the variance of the refined variables.

To see where this method comes from, consider the *inverse* problem, which is that if you have a Wiener path defined on a fine time grid, how do you define the Wiener path on a coarser subgrid? The transformation for the coarse increments ΔW are easy to compute by summing the finer increments, but the transformations for the multiple integrals are more complicated. Consider the multiple integrals over the time interval $[0, t]$, which is divided into N subintervals of time $\Delta = t/N$. Then the multiple integrals

[18] Kevin Burrage, Pamela Burrage, and Taketomo Mitsui, "Numerical solutions of stochastic differential equations — implementations and stability issues," *Journal of Computational and Applied Mathematics* **125**, 171 (2000) (doi: 10.1016/S0377-0427(00)00467-2). Actually, these authors give a more general transformation to two time intervals of *nonequal* duration.

on the full interval in terms of the fine-step integrals may be written (Problem 27.4)

$$
\begin{aligned}
W(t) &= \sum_{n=0}^{N-1} \Delta W(n\Delta t) \\
(I_{10})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{10})_{n\Delta t}^{(n+1)\Delta t} + (N-1-n)\Delta t\, \Delta W(n\Delta t) \Big] \\
(I_{100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^2}{2}(N-1-n)^2 \Delta W(n\Delta t) \Big] \\
(I_{1000})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1000})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{100})_{n\Delta t}^{(n+1)\Delta t} \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 (I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^3}{3!}(N-1-n)^3 \Delta W(n\Delta t) \Big] \\
(I_{11})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
(I_{110})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{110})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
&\qquad + (I_{10})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big].
\end{aligned}
\qquad \text{(coarser Brownian path)} \quad (27.51)
$$

Here, the explicit time ranges are specified for each multiple integral for clarity, so that for example

$$
(I_{10})_{t_1}^{t_2} := \int_{t_1}^{t_2} dt \int_{t_1}^{t} dW(t'). \tag{27.52}
$$

Then the transformations (27.50) follow from writing each refined variable as a linear combination of the coarse variables and the extra random variables, and enforcing consistency of Eqs. (27.51) as well as the proper variances and covariances of the refined variables.

27.8 Random Numbers

The ability to generate quality ''random'' numbers on a computer is obviously an important part of simulating stochastic systems. We will thus spend a bit of time surveying some useful techniques towards these goals.

27.8.1 Uniform Distribution

The workhorse of stochastic simulations is a random-number generator that generates numbers uniformly distributed on the interval $[0, 1)$. These numbers can then be transformed to whatever distribution needed by various methods. Most compilers and canned software packages already include such a generator, so why bother thinking about it? First, you might not want to *trust* just any random-number generator that was handed to you, since there have historically been problems found with random-number generators, some subtle.[19] Second, it is useful to have a *portable* random-number generator, which can generate the exact same sequence of ''random'' numbers under any compiler/architecture, which is impossible using ''built-in'' generators. Doing this generally requires implementing the generator in integer arithmetic to avoid floating-point rounding variations on different architectures. This is very handy when porting your code to

[19] Alan M. Ferrenberg, D. P. Landau, and Y. Joanna Wong, ''Monte Carlo simulations: Hidden errors from 'good' random number generators,'' *Physical Review Letters* **69**, 3382 (1992) (doi: 10.1103/PhysRevLett.69.3382). The authors found that certain Monte-Carlo methods combined with certain random-number generators (which otherwise passed standard statistical tests) produced clear systematic errors in the fourth or fifth decimal places of calculated quantities in the 2D Ising model.

a new platform: the easiest test is just to run it and verify that you get essentially the same output for a particular run. This won't happen if the random numbers are different, however. Finally, having several known methods available allows you to switch methods to make sure you aren't seeing any artifacts due to a particular method.

So just what is a random-number algorithm on a computer? Obviously, it is a *deterministic algorithm*, and ultimately, with the finite precision of any computer, the algorithm will be periodic. The idea is to come up with an algorithm that (1) produces the correct (uniform) distribution, (2) has a period much larger than the number of pseudorandom numbers needed for simulations, and (3) shows no detectable correlations over sets of pseudorandom numbers large enough for useful simulations. The algorithm must generally be chosen carefully to meet these criteria.[20] Useful algorithms may be ''seeded'' with an initial value to produce the same set of random numbers each time. Also, it is desireable that different seeds correspond to different random-number sequences (that is, they should start the generator with initial conditions such that the sequences do not overlap for many iterations. Batteries of statistical tests are available to check the quality of random-number generators, such as Marsaglia's ''Diehard Battery''[21] and the battery of tests described by Knuth.[22] We point out three algorithms here that meet all these criteria, plus some methods for improving the random numbers further.[23]

27.8.1.1 L'Ecuyer's Multiple Recursive Generator

A class of computationally very simple methods go by the name of **linear congruential generators** (LCGs),[24] and implement the recurrence

$$x_{n+1} = (ax_n + c) \mod m \tag{27.53}$$

(linear conguential generator)

in integer arithmetic. These algorithms, implemented with real numbers, would clearly be simple chaotic systems, where the state x_n is stretched by a, shifted by c, and then folded by m. Of course, in integer arithmetic, the sequence is periodic, and the constants here must be chosen carefully to give decent performance (clearly $a = 1$, $c = 1$, and $m = 10$ would be a bad choice!). These methods are popular due to their simplicity, but they can have problems, such as a period that may be short in the context of modern processors, and they can have some problems with statistical tests[25]

A **multiple recursive generator** (MRG) improves on LCGs at the expense of added complexity by expanding the depth of the recurrence. An MRG of order k has the form

$$x_{n+1} = (a_1 x_n + a_2 x_{n-1} + \cdots + a_k x_{n-k+1} + c) \mod m.$$

(multiple recursive generator) (27.54)

L'Ecuyer's combined multiple recursive generator[26] uses *two* MRGs, where the first generator hsa coefficients

$$\begin{aligned} m_1 &= 2\ 147\ 483\ 647 = 2^{31} - 1 \\ a_{11} &= 0 \\ a_{12} &= 63\ 308 \\ a_{13} &= -183\ 326 \\ b_1 &= 0 \end{aligned} \tag{27.55}$$

[20] For a particularly amusing discussion, see Donald E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd ed. (Addison Wesley, 1998), Chapter 3.

[21] George Marsaglia and Wai Wan Tsang, ''Some difficult-to-pass tests of randomness,'' *Journal of Statistical Software* **7**, No. 3, 1 (2002).

[22] Donald E. Knuth, *op. cit.*, Section 3.3, p. 41.

[23] All of the algorithms described here are implemented in the Fortran 90 module/library `RANDOM_PL`, available online at `http://atomoptics.uoregon.edu/~dsteck/computer.html`. Also implemented are the shuffling and combination algorithms for any combination of the three generators, as well as facilities for checkpointing, generating vectors of numbers, generating Gaussian numbers, and running multiple equivalent (but differently seeded) generators in parallel. These algorithms are also implemented in the `SDERK` module mentioned above.

[24] see Donald E. Knuth, *op. cit.*, Sections 3.2 and 3.3 for a detailed discussion of LCGs.

[25] Donald E. Knuth, *op. cit.*, Sections 3.3.3 and 3.3.4.

[26] Pierre L'Ecuyer, ''Combined Multiple Recursive Random Number Generators,'' *Operations Research* **44**, 816 (1996), specifically the generator in Example 4.

while the second has

$$\begin{aligned} m_2 &= 2\ 145\ 483\ 479 \\ a_{21} &= 86\ 098 \\ a_{22} &= 0 \\ a_{23} &= -539\ 608 \\ b_2 &= 0. \end{aligned} \tag{27.56}$$

Some of the coeffiecients are zero as a compromise between speed and quality. The idea then is to run *both* generators simultaneously and *combine* their outputs via

$$x_n = (x_n^{(1)} + x_n^{(2)}) \mod m_1. \tag{27.57}$$

Since the first random-number generator is uniform over the positive range $[0, m_1 - 1]$, we can think of it being a uniform distribution on the circle. Adding the second random number corresponds to a random rotation of the circle, which doesn't affect its distribution. This combination further breaks up any correlations or problems that might occur with each individual generator. This combined generator has a period of about 2^{185}, and outputs integers in the range $[0, 2^{31} - 2]$, which can then be divided by 2^{31} to produce uniform random numbers on $[0, 1)$.

This generator can be seeded in an elegant way that guarantees that different seeds will produce sequences that don't overlap those of other seeds for a maximally long time. First, six initial numbers are needed to start the recurrence, so we can just pick them to be some fixed but otherwise arbitrary, ''random'' numbers. For concreteness, we can choose

$$\begin{aligned} x_1^{(1)} &= 1\ 852\ 689\ 663 \\ x_2^{(1)} &= 1\ 962\ 642\ 687 \\ x_3^{(1)} &= 580\ 869\ 375 \\ x_1^{(2)} &= 2\ 039\ 711\ 750 \\ x_2^{(2)} &= 1\ 671\ 394\ 257 \\ x_3^{(2)} &= 879\ 888\ 250. \end{aligned} \tag{27.58}$$

Now the idea is that with the MRGs, we can efficiently skip ahead to *anywhere* in the random sequence via a divide-and-conquer algorithm.[27] To understand this, first note that the MRGs can be implemented as the linear transformation

$$\begin{bmatrix} x_{n+1} \\ x_n \\ x_{n-1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ a_1 & a_2 & a_3 \end{bmatrix} \begin{bmatrix} x_n \\ x_{n-1} \\ x_{n-2} \end{bmatrix} \mod m =: \mathbf{A} \begin{bmatrix} x_n \\ x_{n-1} \\ x_{n-2} \end{bmatrix} \mod m \tag{27.59}$$

on the three-dimensional state of the generator. Then advancing the generator forward n times is equivalent to instead applying the matrix $(\mathbf{A}^n) \bmod m$ Then since the period is just slightly more than 2^{184}, and if we seed with a single 32-bit integer, there are 2^{32} possible seed values. Thus, we can start different seeds 2^{184-32} or 5.7×10^{45} iterations apart. Thus if s is the seed value, we must apply the matrices $(\mathbf{A}^{2^{184-32}})^s \bmod m$ to the initial state of the random number generators, where s is possibly of the order of 2^{32}. This seems like an absurdly large amount of matrix multiplication just to get started!

The trick behind the divide and conquer algorithm is to compute the large power of the matrix $\mathbf{A}$ *recursively*, according to

$$(\mathbf{A}^n) \mod m = \begin{cases} \mathbf{A} & (n = 1) \\ (\mathbf{A}\mathbf{A}^{n-1}) \mod m & (n > 1,\ n \text{ odd}) \\ (\mathbf{A}^{n/2}\mathbf{A}^{n/2}) \mod m & (n > 1,\ n \text{ even}). \end{cases}$$

(divide-and-conquer algorithm) (27.60)

[27] This algorithm is described in depth by Pierre L'Ecuyer, ''Random Numbers for Simulation,'' *Communications of the ACM* **33**, 85 (1990) (doi: 10.1145/84537.84555).

In this way, $\mathbf{A}^n$ may be computed in only $O(\log_2 n)$ matrix multiplications. Specifically, $(\mathbf{A}^{2^{184-32}}) \bmod m$ may be computed in about 152 matrix multiplications. Then the matrix $(\mathbf{A}^{2^{184-32}})^s \bmod m$ needed to compute the seed can be computed in at most about 32 extra matrix multiplications, which is certainly quite feasible.

27.8.1.2 Knuth's Lagged-Fibonacci Generator

The second generator we mention is Knuth's subtractive, lagged-fibonacci generator.[28] This generator implements the integer recurrence

$$X_j = (X_{j-100} - X_{j-37}) \mod 2^{30}, \tag{lagged–Fibonacci generator} \quad (27.61)$$

which is obviously a multiple recursive generator. However, to improve the random numbers here, 1009 numbers can be generated at a time, but only the first 100 numbers used (the rest to be discarded). Knuth provides an algorithm for initializing the first block of values so that the recurrence can take over; this is done in such a way that each seed in the range $[0, 2^{30} - 3 = 1\ 073\ 741\ 821]$ gives a distinct sequence for roughly at least the first 2^{70} numbers (but take away a factor of 10 to account for discarding numbers). This generator has a period of $2^{29} \cdot (2^{100} - 1)$, or 10^{38} numbers, which is quite good (again, take away a factor of 10 to account for discarding numbers). The generator gives a 32-bit integer output in the range $[0, 2^{31} - 2]$ (note that the least significant bit is always 0, so the number is always odd), which can then be divided by 2^{31} to produce a uniform random number on $[0, 1)$.

27.8.1.3 Mersenne Twister

The Mersenne Twister generator of Matsumoto and Nishimura is rather more complicated than the other algorithms, being of much higher dimension. Thus we do not describe the algorithm, but refer to the original reference, where C source code is included.[29] The reason we mention it is that it has an incredible period of $2^{19937} - 1$, and so by this standard is an extremely good generator, while still being fairly computationally efficient. The state of this generator is represented by 624 32-bit integers, so the same number of initial values are required. This method can be seeded by a single number simply by using it as the first integer, and using a simple LCG method to seed the rest. The period is so long that it is exceedingly unlikely that two seeds will produce overlapping sequences on any reasonable scale.

27.8.1.4 Randomizing Random Numbers

There are a few methods to try to improve the quality of random-number generators like the ones above. The first method is to "shuffle" or "scramble" the output of one random-number generator, based on the output of another.[30] The idea is to maintain a "buffer" table of, say, 64 output numbers filled by the primary generator. Then a random number is chosen from the second generator, and based on this, one of the numbers in the buffer table is chosen to be the output number, and it is replaced by the primary generator. This procedure extends the period of the shuffled sequence to be the least common multiple of the two parent sequences, which for the above methods is to say incredibly long. An alternative is to to a similar shuffle of a random sequence using its own randomness to choose which element of the shuffle table to output. This is the **Bays–Durham shuffle**[31] Then for example, the last output random number is used to select which element of the scramble table to output. Even though only one generator is involved, the method still improves the "randomness" of the parent generator. Both of these shuffling methods break up short-range correlations that may exist in the methods above.

[28] Donald E. Knuth, *op. cit.*, p. 186, Eq. (2) and the following discussion.

[29] Makoto Matsumoto and Takuji Nishimura, "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator," *ACM Transactions on Modeling and Computer Simulation* **8**, 3 (1998) (doi: 10.1145/272991.272995).

[30] Donald E. Knuth, *op. cit.*, p. 33, Algorithm M.

[31] Carter Bays and S. D. Durham, "Improving a Poor Random Number Generator," *ACM Transactions on Mathematical Software* **2**, 59 (1976) (doi: 10.1145/355666.355670); Donald E. Knuth, *op. cit.*, p. 34, Algorithm B.

The other method for combining two random-number generators is the subtraction we mentioned above for the L'Ecuyer algorithm. Combining in a second generator can't make the first one worse, but may help if it has some subtle defects. Combining three random number generators via subtraction and scrambling gives the ultimate security for the paranoid computational scientist.

27.8.2 Gaussian Distribution

To pick random deviates from a *standard normal* distribution, or a Gaussian distribution with zero mean and unit variance, with probability density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{27.62}$$

it turns out to be convenient to do so by generating *two* uniform deviates at a time, and then transform them to Gaussian deviates. The first algorithm we mention for doing this is the **Box–Muller method**,[32] To understand it, we first write the joint distribution for two standard normal deviates as

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \tag{27.63}$$

which we can transform to polar coordinates by setting $f(x,y)\,dx\,dy = f_r(r,\theta)\,dr\,d\theta$ to obtain

$$f_r(r,\theta) = \frac{1}{2\pi} r\, e^{-r^2/2} \tag{27.64}$$

via the usual polar transformation. This is still separable, as we may regard the polar distribution to be the product of

$$f_\theta(\theta) = \frac{1}{2\pi}, \qquad f_r(r) = r\, e^{-r^2/2}, \tag{27.65}$$

where the first distribution is obviously uniform over $[0, 2\pi)$. For the second, we may set

$$du = f_r(r)\,dr = r\, e^{-r^2/2}\,dr = -d\Big(e^{-r^2/2}\Big), \tag{27.66}$$

then it is consistent to identify

$$u = 1 - \Big(e^{-r^2/2}\Big), \tag{27.67}$$

or solving for r,

$$r = \sqrt{-2\log(1-u)}. \tag{27.68}$$

Of course, $1-u$ has the same distribution as u, so we may as well use it instead:

$$r = \sqrt{-2\log u}. \tag{27.69}$$

What this means is that if we take u to be a uniform deviate and use this transformation to find r, then r will have the distribution $f_r(r)$ in Eqs. (27.65). The boundary conditions are that $u = 1$ corresponds to $r = 0$ and $u = 0$ corresponds to $r \longrightarrow \infty$, with the transformation $r(u)$ being a monotonic (and one-to-one) function.

So, to summarize the Box–Muller algorithm, we choose U_1 and U_2 to be uniform deviates on $[0, 1)$, and set $U = U_1$ and $\Theta = 2\pi U_2$. Then $r = \sqrt{-2\log u}$, and since $x = r\cos\theta$ and $y = r\sin\theta$, then the two transformed numbers

$$Z_1 = \sqrt{-2\log(1-U_1)}\,\cos(2\pi U_2), \qquad Z_2 = \sqrt{-2\log(1-U_1)}\,\sin(2\pi U_2), \tag{27.70}$$

(Box–Muller transformation)

[32] G. E. P. Box, Mervin E. Muller, "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics* **29**, 610 (1958).

are two independent, standard-normal deviates. Note here that we were careful to compute the logarithm of $1-U_1$ rather than U_1 here—if U_1 is uniform on $[0,1)$ (as is the case for most random-number generators on computers) it is best to avoid the (rare but problematic) possibility of computing $\log(0)$, whereas $\log(1)$ is no problem at all.

An variation on the Box–Muller method is the **polar Marsaglia method**.[33] Suppose w_1 and w_2 define the coordinates for a uniformly distributed random number inside the *unit circle*. Then let $R^2 = W_1^2 + W_2^2$. From the usual transformation to polar coordinates, the probability density for R is $2R$ (the area element at R has an area that scales as R, and the 2 is fixed by normalization). Then since $d(R^2) = 2R\,dR$, we can identify the probability density for R^2 as being *uniform* on $[0,1)$. Thus, we may transform R to the radius r in the 2D normal distribution via Eq. (27.69) (identifying R with u). Finally, since $\sin\theta = w_2/R$ and $\cos\theta = w_1/R$, we may compute the trigonometric functions in Eq. (27.70) by these simple ratios, so that

$$Z_1 = W_1\sqrt{-\frac{2\log R^2}{R^2}}, \qquad Z_2 = W_2\sqrt{-\frac{2\log R^2}{R^2}}. \tag{27.71}$$
(polar Marsaglia method)

The advantage here is that the sin and cos functions are eliminated, and they may potentially be computationally expensive. Now the remaining issue is, how to pick W_1 and W_2? Simply pick the usual uniform deviates U_1 and U_2 on $[0,1)$, and take $W_{1,2} = 2U_{1,2} - 1$ to create a uniform distribution on the unit box from $[-1,1)$ in both directions. Now compute $R^2 = W_1^2 + W_2^2$. If $R^2 < 1$, then proceed with the transformation; otherwise pick new deviates U_1 and U_2 and try again until you end up with a pair inside the unit circle. The probability on one attempt of succeeding by ending up in the unit circle is $\pi/4 \approx 78.5\%$; on average, the random numbers will need to be picked $(1)(\pi/4) + (2)(1-\pi/4)(\pi/4) + (3)(1-\pi/4)^2(\pi/4) + \cdots = 4/\pi \approx 1.27$ times to successfully generate the two normal numbers.

In principle, the polar Marsaglia method should be faster than the Box–Muller method because the sin and cos functions can be expensive to evaluate. However, this is tempered by the presence of a division and the generation of extra random numbers. To test this, 10^8 normal deviates were generated on a relatively modern computer (2.16 GHz Intel Core Duo) in Fortran 90. Using Knuth's lagged Fibonacci generator, the polar Marsaglia method is indeed faster by about 10-15%. However, using a more complicated generator (the L'Ecuyer combined MRG subtractively mixed with the Mersenne twister, then scrambled by Knuth's generator), the two methods were about equally fast. Thus, either method is acceptable, and the Box–Muller method may even be preferable in that the generator will advance by a known number of iterations when generating normal deviates.

27.8.3 Other Distributions

While not directly useful in simulations of the type of SDEs that we considered at the start of the chapter, the methods so far for generating random numbers are useful in generating random deviates with other distributions useful in quantum optics, so it is worth considering these briefly here.

27.8.3.1 Angular Distributions

One useful class of random-deviate distributions are *angular* distributions, as for choosing random spontaneous-emission directions. A simple idea here is to simply use the rejection method, as we mentioned above in the polar Marsaglia method to convert a uniform variate on the unit box to a uniform variate on the unit circle. In that case, we simply rejected any deviate that fell outside the desired circle. Thus, suppose we have a distribution function $f(\theta,\phi)$ that we wish to simulate. Our procedure will be a "double rejection" as follows:

1. Choose three uniform deviates U_1, U_2, and U_3 on $[0,1)$.
2. Convert them to obtain a uniform deviate in the unit cube (that is, from $[-1,1)$ in all directions) by setting $X = 2U_1 - 1$, $Y = 2U_2 - 1$, and $Z = 2U_3 - 1$.

[33] G. Marsaglia, "Improving the polar method for generating a pair of random variables," Boeing Scientific Research Laboratory report D1-82-0203 (1962); Donald E. Knuth, *op. cit.*, Algorithm P (p. 122).

3. Obtain a uniform deviate in the unit sphere by rejection: Compute $R^2 = X^2 + Y^2 + Z^2$ and continue if $R^2 \le 1$ (and also compute $R = \sqrt{R^2}$); go back to step 1 otherwise.

4. Convert to spherical coordinates by computing

$$\begin{aligned} \Theta &= \tan^{-1} \frac{\sqrt{X^2 + Y^2}}{Z} \\ \Phi &= \tan^{-1} \frac{Y}{X}. \end{aligned} \tag{27.72}$$

   Note that for computing Φ, the arctangent must be defined such that it can distinguish between arguments such as $(X, Y) = (1, 1)$ and $(X, Y) = (-1, -1)$, which would naively give the same answer. This is handled in Fortran and other languages by the `atan2` function.

5. Obtain a deviate with the desired angular distribution by testing to see if

$$U_4 \le \frac{f(\Theta, \Phi)}{\sup\{f(\theta, \phi), \theta \in [0, \pi], \phi \in [0, 2\pi]\}}, \tag{27.73}$$

   where U_4 is another uniform deviate on $[0, 1)$. To save from running the random-number generator again, we can alternately use R^3 in place of U_4, since it is likewise a uniform deviate on $[0, 1)$, and it is independent of the angles. We normalize by the maximum possible value of $f(\theta, \phi)$ to maximize the efficiency of the method. If the condition is true, we're done; if not, start over at step 1.

For example, if we were generating a dipole distribution

$$f(\theta, \phi) = \frac{3}{8\pi} \sin^2 \theta, \tag{27.74}$$

in step 5 we would check to see if

$$R^3 \le \sin^2 \Theta, \tag{27.75}$$

and if so, keep the generated angles. In this example, the success probability for each triple of uniform deviates $U_{1,2,3}$ to generate a uniform deviate in the unit sphere is $\pi/6 \approx 52\%$. Once this happens, the probability of generating a deviate with the dipole distribution pattern is $2/3 \approx 67\%$, for a combined success probability of $\pi/9 \approx 34.9\%$. The first probability is given by the volume ratio of the unit sphere and cube, while the second is given by the volume under the "unwrapped" surface $\sin^3 \theta$ (i.e., *not* the volume inside the boundary $r = \sin^2 \theta$), relative to the same volume under the surface $\sin \theta$, or simply 4π.

Of course, the normalized values X/R, Y/R, and Z/R give the projections of the unit vector pointing in the (Θ, Φ) direction along the x, y, and z directions.

A slightly more efficient method generates the angle values more directly. The normalized probability density for equally-likely emission in any direction is

$$f(\theta, \phi) = \frac{\sin \theta}{4\pi}, \tag{27.76}$$

with normalization convention

$$\int_0^\pi d\theta \int_0^{2\pi} d\phi \, f(\theta, \phi) = 1. \tag{27.77}$$

That is, the ϕ variable is already uniform, so we can handle it by picking a uniform deviate u_2 and rescaling: $\phi = 2\pi u_2$. Thus, we need only consider the θ variable, with normalized density

$$f(\theta) = \frac{1}{2} \sin \theta. \tag{27.78}$$

If we pick a second uniform deviate u_1 and equate probability densities via $du_1 = (1/2) \sin \theta \, d\theta$, note that this is equivalent to equating the cumulative probability functions. For the uniform deviate, the cumulative distribution is simply u_1, while for θ, the cumulative distribution is

$$F(\theta) = \frac{1}{2} - \frac{1}{2} \cos \theta = \sin^2 \theta. \tag{27.79}$$

Equating the cumulative distributions, we obtain $\theta = \sin^{-1}\sqrt{u_1}$. Thus, to summarize, we should choose two uniform deviates U_1 and U_2, and then compute the angles via

$$\Theta = \sin^{-1}\sqrt{U_1}, \qquad \Phi = 2\pi U_2.$$

(transformation from uniform to angular variables) (27.80)

This saves one uniform random deviate compared to the rejection method (plus any rejected deviates), although it requires evaluating the more complicated arcsine function.

27.8.3.2 Exponential Distributions

Another simple example is the exponential distribution,

$$f(x) = \gamma\, e^{-\gamma x}, \tag{27.81}$$

defined on the nonnegative half of the real line. The cumulative density is an exponential rise

$$F(x) = 1 - e^{-\gamma x}, \tag{27.82}$$

and the inverse of this function gives x in terms of a uniform deviate U:

$$X = -\frac{1}{\gamma}\log(1-U). \tag{27.83}$$

(exponential-deviate generation)

Of course, the result is always a positive real number. Also, U can be substituted for U in the logarithm, but if the random-number generator gives $U \in [0,1)$, the above form guards against the possibility of crashing the code by computing $\log(0)$.

27.8.3.3 Power-Law Distributions

Suppose we have a power-law probability density defined between a and b:

$$f(x) = \eta x^{\gamma}, \qquad 0 < a \le x \le b. \tag{27.84}$$

We have chosen the bounds "safely" here, but in some cases they may be negative, too. Here, η is a normalization factor, given by

$$\eta = \begin{cases} \dfrac{\gamma+1}{b^{\gamma+1} - a^{\gamma+1}} & \text{if } \gamma \neq -1 \\ \dfrac{1}{\log b - \log a} & \text{if } \gamma = -1. \end{cases} \tag{27.85}$$

Note that we can safely take $a \longrightarrow 0$ if $\gamma > -1$, otherwise we need $a > 0$. Similarly, we can safely take $b \longrightarrow \infty$ if $\gamma < -1$, otherwise we need $b < \infty$. If $\gamma = 1$, we strictly need $0 < a < b < \infty$.

First, in the case $\gamma \neq -1$, we have a cumulative distribution

$$F(x) = \frac{x^{\gamma+1} - a^{\gamma+1}}{b^{\gamma+1} - a^{\gamma+1}} \qquad (\gamma \neq -1). \tag{27.86}$$

Setting this equal to a cumulative distribution u for a uniform deviate, and solving, we find the algorithm

$$X = \sqrt[\gamma+1]{a^{\gamma+1} + (b^{\gamma+1} - a^{\gamma+1})U} \;=\; \sqrt[\gamma+1]{a^{\gamma+1}(1-U) + b^{\gamma+1}U} \qquad (\gamma \neq -1)$$

(power-law-deviate generation) (27.87)

for a power-law deviate X in terms of a uniform deviate U. However, if $\gamma = -1$, then we have the cumulative distribution

$$F(x) = \frac{\log x - \log a}{\log b - \log a} \qquad (\gamma = -1). \tag{27.88}$$

This leads to the transformation

$$X = \exp\left[\log a + (\log b - \log a)U\right] = \exp\left[(1-U)\log a + U\log b\right] \qquad (\gamma = -1).$$

(power-law-deviate generation) (27.89)

In both cases, the transformation is a weighted average of the endpoints (according to the value U), within a nonlinear transformation (power-law or logarithm).

One context in which the power-law distribution comes up in the Monte-Carlo evaluation of an integral of the form

$$I = \int_0^\infty dx\, x^\gamma f(x) \qquad (\gamma < -1), \tag{27.90}$$

where the restriction on γ is just to simplify the following analysis. Clearly, this integral only converges if $f(x)$ converges to a finite constant as $x \longrightarrow \infty$, and if $f(x)$ converges quickly enough to zero as $x \longrightarrow 0$. For the power-law distribution to do us any good here we must have a lower bound $a > 0$. We will thus further assume that $f(x) = 0$ for all $x < x_{\min}$ (or at least that the region $x < x_{\min}$ contributes negligibly to the integral—we can always find some appropriate boundary that satisfies this condition to any desired precision). Then

$$I = \int_{x_{\min}}^\infty dx\, x^\gamma f(x) \qquad (\gamma < -1), \tag{27.91}$$

and thus we can take advantage of the power-law distribution to turn this into an ensemble average,

$$I = \eta^{-1}\left\langle\!\left\langle f(x)\right\rangle\!\right\rangle_x = \frac{-x_{\min}^{\gamma+1}}{\gamma+1}\left\langle\!\left\langle f(x)\right\rangle\!\right\rangle_x \qquad (\gamma < -1), \tag{27.92}$$

where from Eq. (27.87), we generate the values of x according to

$$X = x_{\min}(1-U)^{1/(\gamma+1)} \qquad (\gamma < -1). \tag{27.93}$$

Note again that we could replace $1-U$ by U to simplify the formula here, but it is safer if U is defined on $[0,1)$ to leave it in this form, so that the (rare) occurrence of the value $U = 0$ does not pose any problem.

27.8.3.4 Gamma Distribution

A more complicated distribution, something of a combination of power-law and exponential distributions, is the **gamma distribution**

$$f(x;\gamma,\beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1}e^{-\beta x} \qquad (x > 0,\ \gamma > 0,\ \beta > 0), \tag{27.94}$$

named thusly due to the gamma function in the normalization factor. In the case of $\gamma = 1$, this is simply an exponential distribution, as in Eq. (27.81).

One useful interpretation of the gamma distribution comes from the property that if $X_1 \sim f(x;\gamma_1,\beta)$ (i.e., X_1 is gamma distributed with parameters γ_1 and β) and $X_2 \sim f(x;\gamma_2,\beta)$, then (Problem 27.2)

$$X = X_1 + X_2 \sim f(x;\gamma_1+\gamma_2,\beta). \tag{27.95}$$

Thus, for example, if γ is an integer, the gamma-distributed deviate X can be interpreted as the sum of γ, independent exponential deviates. This also means that as γ becomes large, $f(x;\gamma,\beta)$ should converge to a Gaussian distribution via the central limit theorem. It is also fairly simple to choose a gamma-distributed deviate if $\gamma = a$ is a relatively small integer, since the combination of exponential deviates

$$X = -\frac{1}{\beta}\sum_{j=1}^{a}\log(1-U_j) = -\frac{1}{\beta}\log\left(\prod_{j=1}^{a}(1-U_j)\right)$$

(gamma-distributed deviate, integer γ) (27.96)

works, for a uniform deviates U_j. The latter form is more efficient on a computer, since it avoids the calculation of multiple (expensive) logarithms.

Another useful special case of the gamma distribution occurs for $\gamma = 1/2$:

$$f(x; 1/2, \beta) = \sqrt{\frac{\beta}{\pi x}}\, e^{-\beta x} \qquad (x > 0,\ \beta > 0). \tag{27.97}$$

Changing variables via $x = z^2/2$ and $dx = z\, dz = \sqrt{2x}\, dz$ leads to

$$f(z; 1/2, \beta) = \sqrt{\frac{2\beta}{\pi}}\, e^{-\beta z^2/2} \qquad (z > 0,\ \beta > 0). \tag{27.98}$$

This is a Gaussian distribution (or at least the positive half) with variance $1/\beta$. Thus, it is also easy to pick a gamma-distributed deviate for any *half*-integer value of γ. That is, if

$$X = \frac{\beta}{2}\, Z^2 - \frac{1}{\beta} \sum_{j=1}^{a} \log(1 - U_j) = \frac{\beta}{2}\, Z^2 - \frac{1}{\beta} \log\left(\prod_{j=1}^{a} (1 - U_j) \right),$$

(gamma-distributed deviate, half-integer γ) (27.99)

where a is a positive integer, Z is a standard-normal deviate, and the U_j are uniform deviates, then X is distributed according to $f(x; \gamma = a + 1/2, \beta)$.

By the same coordinate change, the $\gamma = 3/2$ case,

$$f(x; 3/2, \beta) = \frac{2\beta^{3/2}}{\sqrt{\pi}}\, \sqrt{x}\, e^{-\beta x} \qquad (x > 0,\ \beta > 0), \tag{27.100}$$

becomes

$$f(z; 3/2, \beta) = \sqrt{\frac{\beta^3}{2\pi}}\, z^2\, e^{-\beta z^2/2} \qquad (z > 0,\ \beta > 0). \tag{27.101}$$

Notice that this is the same as a Gaussian distribution of variance $1/\beta$ for three variables x_1, x_2, and x_3, but transformed into spherical coordinates by $z^2 = x_1^2 + x_2^2 + x_3^2$. Thus, for $\gamma = 3/2$, it is possible to create a deviate by summing the squares of three normal deviates (with appropriate scaling), or to add the square of a normal deviate to an exponential deviate as in Eq. (27.99).

For more general γ (i.e., neither integer nor half-integer), it is in general necessary to use more complicated methods, based on various forms of rejection sampling.[34]

[34] Luc Devroye, *Non-Uniform Random Variate Generation* (Springer, 1986), Section IX.3 (ISBN: 0387963057), `http://www.nrbook.com/devroye/`.

27.9 Exercises

Problem 27.1

Let us define ΔW and I_{10} over the same time interval Δt as usual as

$$\begin{aligned} \Delta W &:= \int_0^{\Delta t} dW(t) \\ I_{10} &:= \int_0^{\Delta t} dt \int_0^t dW(t'). \end{aligned} \tag{27.102}$$

Show that the variance of I_{10} and its covariance with ΔW are given by Eq. (27.16),

$$\begin{aligned} \left\langle\!\left\langle I_{10}^2 \right\rangle\!\right\rangle &= \frac{\Delta t^3}{3} \\ \left\langle\!\left\langle I_{10}\, \Delta W \right\rangle\!\right\rangle &= \frac{\Delta t^2}{2} \end{aligned} \tag{27.103}$$

as follows. View I_{10}, defined over a time step Δt, as the solution of the differential equation

$$dy = W(t)\, dt, \tag{27.104}$$

with initial condition $y(0) = 0$, such that $I_{10} = y(\Delta t)$. Then consider a finite-difference approximation to the solution by integrating in N steps of duration $\delta t = \Delta t/N$ according to the Euler method (which becomes exact as $N \longrightarrow \infty$).

Problem 27.2

Prove the addition property of the gamma distribution [Eq. (27.95)]: if $X_1 \sim f(x; \gamma_1, \beta)$ (i.e., X_1 is gamma distributed with parameters γ_1 and β) and $X_2 \sim f(x; \gamma_2, \beta)$, where the gamma probability density is

$$f(x; \gamma, \beta) = \frac{\beta^\gamma}{\Gamma[\gamma]}\, x^{\gamma-1} e^{-\beta x} \qquad (x > 0,\ \gamma > 0,\ \beta > 0), \tag{27.105}$$

then the sum is also gamma-distributed:

$$X = X_1 + X_2 \sim f(x; \gamma_1 + \gamma_2, \beta). \tag{27.106}$$

Problem 27.3

A useful probability density for sampling the $\mathcal{T}$ integral in worldline path integrals is

$$f(\mathcal{T}) = \frac{(2d^2)^{(D+m)/2}}{\Gamma\left[\frac{D+m}{2}\right]} \frac{e^{-2d^2/\mathcal{T}}}{\mathcal{T}^{1+(D+m)/2}}. \tag{27.107}$$

Show that this can be transformed to a gamma distribution, and describe the generation of deviates $\mathcal{T}$ for $D = 4$ and integer $m \geq 0$.

Problem 27.4

Derive the equations (27.51) relating the multiple integrals on the full time interval $[0, t]$ to the multiple

integrals defined on subintervals $\Delta t = t/N$:

$$
\begin{aligned}
W(t) &= \sum_{n=0}^{N-1} \Delta W(n\Delta t) \\
(I_{10})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{10})_{n\Delta t}^{(n+1)\Delta t} + (N-1-n)\Delta t\, \Delta W(n\Delta t) \Big] \\
(I_{100})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{100})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^2}{2}(N-1-n)^2 \Delta W(n\Delta t) \Big] \\
(I_{1000})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{1000})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n)(I_{100})_{n\Delta t}^{(n+1)\Delta t} \\
&\qquad + \frac{\Delta t^2}{2}(N-1-n)^2 (I_{10})_{n\Delta t}^{(n+1)\Delta t} + \frac{\Delta t^3}{3!}(N-1-n)^3 \Delta W(n\Delta t) \Big] \\
(I_{11})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
(I_{110})_0^t &= \sum_{n=0}^{N-1} \Big[(I_{110})_{n\Delta t}^{(n+1)\Delta t} + \Delta t(N-1-n) \Big[(I_{11})_{n\Delta t}^{(n+1)\Delta t} + \Delta W(n\Delta t)\, W(n\Delta t) \Big] \\
&\qquad + (I_{10})_{n\Delta t}^{(n+1)\Delta t} W(n\Delta t) \Big].
\end{aligned}
\tag{27.108}
$$

Index

A gauge, 462
E gauge, 462
P function, 129–130
Q function, 129
T_1 and T_2, 178
σ-algebra, 869
Itō–Taylor expansion, 1169–1172
Itō calculus, 948–953

Abraham–Lorentz force, 39, 451
absolute-value process, 776–777, 802
absorption, 98
absorption coefficient, 57, 102–103
absorption image, 42
absorption oscillator strength, 35–36
ac Stark shift, 162, 272, 273, 397
    due to thermal field, 508
action functional, 403, 910, 914
    expansion, 914, 915
action principle, 403–404
    for Schrödinger equation, 452
    optics, 451
Adams–Bashforth method, 1131
Adams–Moulton methods, 1131
adaptive measurement, 863
adaptive time stepping, 1130
adiabatic approximation, 104, 191, 230–231, 234, 236, 271–273, 528–530
    for momentum, 734–737
    via master equation, 230–231
adiabatic passage, 162–166, 177
adiabatic rapid passage, 166
adiabatic theorem, 163
advanced Green function, 654, 655
advection equation, 239
affine parameter, 993
Aharonov–Bohm effect, 465–467
aliasing, 1139–1140
ancilla, 143
angular momentum, 301–315
anomalous diffusion, 705
antibunching, 76, 834
    resonance fluorescence, 209–211
anticipating calculus, 953–954
antistandard-ordered distribution, 130
arcsine law, 775–777
associated Legendre functions, 412
atom
    two-level, 151–161
atom interferometry, 281–283
atom optics, 39–49
    dipole force, 39–45
    Doppler temperature, 47–49
    magneto-optical trap (MOT), 49
    optical molasses, 46–49
    radiation pressure, 45–49
atom–mirror interaction, 50–53, 57–58, 507, 509, 637–638
atom–photon molecule, 490–491
atomic clock, 171–174
Autler–Townes doublet, 217–221, 256
autocorrelation function, 64
auxiliary functions
    to sine and cosine integrals, 537
averaging operator, 958
avoided crossing, 161

backward differentiation formulae, 1131
backward Euler method, 1125
Baker–Campbell–Hausdorff expansion, 124, 908, 1149
bare states, 161
Bays–Durham shuffle, 1184
beam splitter
    relation among coefficients, 512, 687
Bernoulli numbers, 631
Bessel process, 807–811
Bessel's equation, 413, 455
big-O notation, 1121–1123
Blackman pulse, 277–278
bleaching, 103
Bloch sphere, 166–177
Bloch vector, 168–178
Bloch–Siegert shift, 42, 251–252
Boltzmann statistics, 99
Borel σ-algebra, 870
Born approximation, 145
Born rule, 869
Born–Markov approximation, 145
bound charge density, 579, 1010

bound current density, 580
boundary crossing, 748–754, 758–761
Box–Muller method, 1185–1186
Bragg scattering, 232–235
    rate, 235
branch prediction, 1115–1116
Breit–Rabi formula, 397
Brownian area, 732
Brownian Bridge, 827–828
Brownian bridge, 737–807, 913, 925, 927, 930, 931, 961, 980, 982, 983, 985, 986
    boundary crossing, 750–754, 758–761, 985, 986
    escape, 754–758
    escape probability, 983
    local time, 797–807
    sojourn time, 777–807
Brownian bridges, 988
Brownian excursion, 738
Brownian meander, 738
Brownian motion, 719–720, 925
    reflected, 776–777, 802
    skew, 802–807
Brownian path
    consistent refinement, 1177–1181
bunching, 73–76

cache, 1111–1112
    line, 1111
    thrashing, 1112
Cameron–Martin transformation, 949
Cameron–Martin–Girsanov transformation, 949–950
canonical cosymplectic form, 135–137
canonical quantization, 409
canonical transformation, 136–137, 463
    generating function, 463
Casimir effect
    path integral, 965–989
Casimir energy, 638–647
    temperature dependence, 646–647
Casimir–Polder potential, 42, 464, 533–557, 563–578, 596–635
    atom–atom potentials, 616–621
    comparison to classical dipole, 543–544
    dielectric and general media, 596–635
    excited levels, 632–635
    general planar interface, 601–614
    Lifshitz expression, 608
    path integral, 965–989
    rarefied dielectric, 563–573, 610
    spherical cavity, 614–616
    temperature dependence, 621–632, 988–989
    three atoms, 620–621
Cauchy integral formula, 589, 601, 628
Cauchy principal value, 289, 499, 542, 590, 650
Cauchy principle value, 543
Cauchy probability distribution, 704, 710
Cauchy's integral formula, 635
caustics, 921
cavity damping, 492–493
cavity decay, 146, 491–492, 509–519, 531
    decay rate, 516–519
    nonexponential, 682
cavity driving field, 491
cavity photon number
    measurement of, 492–493
cavity QED coupling constant, 486
center-of-mass coordinates, 475–477
central limit theorem, 697–706
central-limit theorem, 711
Chapman–Kolmogorov equation, 717, 908
characteristic function, 702
chronological operator, 112
Church of the Larger Hilbert Space, 142–143
Clebsch–Gordan coefficients, 304–309
cloning, 131–132
closed set, 870
CNOT gate, 132
coarse-graining approximation, 145
coherence, 63–76
    length, 70–72
    time, 70–72
coherent population trapping, 273, 283–295
    velocity-selective, 286–287
coherent state, 128, 187, 491
collapses and revivals, 487–489
collisions
    dephasing due to, 178
commutator
    with space-dependent mass, 934
composite systems, 130–131
conditional probability
    density, 716
Condon–Shortley phase convention, 307
conductivity, 59
conjugate momentum, 404
constitutive relations, 1009
continuous function, 1122
controlled-NOT gate, 132
convective derivative, 457
convergence
    definition, 1121–1123
convolution, 698–706, 709–710
    of box functions, 699
    of two Gaussians, 700–701
    with δ-function, 698–699
convolution kernel, 699

convolution theorem, 699–700, 1133
cosine integral Ci(x), 537, 574–575
cosymplectic form, 135–137
Coulomb gauge, 401, 422–423, 459
counters, hardware performance, 1119
counting process, 728, 833–835
coupled-mode theory, 511–519
critical sampling, 1134–1135
cross section
  absorption, 35, 101
  natural, on resonance, 101, 102, 192
  scattering, 195–197
crossover resonance, 225–228

D line, 397
dark state, 285
decay rate
  modified by macroscopic body, 50–53, 635–638
decay rate, natural, 271
degree of coherence, 94
  degree of first-order temporal, 694
  first-order, 66
  first-order quantum, 686
  first-order temporal, 66
  second-order quantum, 686
  second-order temporal, 72–76
delta function
  chain rule, 706, 828–829
  derivative of, 794, 828–829
  longitudinal, 423–425
  transformation for surface constraint, 706–710
  transverse, 423–425
density matrix, 109–110
  path integral, 928–930
density operator, 107–110
  purity, 110, 255
derivative process, 820
detector efficiency, 839–840, 853–855, 878–879
diamagnetic interaction, 473, 475
Dicke superradiance, 298–299
dielectric constant, 581
diffusion
  spontaneous emission, 247
diffusion coefficient, 239, 698
diffusion equation, 239
diffusion process, 697
diffusion rate
  discontinuity, 802–807
dipole approximation, 151, 460
dipole force, 39–45, 237–238, 254
  dressed-atom picture, 256–258
dipole interaction Hamiltonian, 151–157, 457, 460–465, 485–486, 495
dipole matrix element
  relation to decay rate, 192
dipole moment
  electric, 474
  magnetic, 474
dipole operator, 152–153, 229
  relation to Bloch vector, 170
direct product, 130–131
direction cosines, 425–426
Dirichlet problem
  Laplace equation, 758–760
  Poisson equation, 760–761
discrete Fourier Transform, 1136–1147
discrete spectrum, 488
dispersion, 580–581
divergence theorem, 706
divergent sequence, 1121
divide-and-conquer algorithm, 1183–1184
domination
  of sequences, 1121
Doppler broadening, 57, 100, 174
Doppler limit, 250–251
Doppler shift, 259, 273, 280
Doppler temperature, 47–49
Doppleron resonance, 224
dressed states, 161–166, 254, 489
drift coefficient, 239
Drude–Lorentz model, 59
dynamical tunneling, 234

effective potential, 150
effective, non-Hermitian Hamiltonian, 185
Einstein A and B coefficients, 98–100, 104–105
Einstein A coefficient, 178
Einstein rate equations, 98–105
EIT, 287–290
electric dipole moment, 474
electric displacement, 435, 579
electric displacement field, 462
electric flux density, 435, 579
electric octupole moment, 474
electric quadrupole moment, 474
electromagnetic field
  commutators, 425–430
  gauge freedom, 431–434
  Hamiltonian structure, 402–405, 431–434
  mass coupling, 455–456
  quantization, 406–420
electromagnetically induced transparency, 287–290
  as position measurement, 899–904
energy spectral density, 64
entanglement, 131–140, 142–143
environment, 143, 144

error tolerance, numerical, 1130
escape probability, 754–758, 930
Euler method, 1123–1125, 1167
    implicit, 1123–1125
    stochastic, 1167–1168
Euler's constant, 537
Euler–Cromer method, 1162–1163, 1165
Euler–Lagrange equation, 403, 405, 451–452, 458
Euler–Maclaurin summation formula, 631
Euler–Maruyama method, 1167–1168
Euler–Mascheroni constant, 575, 671
exchange interaction, 141–142
exponential integral $E_1(x)$, 575, 667, 681

Fabry–Perot cavity, 261, 296, 490, 517, 519, 755
Fano profile, 286
far-off resonance trap (FORT), 42
Fermat's principle, 451
Fermi's Golden Rule, 509–510, 531
Fermi's golden rule, 500–502, 635, 637
Feynman–Kac formula, 761–770, 923, 927–928
Feynman–Vernon–Hellwarth representation, 169
fidelity, 117
finesse, 517
finite-difference method, 1123
finite-difference operator, 916
first variation, 402, 406, 914
first-passage time
    Brownian bridge, 752–754
    Wiener path, 750
fixed-point arithmetic, 1109
fixed-point iteration, 1124
floating-point arithmetic, 1109–1110
fluctuation–dissipation relation, 48, 267, 522, 621–626
fluorescence, resonance, 194–211
Fock state, 187
Fock—Schwinger proper time, 975
Fokker–Planck equation, 238–240, 258–259, 716–718, 826
    discontinuity in, 802–807
    multidimensional, 718, 826
forbidden transition, 343
forward Euler method, 1125
Fourier Transform
    discrete, 1136–1147
Fourier transform, 1133–1147
free charge density, 1010
free spectral range, 517
free-electron gas, 58–60
Fresnel reflection coefficients, 564, 603
FTIR spectroscopy, 67, 94
functional, 402, 406
functional derivative, 402–406, 452, 914
    second, 914
functional determinant, 915, 971
    as eigenvalue product, 973–975
functional integral, 907–962
fundamental charge, 152

Göppert-Mayer transformation, 462–463
gain coefficient, 105
gamma distribution, 1189
gauge transformation, 461, 463–467
Gaussian fluctuations
    complex, 74–76
    exponentiated, 95
Gel'fand–Yaglom method, 916–918
general Legendre equation, 412
generalized Rabi frequency, 159
generalized susceptibility, 592–596
generating function
    for canonical transformation, 463
generator
    for SDE, 717
geometric Brownian motion, 826
geometric optics, 451
Girsanov transformation, 949
Glauber–Sudarshan P function, 129–130
Green function, 648–649, 653–655
    classical harmonic oscillator, 648–649
    for Schrödinger equation, 653–655
Green tensor, 563, 581–586
    for free space, 583–586
    Kramers–Kronig relations, 596
    planar interface, 603–604
    related to field correlation, 595, 596
    scattering part, 601

Hamilton's equations, 404, 405
Hamilton's principle, 403
Hamiltonian, 404–406
Hanbury-Brown–Twiss experiment, 75–76, 685, 688
hardware performance counters, 1119
harmonic function, 759
harmonic oscillator, 31–36
    coherent state, 187
    damped, 33–36, 150
    Fock state, 187
    Green function, 648–649
    quantum, 186–191
    quantum damped, 187–191, 261, 266–267
heat bath, 144
Heisenberg picture, 112–114
Hellmann–Feynman theorem, 640
Helmholtz equation, 407
    Green function, 681, 1002–1003

scalar vs. vector solutions, 414–418, 454
Helmholtz theorem, 421–422
Hermite–Gaussian functions, 121
heterodyne detection, 851–853, 859–863
heterodyne spectroscopy, 79–81
Hilbert transform, 289
homodyne detection, 840–851, 858–859, 861–863, 900
electromagnetically induced transparency, 899–900
Hong–Ou–Mandel dip, 694
Hong–Ou-Mandel effect, 688–694
Husimi distribution, 129
upper bound, 149
hydrogen maser, 174
hyperfine structure, 551–553

IEEE 754 standard, 1109–1110
impedance
of vacuum, 40
implicit Euler method, 1123–1125
stochastic, 1174–1175
indistinguishability, 140–142
inhomogeneous broadening, 174
inlining, procedure, 1117
input–output theory, 519–531
integer arithmetic, computer, 1109
integrated Brownian motion, 732
intensity
of Poisson process, 726
interaction picture, 114, 515
for atom–field interaction, 156–157
interference
between partially coherent sources, 72
coherence, 63–76
two-photon, 688–694
visibility, 69–70
interferometer
Michelson, 66–67, 94
Itō calculus, 712–726
ensemble average, 714–715
Itō rule, 713–714

Jaynes–Cummings model, 485–491, 515
as model for spontaneous emission, 487–489
Johnson noise, 624–626
joint probability density, 717
Jordan–Pauli commutators, 426–428
jump process, 728

Kirkwood distribution, 130
Klein–Gordon equation, 998–1002
Knuth lagged Fibonacci generator, 1184
Kolmogorov backward equation, 718, 769–770
Kolmogorov forward equation, 718
Kramers–Heisenberg formula, 596–600
Green tensor, 599–600
polarizability, 596–599
Kramers–Kronig relations, 289, 589–592, 648–649
Green tensor, 596
Kraus operator, 868

L'Ecuyer multiple recursive generator, 1182–1184
Lévy flight, 280
Lévy's arcsine law, 775–777
ladder operators, 302
ladder structure, 232
lagged Fibonacci generator, 1184
Lagrangian, 403–406
for string, 451–452
Laguerre–Gaussian functions, 121
Lamb dip, 221–228
Lamb shift, 42, 464, 465, 499, 508–509, 533, 536, 557–563, 663, 982
Lamb–Dicke effect, 507
lamp shift, 162
Lamperti transform, 747–748
temporal scaling, 748
Landau–Zener crossing, 162–166, 254–255, 298
Laplace equation, 422, 758–760
Laplace transform, 180–181, 656, 681, 772, 1047–1050
Laplace–Beltrami operator, 935, 937
Laplace–Beltrami ordering, 933
Laplacian
as spherical average, 759–760
spherical coordinates, 411
Larmor precession, 168
laser
Lorenz–Haken model, 261–266
laser medium
three-level, 104–105
Lebesgue measure, 870
level-shift operator, 659
Levi-Civita symbol, 301, 395, 582
light
thermal, 74–76
likelihood-ratio estimator
for derivative, 813
limit
from above, 1122
from below, 1122
of function, 1122
of sequence, 1121
Lindblad form, 185, 866
Lindblad superoperator, 185, 187
interference represented by, 292
line shape, 100
natural, 101

linear congruential generator, 1182
linewidth
    measurement of, 78–93
Liouville distribution, 127
Liouville–von Neumann equation, 108, 180
Liouvillian superoperator, 108, 180
little-*o* notation, 1121–1123
local interactions, 465–467
local realism, 131
local time, 793–807
long-wavelength approximation, 151, 460
longitudinal decay rate, 178
loop unrolling, 1114–1115
Lorentz force, 457–458
    Hamiltonian, 458
    Lagrangian, 458
Lorentz model, 33–36, 186–189
    damping coefficient, 38–39
Lorentzian
    absorption, 35
    line shape, 100, 101, 665
    lineshape, 68
    noise spectrum, 78, 80–85
Lorenz gauge, 401
Lorenz–Haken model, 261–266
low-pass filter, 720
Lévy area, 1178–1180

magic wavelength, 174, 252–253
magnetic dipole moment, 474
magnetic flux density, 435, 579
magnetic-dipole transition, 397
magnetizability, 1018
magnetization field
    atomic, 468–469, 473, 480
magneto-optical trap (MOT), 49
Malliavin derivative, 819
Markov approximation, 145, 521
Markovian
    evolution, 717
martingale, 715
master equation, 143–148, 185, 187, 508
    atom in thermal field, 508
    Born–Markov, 143–148, 508
    Lindblad form, 866
    stochastic, 833–863, 865–868, 875–904
    unconditioned, 833
matrix elements
    of **r** and **p**, 463–464
    of atom–field interaction Hamiltonian, 464
Matsubara frequencies, 628
maximum-value process, 776
Maxwell equations, 401–402, 579–580
measurable set, 870
measure, 869–871
    positive, 870
measurement record, 835
mechanical effects of light, 228–253
Mellin transform, 1047–1049
Mersenne Twister, 1184
methane-stabilized He–Ne laser, 225
method of images
    for diffusion, 805
Michelson interferometer, 66–67, 94
midpoint method, 1126
Milstein method, 1168–1169
minimal coupling, 458–460
    Hamiltonian, 460
    replacement, 460
mixed state, 107
Mollow triplet, 201–209, 255–256
    nonclassical correlations, 208–209
moment-generating function, 1048–1049
moments
    inverse, 1048–1049
momentum-shift operator, 232, 272
Moore machine, 1115
Moyal bracket, 127–128, 150
multiple recursive generator, 1182–1184
multipole expansion, 473–475, 583
    of localized current, 583–584
multipole fields, vector, 418–420
multistep methods, 1131

Naimark's theorem, 872
natural line width, 102
Neumark's theorem, 872
Newton's Second Law, 404
NIST-7 cesium-beam clock, 174, 255
NIST-F1 cesium fountain clock, 174
no-cloning theorem, 131–132
nonexponential decay, 503–506, 666–675, 679–680
    cavity, 682
normal modes
    of atom–cavity system, 489
normal-mode decomposition, 406, 972

octupole moment
    electric, 474
Ohm's law, 625
one-photon Rabi frequency, 486, 487, 496
open quantum systems, 142–148
open set, 870
optical Bloch equations, 177–186, 189, 191–194, 230, 275, 500
    Torrey's solutions, 181–185, 255

optical lattice
accelerating, 259
optical molasses, 46–49
optical theorem, 43–44
optics
geometric, 451
Ornstein–Uhlenbeck process, 718–720
correlation function, 718–719
low-pass filter and, 720
oscillator strength, 35–36
oscillatory sequence, 1121
out-of-order execution, 1114

Paley–Wiener theorem, 506
parabolic-cylinder functions, 165
parallel programming, 1118
parametric downconversion, 688
parity, 152–153
Parseval's theorem, 72
partition function, 1002–1003
path integral, 922–931, 936
path integral, 907–963, 965–989
Itō, 941–944, 946–950
ensemble-average normalization, 923–927, 960–962
midpoint, 938–941, 946–950
moment relations, 944–946
Monte–Carlo method, 922–933, 960–962
operator-ordering issues, 933–962
partition function, 922–931
phase-space, 910–911
prepoint, 941–944, 946–950
product ordering, 963
Stratonovich, 938–941, 946–950
Pauli operators, 166
Peres–Horodecki criterion, 132–140
perimittivity
imaginary frequencies, 591–592
permittivity, 581, 586–592
imaginary part as loss, 587–589
phase diffusion, 826–827
phase estimation, 861–863
phase lag
of circuit response, 650
phase modulation, 280, 296
phase noise, 76–95
phase space, 115
phase-space action, 404
photon, 156
photon blockade, 489–490
photon echo, 174–177
photon scattering rate, 42–45
photon-recoil momentum, 272
pipeline, cpu, 1113–1116
Planck blackbody distribution, 100
plasma frequency, 58
plasma model, 58
Poincaré cone condition, 758
Poisson bracket, 127, 128, 135–137, 934
Poisson distribution, 726–728, 826
Poisson equation, 421, 760–761
Poisson process, 726–731, 826, 839, 843
inhomogeneous, 728
intensity, 726
white-noise limit, 728–729
Poisson sum rule, 931
polar Marsaglia method, 1186
polarizability, 32–33, 255
diagonal in principle coordinates, 595
related to dipole correlation, 594, 595
scalar, 595
static, 542
polarization field, 33, 579, 581
atomic, 461, 467–468, 470, 484
coupling to electric field, 60–61
effective sources, 579–580
polarization sum, 425–426
polarizer, 507–508
pole approximation, 660, 662, 663
position measurement
electromagnetically induced transparency, 899–904
probe excitation, 895–899
resonance fluorescence, 885–895
positive map, 865–866
positive measure, 870
positive partial transpose (PPT), 132–140
positive semidefinite, 138
positive semidefinite operator, 133
positive-operator-valued measure, 868–875
imprecise measurement, 873–875
spontaneous emission, 872–873
POVM, 868–875
imprecise measurement, 873–875
spontaneous emission, 872–873
power broadening, 197
power spectral density, 65
one- vs. two-sided, 68–69
Power–Zienau transformation, 460–465, 467, 469–484, 544–548
power-equivalent width, 71
predictor–corrector methods, 1131
probability density
for combinations of random variables, 709–710
probability measure, 871
probe absorption
driven, two-level atom, 211–217, 259–260
problem:magdipdecay, 509

procedure inlining, 1117
product ordering, 959–963
profiler, 1118
projection theorem, 336–337
projection-valued measure, 869
propagator, 655, 907–962
  classical limit, 913–922
  free particle, 911–913
  imaginary-time, 922–933
  semiclassical, 913–922
Purcell factor, 682
pure state, 107
purification, 142–143
purity, 149–150

quadratures, 857–863
quadrupole moment, 484
  electric, 474
quantum beats, 291–295
  steady-state, 293–297
quantum jumps, 833–840
quantum Langevin equation, 521
quantum measurement, 868
  continuous, of position, 875–904
  strong, 874
  weak, 874
quantum regression theorem, 199–201, 209–210, 213–214, 218, 241–243, 256, 260–261
quantum Zeno effect, 505, 507–508, 678
qubit, 107
quotient
  of two normal deviates, 710

Röntgen interaction
  center-of-mass, 475–482
Rabi flopping, 157–162, 487, 657–658
  damped, 180–185
  resolvent approach, 657–658
Rabi frequency, 153–154, 185–186, 229, 230, 270, 487
  dependence on photon number, 487
  for quadrupole interaction, 484
  generalized, 230
  Raman, 233, 272
radiation
  atom near macroscopic body, 635–638
  atom near mirror, 50–53
  two atoms, 53–56
radiation pressure, 45–49, 236
radiation reaction, 38–39
  Abraham–Lorentz force, 39
radiation, dipole, 36–39, 194–195
  angular distribution, 38
Radon–Nikodym derivative, 950
Raman scattering
  spontaneous, 675–683
Ramsey fringes, 171–177, 255
random number generation, 1181–1190
  angular distributions, 1186–1188
  exponential distribution, 1188
  gamma distribution, 1189–1190
  power-law distribution, 1188–1189
random process
  complex Gaussian, 74–76
random walk, 48, 697–706
random-number generation
  Gaussian deviates, 1185–1186
  normal deviates, 1185–1186
  shuffling, 1184–1185
  uniform deviates, 1181–1185
rate equations
  Einstein's, 98–105, 178, 191–194
Rayleigh's formula, 418, 419
recoil energy, 233
  Raman, 273
recoil frequency, 234, 235
recoil temperature, 49
recoil velocity, 45
reduced matrix element, 334
reflectance, 441
reflected Brownian motion, 776–777, 802
reflection coefficient, 58, 602–603
  Fresnel, 564, 603
Reflection Principle, 749–751
registers, cpu, 1110–1111
regularization
  dimensional, 1028
reparameterization independence, 992
reservoir, 143, 144
residue, 589
resolved-sideband Raman cooling, 278–281
resolvent
  of the Liouvillian, 181
resolvent operator, 653–681
resonance fluorescence, 194–211, 255–256
  as position measurement, 885–895
  coherent and incoherent, 197–199
retarded Green function, 654, 655
Richardson extrapolation, 1151–1154, 1156–1158
Riemann sheet, 666
Rihaczek distribution, 130
rotating frame, 154–157, 254, 270–271, 283
rotating-wave approximation, 34, 153, 155, 157, 515, 534
round-trip time, 511
rubidium 87
  transition wavelengths and lifetimes, 553–554

Runge–Kutta methods, 1125–1131
adaptive stepping, 1130
implicit, 1129–1130
second-order, 1126
stochastic, 1172–1176
Runge–Kutta–Fehlberg method, 1130

sample mean, 703
sampling theorem, 1133–1136
saturated-absorption spectroscopy, 221–228
crossover resonance, 225–228
saturation, 102–103
saturation intensity, 103, 105, 192–194
for two-level atom, 41
saturation parameter, 179
saturation spectroscopy, 221–228
crossover resonance, 225–228
scattering cross section, 195–197
scattering rate, photon, 42–45
Schrödinger equation
with curvature, 935
Schrödinger picture, 112–114
Schrödinger–von Neumann equation, 108, 177
Schrödinger-cat state, 133
Schwinger proper time, 975
second
definition of, 174
second quantization, 409, 972–973
second variation, 914
selection rules, 342–344
self-heterodyne spectroscopy, 81–93
separability, 131
shot noise, 729–731
sine integral Si(x), 537, 574–575
single-precision, IEEE, 1109, 1110
singular D function, 427
skew Brownian motion, 802–807
slow light, 290
slowly varying coherences, 167
Smoluchowski equation, 735
Snell's Law, 565
sojourn time, 770–807
spectral density
of frequency fluctuations, 76
of frequency fluctuations(, 94
of frequency fluctuations), 95
of phase fluctuations, 76
of phase fluctuations(, 94
of phase fluctuations), 95
spherical Bessel functions, 413
spherical cavity modes
normalization, 454–456
scalar, 411–414
vector, 414–418
spherical Hankel functions, 419
spherical harmonics, 412–413
spin echo, 174–177
split-operator methods, 1149–1165
spontaneous decay rate, 186, 498–499
between planar mirrors, 509
near planar mirror, 507, 637–638
with angular-momentum degeneracy, 344–346
spontaneous emission, 98, 146, 246–247, 495–506, 508–509, 662–675
enhanced by cavity, 682
Fermi's golden rule, 502–503
in stimulated Raman scattering, 273
in thermal field, 508–509
into 1D waveguide, 509
into 2D waveguide, 509
line shape, 663–665
magnetic-dipole, 509
master equation, 499–500
near macroscopic body, 635–638
nonexponential decay, 503–506, 666–675
resolvent operator, 662–675
Weisskopf–Wigner approach, 495–499, 507, 509
spontaneous Raman scattering, 675–683
line shape, 682–683
square-well potential, 930
squeezed state, 454
squeezing, 858
Stückelberg angle, 161, 489
standard deviation, 697, 701
of mean, 703
standard topology, 870
standard-ordered distribution, 130
Stark shift
ac, 162
due to thermal field, 508
stationary noise process, 88
stationary-phase approximation, 913
steady-state quantum beats, 293–297
stiff equation, 1124
stimulated emission, 98
stimulated Raman adiabatic passage, 290–291, 298
stimulated Raman cooling, 278–281
stimulated Raman transition, 232, 269–283, 298
velocity selection, 274–280
STIRAP, 290–291, 298
stochastic derivative, 819
stochastic differential equation, 712–726, 1167–1181
implicit methods for, 1174–1176
stochastic force
on damped particle, 731–737
stochastic master equation, 833–868, 875–904

diffusion form, 840–851
jump form, 833–840, 864
linear, 881–885
multiple observers, 853–855, 878–879
position measurement, 865–904
stochastic Schrödinger equation, 835–836, 847, 864, 875–877
stochastic Taylor expansion, 1169–1172
stochastic-dipole force, 238–246
dressed-state model, 244–245
Stokes relations, 512
Stratonovich calculus, 720–726, 948–950, 954–957
validity of chain rule in, 724–725
strong convergence
SDE methods, 1176–1177
strong coupling, 491
strong measurement, 874
Struve function, 276
subradiance, 56, 298–299
superradiance, 52, 56, 298–299
survival probability, 517
susceptibility, 33, 435, 581, 1009
generalized, 592–596
magnetic, 1009
symmetrized ordering, 957–959, 959, 963
symplectic integration, 1161–1165
symplectic matrix, 136–137, 149

Tanaka formula, 802
Taylor expansion
stochastic, 1169–1172
tensor product, 130–131
thermal field, 508
thermal light, 74–76
thermal state
Wigner distribution, 149
Thomas–Reiche–Kuhn sum rule, 36, 539–540, 543, 549, 550
time slicing, 911
topology, 870
trace, 109
transmittance, 441
transverse decay rate, 178
truncation error, 1122
tunneling, 254
two-level atom, 151–161
two-mode squeezed state, 140
two-photon absorption, 297–298
two-photon interference, 688–694

uncertainty relation, 70, 72
generalized, 135–137
unit-stride access, 1111
unitary time-evolution operator, 111–112, 907
unitary transformation, 155–156

vacuum Rabi doublet, 489
van der Waals–London potential, 616–621
Van Vleck–Morette determinant, 917–921
harmonic oscillator, 921–922
variation process, 820
variational calculus, 402–406
vector spherical harmonics, 415
Verlet method, 1163–1165
visibility, 69–70, 94
two-detector, 94
Voigt profile, 100
VSCPT, 286–287

wave equation
for string, 451–452
weak convergence
SDE methods, 1176–1177
weak measurement, 874
Weber's equation, 164
Weisskopf–Wigner approximation, 495–499, 517–518
Weyl correspondence, 122–128, 149
Weyl ordering, 125–126, 933, 954–957, 963
Weyl's quantization rule, 123
white noise, 712
Wick rotation, 922, 970
Wiener differential, 712
Wiener increment, 712
Wiener path
reflected, 776–777, 802
Wiener process, 710–712, 737–811, 925
boundary crossing, 749–754, 758–761
correlation function, 715
escape, 754–758
local time, 793–807
recurrence, 807–811
sojourn time, 770–807
Wiener–Khinchin theorem, 64–69, 715
optical, 65–69, 194
Wigner 3-j symbol, 309
Wigner distribution, 115–129
area theorem, 117–118, 134
discrete Wigner transform, 1145–1147
Gaussian, 118–119
harmonic oscillator eigentstate, 121–122
superposition state, 119–121
thermal state, 149
upper bound, 149
Wigner–Eckart theorem, 333–337, 341–342
world line, 965
world lines, 980