Lecture Notes on Condensed Matter Physics (A Work in Progress)

Daniel Arovas
Department of Physics
University of California, San Diego

July 14, 2025

Contents

Contents i

List of Figures xvii

List of Tables xxvi

1 Broken Symmetry 1

1.1 Introduction . . . 1

1.1.1 The mother of all Hamiltonians . . . 1

1.2 Classical and Quantum Statistical Physics . . . 4

1.2.1 Correspondence between quantum and classical statistical mechanics . . . 6

1.3 Phases of matter . . . 9

1.3.1 Spontaneous symmetry breaking . . . 9

1.3.2 Beyond the Landau paradigm . . . 12

1.4 Landau Theory of Phase Transitions . . . 14

1.4.1 Quartic free energy with Ising symmetry . . . 14

1.4.2 Cubic terms in Landau theory : first order transitions . . . 16

1.4.3 Magnetization dynamics . . . 18

1.4.4 Sixth order Landau theory : tricritical point . . . 20

1.4.5 Hysteresis for the sextic potential . . . 22

1.4.6 Weak crystallization . . . 25

1.5 Four Vignettes . . . 26
1.5.1 Lower critical dimension . . . 26
1.5.2 Random systems : Imry-Ma argument . . . 29
1.5.3 Hohenberg-Mermin-Wagner theorem . . . 31
1.5.4 Goldstone's theorem . . . 37
1.6 Appendix : The Foldy-Wouthuysen Transformation . . . 41
1.6.1 The Dirac Hamiltonian . . . 41
1.6.2 Emergence of the spin-orbit and Zeeman interaction terms . . . 42
1.7 Appendix : Ideal Bose Gas Condensation . . . 44
1.8 Appendix : Asymptotic Series in a Zero-Dimensional Field theory . . . 47
1.9 Appendix : Derivation of Ginzburg-Landau Functional . . . 55
1.9.1 Discrete symmetry : $\mathbb{Z}_2$ . . . 55
1.9.2 Continuous symmetry : O(n) . . . 58

2 Crystal Math **61**
2.1 Classification of Crystalline Structures . . . 61
2.1.1 Bravais Lattices . . . 61
2.1.2 Miller indices . . . 64
2.1.3 Crystallographic restriction theorem . . . 66
2.1.4 Enumeration of two and three-dimensional Bravais lattices . . . 66
2.1.5 Crystals . . . 69
2.1.6 Trigonal crystal system . . . 73
2.1.7 Point groups, space groups and site groups . . . 74
2.2 More on Point Groups . . . 77
2.2.1 Standard notation for point group operations . . . 77
2.2.2 Proper point groups . . . 77
2.2.3 Commuting operations . . . 79
2.2.4 Improper point groups . . . 80

2.2.5 The ten two-dimensional point groups . . . 83
2.2.6 The achiral tetrahedral group, T_d . . . 83
2.2.7 Tetrahedral *vs.* octahedral symmetry . . . 86
2.2.8 The 32 crystallographic point groups . . . 87
2.2.9 Hermann-Mauguin (international) notation . . . 88
2.2.10 Double groups . . . 92
2.2.11 The three amigos : D_4 , C_{4v} , D_{2d} . . . 96
2.3 Space Groups . . . 101
2.3.1 Space group elements and their properties . . . 101
2.3.2 Factor groups . . . 105
2.3.3 How to make a symmorphic space group . . . 107
2.3.4 Nonsymmorphic space groups . . . 108
2.3.5 Translations and their representations . . . 109
2.3.6 Space group representations . . . 111
2.4 Fourier Space Crystallography . . . 113
2.4.1 Space group symmetries . . . 114
2.4.2 Extinctions . . . 115
2.4.3 Sticky bands . . . 117

3 Deformations of Crystals **123**
3.1 Elasticity . . . 123
3.1.1 Stress and strain tensors . . . 123
3.1.2 Elasticity and symmetry . . . 125
3.2 Phonons in Crystals . . . 128
3.2.1 One-dimensional chain . . . 128
3.2.2 General theory of lattice vibrations . . . 130
3.2.3 Translation and rotation invariance . . . 134
3.2.4 Phonons in an fcc lattice . . . 134

3.2.5 Phonons in the hcp structure . . . 135
3.2.6 Phonon density of states . . . 138
3.2.7 Einstein and Debye models . . . 141
3.2.8 Phenomenological theory of melting . . . 142
3.2.9 Goldstone bosons . . . 146
3.2.10 Elasticity theory redux : Bravais lattices . . . 148
3.2.11 Elasticity theory in cases with bases . . . 151
3.3 Neutron diffraction . . . 152
3.3.1 Inelastic differential scattering cross-section . . . 152
3.3.2 Evaluation of $S_{ll'}(\boldsymbol{q},\omega)$ . . . 154
3.3.3 Dynamic structure factor for Bravais lattices . . . 155
3.3.4 Debye-Waller Factor . . . 156
3.3.5 The Mössbauer effect . . . 157

4 Electronic Band Structure of Crystals 159
4.1 Energy Bands in Solids . . . 159
4.1.1 Bloch's theorem . . . 159
4.1.2 Schrödinger equation . . . 161
4.1.3 $V = 0$: empty lattice . . . 162
4.1.4 Perturbation theory . . . 162
4.1.5 Solvable model : one-dimensional Dirac comb . . . 165
4.1.6 Diamond lattice bands . . . 167
4.2 Metals and Insulators . . . 169
4.2.1 Density of states . . . 169
4.2.2 Fermi statistics . . . 173
4.2.3 Metals and insulators at $T = 0$ . . . 174
4.3 Tight Binding Model . . . 175
4.3.1 Bands from atomic orbitals . . . 175

4.3.2 Wannier functions . . . 178
4.3.3 Tight binding redux . . . 180
4.3.4 Interlude on Fourier transforms . . . 181
4.3.5 Examples of tight binding dispersions . . . 183
4.3.6 Bloch's theorem, again . . . 186
4.3.7 Go flux yourself : how to add magnetic fields . . . 187
4.3.8 General flux configuration on the square lattice . . . 193
4.4 Topological Band Structures . . . 194
4.4.1 SSH model . . . 194
4.4.2 Polarization and geometric phase . . . 199
4.4.3 Domain wall states in the Dirac equation . . . 202
4.4.4 The adiabatic theorem and Berry's phase . . . 206
4.4.5 Connection, curvature, and Chern numbers . . . 208
4.4.6 Two-band models . . . 211
4.4.7 Quantum Metric and Geometric Curvature Tensors . . . 215
4.4.8 The TKNN formula . . . 217
4.5 Semiclassical Dynamics of Bloch Electrons . . . 221
4.5.1 Adiabatic evolution . . . 222
4.5.2 Violation of Liouville's theorem and its resolution . . . 224
4.5.3 Bloch oscillations . . . 226
4.6 *Ab initio* Calculations of Electronic Structure . . . 226
4.6.1 Orthogonalized plane waves . . . 226
4.6.2 The pseudopotential . . . 229
4.7 Appendix I : Gauss-Bonnet and Pontrjagin . . . 231
4.7.1 Gauss-Bonnet theorem . . . 231
4.7.2 The Pontrjagin index . . . 233
4.8 Appendix II : Derivation of Eqn. 4.268 . . . 234

5 Metals **237**

5.1 Introduction . . . 237

5.2 $T = 0$ and the Fermi Surface . . . 237

5.2.1 Definition of the Fermi surface . . . 237

5.2.2 Fermi surface vs. Brillouin zone . . . 240

5.2.3 Spin-split Fermi surfaces . . . 242

5.3 Quantum Thermodynamics of the Electron Gas . . . 244

5.3.1 Fermi distribution . . . 244

5.3.2 Sommerfeld expansion . . . 244

5.3.3 Chemical potential shift . . . 247

5.3.4 Specific heat . . . 248

5.4 Effects of External Magnetic Fields . . . 248

5.4.1 Magnetic susceptibility and Pauli paramagnetism . . . 248

5.4.2 Landau diamagnetism . . . 251

5.4.3 de Haas-van Alphen oscillations . . . 254

5.4.4 de Haas-von Alphen effect for anisotropic Fermi surfaces . . . 258

5.5 Simple Theory of Electron Transport in Metals . . . 261

5.5.1 Drude model . . . 261

5.5.2 Magnetoresistance and magnetoconductance . . . 263

5.5.3 Hall effect in high fields . . . 265

5.5.4 Cyclotron resonance in semiconductors . . . 265

5.5.5 Magnetoresistance in a two band model . . . 266

5.5.6 Optical reflectivity of metals and semiconductors . . . 268

5.5.7 Theory for Bloch wavepackets . . . 271

5.6 Boltzmann Equation in Solids . . . 272

5.6.1 Semiclassical dynamics and distribution functions . . . 272

5.6.2 Local equilibrium . . . 277

5.7 Conductivity of Normal Metals . . . 279
5.7.1 Relaxation time approximation . . . 279
5.7.2 Optical conductivity and the Fermi surface . . . 282
5.8 Calculation of the Scattering Time . . . 283
5.8.1 Potential scattering and Fermi's golden rule . . . 283
5.8.2 Screening and the transport lifetime . . . 286
5.9 Dynamics of Holes . . . 289
5.9.1 Properties of holes . . . 289
5.9.2 Boltzmann equation for holes . . . 291
5.10 Magnetoresistance and Hall Effect . . . 292
5.10.1 Boltzmann theory for $\rho_{\alpha\beta}(\omega, \boldsymbol{B})$ . . . 292
5.10.2 Hall effect in high fields . . . 295
5.11 Thermal Transport . . . 296
5.11.1 Boltzmann theory . . . 296
5.11.2 The heat equation . . . 299
5.11.3 Calculation of transport coefficients . . . 300
5.11.4 Onsager relations . . . 302
5.12 Electron-Phonon Scattering . . . 304
5.12.1 Introductory remarks . . . 304
5.12.2 Electron-phonon interaction . . . 304
5.12.3 Boltzmann equation for electron-phonon scattering . . . 307

6 Semiconductors and Insulators **311**
6.1 Introduction . . . 311
6.1.1 Band gaps and transport . . . 311
6.1.2 Hall effect . . . 313
6.1.3 Optical absorption . . . 315
6.1.4 Direct *versus* indirect gaps . . . 316

6.1.5 Mobility . . . 317
6.1.6 Effective mass . . . 318
6.2 Number of Carriers in Thermal Equilibrium . . . 318
6.2.1 Intrinsic semiconductors . . . 320
6.2.2 Extrinsic semiconductors . . . 320
6.3 Donors and Acceptors . . . 323
6.3.1 Impurity charges in a semiconductor . . . 323
6.3.2 Donor and acceptor quantum statistics . . . 324
6.3.3 Chemical potential *versus* temperature in doped semiconductors . . . 326
6.4 Inhomogeneous Semiconductors . . . 327
6.4.1 Modeling the p-n junction . . . 328
6.4.2 Rectification . . . 331
6.4.3 MOSFETs and heterojunctions . . . 334
6.4.4 Heterojunctions . . . 336
6.5 Insulators . . . 338
6.5.1 Maxwell's equations in polarizable media . . . 339
6.5.2 Clausius-Mossotti relation . . . 341
6.5.3 Theory of atomic polarizability . . . 342
6.5.4 Electromagnetic waves in a polar crystal . . . 345

7 Mesoscopia **347**
7.1 Introduction . . . 347
7.2 The Landauer Formula . . . 347
7.2.1 Example: potential step . . . 350
7.3 Multichannel Systems . . . 352
7.3.1 Transfer matrices: the Pichard formula . . . 356
7.3.2 Discussion of the Pichard formula . . . 358
7.3.3 Two quantum resistors in series . . . 360

7.3.4 Two quantum resistors in parallel . . . 363
7.4 Universal Conductance Fluctuations in Dirty Metals . . . 371
7.4.1 Weak localization . . . 375
7.5 Anderson Localization . . . 377
7.5.1 Characterization of localized and extended states . . . 379
7.5.2 Numerical studies of the localization transition . . . 380
7.5.3 Scaling theory of localization . . . 381
7.5.4 Finite temperature . . . 386

8 Hartree-Fock and Density Functional Theories 389
8.1 Second Quantization . . . 389
8.1.1 Basis states and creation/annihilation operators . . . 389
8.1.2 The second quantized Hamiltonian . . . 393
8.2 Hartree-Fock Theory . . . 395
8.3 Density Functional Theory . . . 402
8.3.1 Hohenberg-Kohn theorems . . . 402
8.3.2 Kohn-Sham equations . . . 404
8.4 Response Functions . . . 408
8.4.1 Linear response theory . . . 408
8.4.2 Static screening . . . 410
8.4.3 Approximate forms for the polarization function . . . 412

9 Landau Fermi Liquid Theory 415
9.1 Normal ^{3}He Liquid . . . 415
9.2 Fermi Liquid Theory : Statics and Thermodynamics . . . 417
9.2.1 Adiabatic continuity . . . 417
9.2.2 First law of thermodynamics for Fermi liquids . . . 421
9.2.3 Low temperature equilibrium properties . . . 424

9.2.4 Thermodynamic stability at $T = 0$ . . . 430
9.3 Collective Dynamics of the Fermi Surface . . . 432
9.3.1 Landau-Boltzmann equation . . . 432
9.3.2 Zero sound : free FS oscillations in the collisionless limit . . . 434
9.4 Dynamic Response of the Fermi Liquid . . . 436

10 Linear Response of Quantum Systems **439**
10.1 Response and Resonance . . . 439
10.1.1 Forced damped oscillator . . . 439
10.1.2 Energy dissipation . . . 441
10.1.3 Kramers-Kronig relations . . . 441
10.2 Quantum Mechanical Response Functions . . . 443
10.2.1 First order perturbation theory . . . 443
10.2.2 Spectral representation . . . 446
10.2.3 Energy dissipation . . . 448
10.2.4 Correlation functions . . . 450
10.2.5 Continuous systems . . . 451
10.3 A Spin in a Magnetic Field . . . 451
10.3.1 Bloch Equations . . . 452
10.4 Density Response and Correlations . . . 454
10.4.1 Sum rules . . . 455
10.4.2 Dynamic Structure Factor for the Electron Gas . . . 457
10.5 Charged Systems: Screening and Dielectric Response . . . 463
10.5.1 Definition of the charge response functions . . . 463
10.5.2 Static screening: Thomas-Fermi approximation . . . 465
10.5.3 High frequency behavior of $\epsilon(\boldsymbol{q}, \omega)$ . . . 466
10.5.4 Random phase approximation (RPA) . . . 467
10.5.5 Plasmons . . . 469

10.5.6 Jellium . . . 470
10.6 Electromagnetic Response . . . 472
10.6.1 Gauge invariance and charge conservation . . . 474
10.6.2 A sum rule . . . 475
10.6.3 Longitudinal and transverse response . . . 475
10.6.4 Neutral systems . . . 476
10.6.5 The Meissner effect and superfluid density . . . 477

11 Phenomenological Theories of Superconductivity **481**
11.1 Basic Phenomenology of Superconductors . . . 481
11.2 Thermodynamics of Superconductors . . . 484
11.3 London Theory . . . 488
11.3.1 De Gennes' derivation of London Theory . . . 490
11.3.2 Pippard's nonlocal extension . . . 490
11.4 Ginzburg-Landau Theory . . . 491
11.4.1 Landau theory for superconductors . . . 492
11.4.2 Ginzburg-Landau free energy . . . 493
11.4.3 Equations of motion . . . 494
11.4.4 Critical current . . . 495
11.4.5 Ginzburg criterion . . . 497
11.4.6 Domain wall solution . . . 499
11.4.7 Scaled Ginzburg-Landau equations . . . 501
11.5 Applications of Ginzburg-Landau Theory . . . 501
11.5.1 Domain wall energy . . . 502
11.5.2 Thin type-I films : critical field strength . . . 504
11.5.3 Critical current of a wire . . . 507
11.5.4 Magnetic properties of type-II superconductors . . . 509
11.5.5 Lower critical field of a type-II superconductor . . . 511

11.5.6 Abrikosov vortex lattice 512

12 BCS Theory of Superconductivity **519**

12.1 Binding and Dimensionality 519

12.2 Cooper's Problem 521

12.3 Effective Attraction Due to Phonons 525

12.3.1 Electron-phonon Hamiltonian 525

12.3.2 Effective interaction between electrons 526

12.4 Reduced BCS Hamiltonian 527

12.5 Solution of the Mean Field Hamiltonian 529

12.6 Self-Consistency 531

12.6.1 Solution at zero temperature 532

12.6.2 Condensation energy 533

12.7 Coherence factors and quasiparticle energies 534

12.8 Number and Phase 535

12.9 Finite Temperature 536

12.9.1 Isotope effect 537

12.9.2 Landau free energy of a superconductor 538

12.10 Paramagnetic Susceptibility 541

12.11 Finite Momentum Condensate 542

12.11.1 Gap equation for finite momentum condensate 544

12.11.2 Supercurrent 544

12.12 Effect of Repulsive Interactions 546

12.13 Appendix I : General Variational Formulation 548

12.14 Appendix II : Superconducting Free Energy 549

13 Applications of BCS Theory **553**

13.1 Quantum XY Model for Granular Superconductors 553

13.1.1 No disorder . . . 554
13.1.2 Self-consistent harmonic approximation . . . 555
13.1.3 Calculation of the Cooper pair hopping amplitude . . . 557
13.2 Tunneling . . . 559
13.2.1 Perturbation theory . . . 560
13.2.2 The single particle tunneling current I_N . . . 561
13.2.3 The Josephson pair tunneling current I_J . . . 570
13.3 The Josephson Effect . . . 573
13.3.1 Two grain junction . . . 573
13.3.2 Effect of in-plane magnetic field . . . 574
13.3.3 Two-point quantum interferometer . . . 575
13.3.4 RCSJ Model . . . 576
13.4 Ultrasonic Attenuation . . . 588
13.5 Nuclear Magnetic Relaxation . . . 590
13.6 General Theory of BCS Linear Response . . . 592
13.6.1 Case I and case II probes . . . 596
13.6.2 Electromagnetic absorption . . . 597
13.7 Electromagnetic Response of Superconductors . . . 600
13.7.1 Computation of the electromagnetic response kernel . . . 601
13.7.2 Finite q response and the true penetration depth . . . 604

14 Magnetism **607**
14.1 Introduction . . . 607
14.1.1 Absence of orbital magnetism within classical physics . . . 609
14.2 Basic Atomic Physics . . . 609
14.2.1 Single electron Hamiltonian . . . 609
14.2.2 The Darwin term . . . 610
14.2.3 Many electron Hamiltonian . . . 610

14.3 The Periodic Table . . . 612
14.3.1 Aufbau principle . . . 613
14.3.2 Splitting of configurations: Hund's rules . . . 614
14.3.3 Spin-orbit interaction . . . 616
14.3.4 Crystal field splittings . . . 618
14.4 Magnetic Susceptibility of Atomic and Ionic Systems . . . 620
14.4.1 Filled shells: Larmor diamagnetism . . . 620
14.4.2 Partially filled shells: van Vleck paramagnetism . . . 621
14.5 Moment Formation in Interacting Itinerant Systems . . . 625
14.5.1 The Hubbard model . . . 625
14.5.2 Stoner mean field theory . . . 626
14.5.3 Antiferromagnetic solution . . . 630
14.5.4 Mean field phase diagram of the Hubbard model . . . 631
14.6 Interaction of Local Moments: the Heisenberg Model . . . 633
14.6.1 Ferromagnetic exchange of orthogonal orbitals . . . 634
14.6.2 Heitler-London theory of the H_2 molecule . . . 636
14.6.3 Failure of Heitler-London theory . . . 638
14.6.4 Herring's approach . . . 638
14.7 Mean Field Theory . . . 639
14.7.1 Ferromagnets . . . 642
14.7.2 Antiferromagnets . . . 642
14.7.3 Susceptibility . . . 643
14.7.4 Variational probability distribution . . . 645
14.8 Magnetic Ordering . . . 647
14.8.1 Mean field theory of anisotropic magnetic systems . . . 650
14.8.2 Quasi-1D chains . . . 650
14.9 Spin Wave Theory . . . 652

14.9.1 Ferromagnetic spin waves . . . 653
14.9.2 Static correlations in the ferromagnet . . . 654
14.9.3 Antiferromagnetic spin waves . . . 655
14.9.4 Specific heat due to spin waves . . . 660
14.10 Appendix : Generalized Spin Wave Theory for Isotropic Systems . . . 661
14.10.1 General form of Heisenberg Hamiltonian . . . 661
14.10.2 Planar spiral phases . . . 663
14.10.3 Sublattices . . . 664
14.10.4 Diagonalization . . . 666
14.11 Appendix: The Foldy-Wouthuysen Transformation . . . 668
14.11.1 Derivation of the spin-orbit interaction . . . 669

15 Spins, Coherent States, Path Integrals, and Applications 673
15.1 The Coherent State Path Integral . . . 673
15.1.1 Feynman path integral . . . 673
15.1.2 Coherent state path integral for the 'Heisenberg-Weyl' hroup . . . 675
15.2 Coherent States for Spin . . . 679
15.2.1 Coherent state wavefunctions . . . 683
15.2.2 Valence bond states . . . 684
15.2.3 Derivation of spin path integral . . . 686
15.2.4 Gauge field and geometric phase . . . 688
15.2.5 Semiclassical dynamics . . . 688
15.3 Other Useful Representations of the Spin Path Integral . . . 690
15.3.1 Stereographic representation . . . 690
15.3.2 Recovery of spin wave theory . . . 691
15.4 Quantum Tunneling of Spin . . . 691
15.4.1 Model Hamiltonian . . . 691
15.4.2 Instantons and tunnel splittings . . . 692

15.4.3 Garg's calculation (1993) . . . 693
15.5 Haldane's Mapping to the Nonlinear Sigma Model . . . 695
15.5.1 Hamiltonian . . . 696
15.5.2 Geometric phase . . . 697
15.5.3 Emergence of the nonlinear sigma model . . . 698
15.5.4 Continuum limit of the geometric phase: $d = 1$ . . . 699
15.5.5 The geometric phase in higher dimensions . . . 701
15.6 Large-N Techniques . . . 701
15.6.1 $1/N$ expansion for an integral . . . 702
15.6.2 Large-N theory of the nonlinear sigma model . . . 702
15.6.3 Correlation functions . . . 709

16 Notes on Line Graphs **711**
16.1 Line graphs: Kagomé and Checkerboard Lattices . . . 711
16.1.1 Kagomé lattice . . . 711
16.1.2 Checkerboard lattice . . . 712
16.1.3 Square-octagon lattice line graph . . . 713
16.2 Pyrochlore Lattice . . . 714
16.2.1 Adjacency Matrix . . . 717
16.2.2 The FCC lattice Brillouin zone . . . 720
16.3 Depleted Pyrochlores . . . 722
16.3.1 Pyrochlore with 16 element basis . . . 722
16.3.2 Depleted pyrochlore . . . 725
16.4 References . . . 728

17 Quadratic Hamiltonians **729**
17.1 Bosonic Models . . . 729
17.1.1 Bogoliubov equations . . . 730

17.1.2 Ground state 731
17.1.3 A final note on the boson problem 732
17.2 Fermionic Models 733
17.2.1 Ground state 735
17.3 Majorana Fermion Models 735
17.3.1 Majorana chain 736
17.4 Jordan-Wigner Transformation 739
17.4.1 Anisotropic XY model 740
17.4.2 Majorana representation of the JW transformation 743

List of Figures

1.1 What is the world made of? . . . 2
1.2 What is the world made of? Another point of view. . . . 3
1.3 Space and time in quantum critical phenomena. . . . 8
1.4 Phase diagrams for H_2O and the quantum Hall effect. . . . 10
1.5 Phase diagram for the quartic Landau free energy. . . . 14
1.6 Behavior of the quartic free energy $f(m) = \frac{1}{2}am^2 - \frac{1}{3}ym^3 + \frac{1}{4}bm^4$. . . . 17
1.7 Fixed points for $\varphi(u) = \frac{1}{2}ru^2 - \frac{1}{3}u^3 + \frac{1}{4}u^4$ and flow $\dot{u} = -\varphi'(u)$. . . . 19
1.8 Behavior of the sextic free energy $f(m) = \frac{1}{2}am^2 + \frac{1}{4}bm^4 + \frac{1}{6}cm^6$. . . . 21
1.9 High temperature cubic perovskite crystal structure of $BaTiO_3$. . . . 22
1.10 Sextic free energy $\varphi(u) = \frac{1}{2}ru^2 - \frac{1}{4}u^4 + \frac{1}{6}u^6$ for different values of r. . . . 23
1.11 Fixed points $\varphi'(u^*) = 0$ and flow $\dot{u} = -\varphi'(u)$ for the sextic potential. . . . 24
1.12 Domain walls in the two and three dimensional Ising model. . . . 27
1.13 A domain wall in an XY ferromagnet. . . . 29
1.14 Imry-Ma domains and free energy *versus* domain size. . . . 30
1.15 The double well and Mexican hat potentials. . . . 40
1.16 Relative error *versus* number of terms kept for the quartic theory. . . . 50
1.17 Cluster symmetry factors for the quartic theory. . . . 51
1.18 Logarithm of ratio of remainder after N terms for the sextic theory. . . . 53
1.19 Diagrams and their symmetry factors for the sextic theory. . . . 54

2.1 Brillouin zones . . . 63

2.2 Simple cubic (sc), body-centered cubic (bcc), and face-centered cubic (fcc) lattices 63
2.3 Examples of Miller planes . . . 64
2.4 Fivefold symmetry . . . 67
2.5 The five two-dimensional Bravais lattices . . . 67
2.6 Cubic, trigonal, and hexagonal Bravais lattices . . . 69
2.7 Orthorhombic Bravais lattices . . . 69
2.8 Tetragonal, monoclinic, and triclinic Bravais lattices . . . 70
2.9 Tetrahedral, cubic, and icosahedral group symmetry operations . . . 71
2.10 Unit cells of four high temperature cuprate superconductors . . . 72
2.11 Two nonsymmorphic crystal lattices . . . 76
2.12 The symmetry operations of T and O . . . 79
2.13 Stereograms of point groups (I) . . . 80
2.14 Stereograms of point groups (II) . . . 81
2.15 Stereograms of point groups (III) . . . 83
2.16 The zincblende structure and the group T_d . . . 84
2.17 Subgroup relations among the 32 crystallographic point groups . . . 88
2.18 Schematic diagram of (common axis) double group rotation generators . . . 93
2.19 Structure of hexagonal H_2O ice . . . 102
2.20 Unit cells for the 17 two-dimensional space groups . . . 104
2.21 Some common AB crystal structures and their space groups . . . 108
2.22 Stickiness of tight binding energy bands in an hcp crystal . . . 118
2.23 Examples of one-dimensional symmorphic and nonsymmorphic lattices . . . 120
2.24 Examples of space groups and their nonsymmorphic ranks . . . 121

3.1 A linear chain of masses and springs . . . 130
3.2 Phonons in elemental rhodium and in gallium arsenide . . . 132
3.3 Classical lattice energy for hcp ^{4}He . . . 137
3.4 Phonon dispersions for hcp ^{4}He . . . 138

3.5 Phonon spectra in elemental rhodium and in gallium arsenide 140

4.1 Band structure for an empty one-dimensional lattice 163
4.2 Energy bands for the Dirac comb potential 166
4.3 The zincblende structure and its Brillouin zone 167
4.4 Diamond lattice band structures 168
4.5 Square and cubic lattice densities of states 171
4.6 Three one-dimensional band structures 172
4.7 The Fermi distribution 174
4.8 Atomic energy levels and crystalline energy bands 176
4.9 The honeycomb lattice 184
4.10 Density of states for triangular and honeycomb lattices 185
4.11 Matrix elements for neighboring tight binding p-orbitals 187
4.12 Gauges for the square lattice Hofstadter model 189
4.13 Magnetic subbands for the square lattice Hofstadter model 191
4.14 T-breaking models with zero net flux per unit cell 192
4.15 Lattice gauge field configuration for a general flux configuration 194
4.16 Structure of polyacetylene, CH_x 195
4.17 Spectrum of the SSH Hamiltonian on a finite chain 199
4.18 Winding of $t(k)$ in the SSH model. 200
4.19 A Hermitian line bundle 209
4.20 Topological phase diagram for the Haldane honeycomb lattice model 214
4.21 Hofstadter's butterfly with gaps color-coded by Chern number 218
4.22 Colored Hofstadter butterfly for the isotropichoneycomb lattice system 219
4.23 Pseudopotentials and pseudopotential band structure of Si 229
4.24 Two smooth vector fields on the sphere 232
4.25 Smooth vector fields on the torus and on a $g = 2$ manifold 232
4.26 Composition of two circles 234

5.1 Fermi surfaces for two and three-dimensional structures . . . 239
5.2 Extended zones and their folding for the square lattice . . . 241
5.3 Brillouin zones and free electron Fermi seas for the square lattice . . . 242
5.4 First Brillouin zones for bcc and fcc solids . . . 243
5.5 The Fermi distribution . . . 245
5.6 Deformation of the complex integration contour in Eqn. 5.23 . . . 246
5.7 Fermi distributions in the presence of an external magnetic field . . . 250
5.8 DOS in a 3D electron gas in a magnetic field . . . 252
5.9 de Haas-van Alphen oscillations . . . 256
5.10 de Haas - van Alphen oscillations in a 2D Fermi gas . . . 257
5.11 Electron orbits in the calculation of the de Haas - van Alphen effect . . . 259
5.12 Electron scattering by impurities with a biasing electric field . . . 262
5.13 Cyclotron resonance peaks . . . 266
5.14 Electron-phonon vertices . . . 276
5.15 Frequency-dependent conductivity of liquid sodium . . . 281
5.16 Residual resistivity per percent impurity . . . 288
5.17 Current and momentum carried by holes . . . 290
5.18 Energy bands in aluminum . . . 294
5.19 Fermi surfaces for electron and hole bands in Aluminum . . . 296
5.20 The thermocouple . . . 299
5.21 The Peltier effect . . . 300
5.22 QT product for p-type and n-type Ge . . . 302
5.23 Transverse and longitudinal phonon polarizations . . . 305

6.1 Valence and conduction bands in semiconductors . . . 312
6.2 Pseudopotential band structures of diamond and zincblende semiconductors . . . 314
6.3 Resistivity of antimony-doped germanium . . . 315
6.4 Optical processes in Si . . . 316

6.5 Direct and indirect gap semiconductors . . . 317
6.6 Bands in metals, semiconductors, and insulators . . . 321
6.7 Relevant group II through group VI elements . . . 322
6.8 Donors and acceptors . . . 324
6.9 Evolution of chemical potential with temperature . . . 327
6.10 The p - n junction . . . 327
6.11 A GaAs - Al_xGa_xAs heterostructure . . . 328
6.12 The p - n junction in equilibrium . . . 329
6.13 Electrostatics of the p - n junction . . . 330
6.14 The biased p - n junction . . . 332
6.15 $j(V)$ for a biased p - n junction . . . 333
6.16 Zero and negative bias in a p-semiconductor - metal junction . . . 334
6.17 Strong positive bias in a p-semiconductor - metal junction . . . 335
6.18 The MOSFET . . . 336
6.19 GaAs$-$$Al_xGa_{1-x}As$ heterojunction . . . 337
6.20 Accumulation layer formation in an n-n heterojunction . . . 337
6.21 Accumulation and inversion in semiconductor heterojunctions . . . 338
6.22 Dielectric behavior in a polar crystal . . . 345

7.1 Scattering at a potential step . . . 351
7.2 Dimensionless two-terminal conductance for a potential step . . . 352
7.3 Two quantum scatterers in series . . . 357
7.4 $E(\phi)$ for free electrons on a ring . . . 365
7.5 Scattering problem for a ring enclosing a flux Φ . . . 366
7.6 Two-probe conductance $G(\phi, \kappa)$ of a ring with two scatterers (I) . . . 368
7.7 Two-probe conductance $G(\phi, \kappa)$ of a ring with two scatterers (II) . . . 369
7.8 Two-probe conductance $G(\phi, \kappa)$ of a ring with two scatterers (III) . . . 370
7.9 Localized and extended electronic states in a disordered system . . . 378

7.10 Localization length for $d = 2$ (left panel) and $d = 3$ (right panel) systems 381
7.11 Scaling function λ_M/M *versus* λ_∞/M 382
7.12 Sketch of the β-function for the localization problem for $d = 1, 2, 3$ 383

8.1 The *Aufbau* principle and the diagonal rule. 400

9.1 Phase diagrams of ^{3}He (left) and ^{4}He (right) 416
9.2 Specific heat and magnetic susceptibility of normal liquid ^{3}He 417
9.3 Two particle, two hole excitation of the state $|\mathsf{F}\rangle$ 418
9.4 A real horse and a quasi-horse 420
9.5 Distribution $\delta n_{\boldsymbol{k}\sigma}$ for a swollen Fermi surface 426
9.6 Distribution $\delta n_{\boldsymbol{k}\sigma}$ in the presence of a magnetic field 427
9.7 Distribution of quasiparticles in a frame moving with velocity $\boldsymbol{u}$ 429

10.1 The complex integration contour $\mathcal{C}$ 442
10.2 Particle-hole continuum of the noninteracting electron gas 459
10.3 Two pair excitations of the noninteracting electron gas 460
10.4 Low and high energy particle-hole excitations 461
10.5 $S(k, \omega)$ of the noninteracting electron gas 462
10.6 Perturbation expansion for RPA susceptibility bubble 467

11.1 Timeline of superconductors and their transition temperatures 482
11.2 Flux expulsion from a superconductor in the Meissner state 484
11.3 Phase diagrams for type I and type II superconductors 485
11.4 STM image of a vortex lattice in $NbSe_2$ 486
11.5 Dimensionless energy gap $\Delta(T)/\Delta_0$ in Nb, Ta, and Sn 487
11.6 Numerical solution to a Ginzburg-Landau domain wall 503
11.7 $\Delta g = g_{\rm SC} - g_{\rm N}$ for a thin extreme type-I film 506

12.1 Graphical solution to the Cooper problem 523

12.2 Feynman diagrams for electron-phonon processes . . . 526
12.3 John Bardeen, Leon Cooper, and J. Robert Schrieffer . . . 528
12.4 BCS coherence factors . . . 535
12.5 Temperature dependence of the energy gap in Pb . . . 537
12.6 Heat capacity in aluminum at low temperatures . . . 540
12.7 Contours for complex integration . . . 550

13.1 Graphical solution to the SCHA equation . . . 556
13.2 NIS tunneling for positive bias, zero bias, and negative bias . . . 566
13.3 Tunneling data by in a Pb/MgO/Mg sandwich junction . . . 567
13.4 SIS tunneling for positive bias, zero bias, and negative bias . . . 569
13.5 Current-voltage characteristics for a current-biased Josephson junction . . . 572
13.6 Fraunhofer pattern of Josephson junction subject to an in-plane field . . . 575
13.7 Phase flows for the equation $\ddot{\phi} + Q^{-1}\dot{\phi} + \sin\phi = j$ . . . 577
13.8 Dimensionless washboard potential and the RCSJ model . . . 578
13.9 I/I_c *versus* $\langle V\rangle/I_cR$ for different damping/temperature ratios . . . 583
13.10 Shapiro steps in the AC Josephson effect . . . 584
13.11 Graphical solution of $\dot{\psi} = -\delta + \varepsilon\, G(\psi)$ . . . 587
13.12 Phonon absorption and emission processes . . . 589
13.13 Ultrasonic attenuation in tin, compared with predictions of the BCS theory . . . 590
13.14 NMR relaxation rate $1/T_1$ *versus* temperature as predicted by BCS theory . . . 592
13.15 Real (σ_1) and imaginary (σ_2) parts of the conductivity of a superconductor . . . 598
13.16 Real part of the conductivity $\sigma_1(\omega, T)$ in $CeCoIn_5$ . . . 601

14.1 The *Aufbau* principle and the diagonal rule. . . . 613
14.2 Variation of L, S, and J among the 3d and 4f series . . . 615
14.3 Effect on s, p, and d levels of a cubic crystal field . . . 618
14.4 Splitting of one-electron states in different crystal field environments. . . . 619

14.5 Reduced magnetization curves for three paramagnetic salts *vs.* Brillouin theory . 622
14.6 A graduate student experiences the Stoner enhancement 629
14.7 Mean field phase diagram of the Hubbard model 633

16.1 The Kagomé lattice is a line graph of the honeycomb lattice 712
16.2 The checkerboard lattice is a line graph of the square lattice 713
16.3 Line graph of the square-octagon lattice 714
16.4 Energy bands for the line graph of the square-octagon lattice. 715
16.5 The pyrochlore lattice is an FCC Bravais lattice with a four element basis 716
16.6 A tetrahedron inscribed inside a cube 717
16.7 The α and β tetrahedra are related by inversion 718
16.8 First Brillouin zone for an FCC structure 720
16.9 Supertetrahedron basis for the pyrochlore lattice 722
16.10Pyrochlore lattice band structures in two bases 724
16.11Comparison of depleted pyrochlore band structures 725
16.12Two additional depleted pyrochlore band structures 727

List of Tables

1.1 $F(\lambda)$ and $n^*(\lambda)$ for part d 52

2.1 True Facts about two and three-dimensional crystallography 74
2.2 Standard notation for point group operations 78
2.3 The ten two-dimensional point groups 83
2.4 Character table for the group T_d 84
2.5 Elements and classes for T_d 85
2.6 Irreducible representations and basis functions for T_d and O 86
2.7 The 32 three-dimensional crystallographic point groups 87
2.8 Two element point group notation 89
2.9 Notations for simple crystallographic point groups 90
2.10 Notations for multi-axis point groups 90
2.11 HM and Schoenflies notation for the 32 crystallographic point groups 92
2.12 Character table for the double group of T_d 95
2.13 Character table for the point groups D_4, C_{4v}, and D_{2d} 97
2.14 Character table for the double groups of D_4 , C_{4v} , and D_{2d} 100
2.15 The 17 wallpaper groups and their short notation 103
2.16 The 73 symmorphic three-dimensional space groups 106
2.17 The 157 nonsymmorphic three-dimensional space groups 110

3.1 Abbreviation for symmetric compound indices $(\alpha\beta)$ 125

3.2 Debye temperatures and melting points for common elements 143
3.3 Composite indices for symmetric rank two tensors. 149

4.1 Common semiconductors and their band gaps 169

5.1 Effective masses of monovalent metals 283
5.2 Residual resistivity of copper per percent impurity 289
5.3 Electron-phonon interaction parameters for some metals 306

6.1 Common semiconductors and their properties at $T = 3000\,\mathrm{K}$ 313
6.2 Donor and acceptor binding energies in Si and Ge (in meV) 324
6.3 Donor states and their energies 325

9.1 Fermi liquid parameters for $^3\mathrm{He}\,\mathsf{N}$ 424

13.1 Frequency dependence of the BCS coherence factors 597

14.1 Rough order in which shells of the Periodic Table are filled 613
14.2 Electronic configuration of 3d-series metals 614
14.3 Hund's rules applied to p, d, and f shells 616
14.4 Molar susceptibilities of noble gas atoms and alkali and halide ions 621
14.5 Calculated and measured effective magneton numbers for rare earth ions 623
14.6 Calculated and measured effective magneton numbers for transition metal ions . 624

Chapter 1

Broken Symmetry

1.1 Introduction

What is the world made of? To a philosopher, this question lies at the intersection of ontology (the 'study' of being and existence) and mereology (the 'study' of parts and wholes). To a physicist, the answer very much depends on whom you ask, because ultimately it is a matter of energy scales. To a condensed matter physicist, the world consists of electrons, nuclei, and photons. That's pretty much it[1]. The characteristic energy scales in condensed matter typically range from milli-electron volts (meV) to electron volts (eV)[2]. By contrast, the protons and neutrons which constitute nuclei typically have binding energies on the order of MeV. The nucleons themselves consist of quarks and gluons. Quarks acquire their masses from coupling to the Higgs field, but some 99% of the nucleon mass is due to gluons and virtual quark-antiquark pairs, *i.e.* to the physics of QCD binding[3].

1.1.1 The mother of all Hamiltonians

The Hamiltonian for a single electron in the presence of a static external potential $V(\boldsymbol{r})$ and electromagnetic vector potential $\boldsymbol{A}(\boldsymbol{r})$ is given by

$$H = \frac{\boldsymbol{\pi}^2}{2m} + V(\boldsymbol{r}) + \frac{e\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{H} + \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}V\times\boldsymbol{\pi} + \frac{\hbar^2}{8m^2c^2}\boldsymbol{\nabla}^2V + \frac{(\boldsymbol{\pi}^2)^2}{8m^3c^2} + \dots\,, \tag{1.1}$$

[1] Sometimes experimentalists use muons or even positrons to probe their samples, but the samples themselves consist of electrons and nuclei.

[2] Sometimes higher energy probes on the order of keV are used, for example in photoemission spectroscopy.

[3] So the next time you overhear someone holding forth about how the Higgs field (or, worse still, the Higgs boson) gives mass to everything in the universe, you can tell them that they are full of shit.

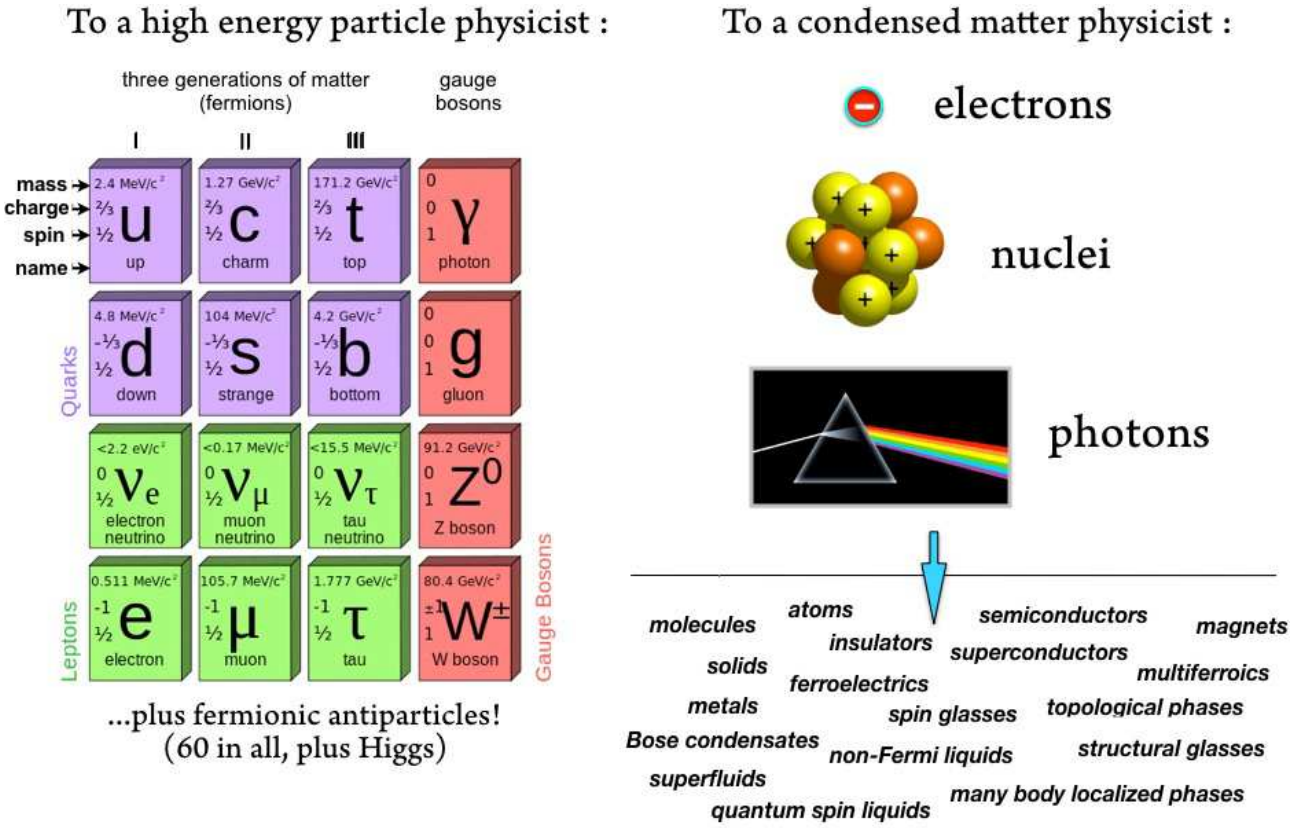


Figure 1.1: What is the world made of? It depends on whom you ask.

where $\boldsymbol{\pi} = \boldsymbol{p} + \frac{e}{c}\boldsymbol{A}$. Where did this come from? From the Dirac equation,

$$i\hbar\frac{\partial\Psi}{\partial t} = \begin{pmatrix} mc^2 + V & c\,\boldsymbol{\sigma}\cdot\boldsymbol{\pi} \\ c\,\boldsymbol{\sigma}\cdot\boldsymbol{\pi} & -mc^2 + V \end{pmatrix}\Psi = E\Psi\ . \tag{1.2}$$

The wavefunction Ψ is a four-component Dirac spinor. Since mc^2 is the largest term for our applications, the upper two components of Ψ are essentially the positive energy components. However, the Dirac Hamiltonian mixes the upper two and lower two components of Ψ. One can 'unmix' them by making a canonical transformation,

$$H \longrightarrow \hat{H}' \equiv e^{iS}\,\hat{H}\,e^{-iS}\ , \tag{1.3}$$

where S is Hermitian, to render $\hat{H}'$ block diagonal. With $E = mc^2 + \varepsilon$, the effective Hamiltonian is given by (14.3). This is known as the Foldy-Wouthuysen transformation, the details of which may be found in many standard books on relativistic quantum mechanics and quantum field theory (*e.g.* Bjorken and Drell, Itzykson and Zuber, *etc.*) and are recited in §14.11 below. Note that the Dirac equation leads to $g = 2$. If we go beyond "tree level" and allow for radiative corrections within QED, we obtain a perturbative expansion,

$$g = 2\left\{1 + \frac{\alpha}{2\pi} + \mathcal{O}(\alpha^2)\right\}\ , \tag{1.4}$$

הסיכום

הוגה אחד זקן
מחדרו אל תלמידיו יצא
ואמר בנקשו במקל:
שמעו אחרון דברי אשר אשא
כי הגיע הזמן לסכם,
לאחר שנות בינה ועיצה,
מה הם, לסוף מיצוי,
הדברים שמהם העולם עשוי.

העולם, נכבדי – שאחרי כל תהיות
ופליאות בו דברי אמורים –
עשוי ארץ, שמים, שיחים, סנוניות,
אוניות, גלי ים, נהרות, ערים
גדולות וקטנות, ירידים, צנוניות,
כל מיני פטישים, נייר, משורים,
רחובות, חדרים, – כך הוסיף למנות

ולמנות – גאיות, שפלות, הרים
וגבעות – ושעה שרופא ואחות
כבר עליו גחנו בבית–החולים,
עדיין היו שפתיו נעות
ומונות – קלחות, כפיות, ספרים,
מגרות, צנצנות – ומילים אחרונות
שהשמיע היו: נעצים, כפתורים.

Summary

One old thinker
out from his chamber to his disciples came
and he said as he rapped with his cane:
harken to the last of my words
for it is now time to summarize,
after years of reason and deliberation,
what, in the last analysis,
are the things of which the world is made.



Natan Alterman
(1910 - 1970)

The world, my distinguished sirs – which after all wondering
and marveling my words concern –
is made of land, sky, bushes, sparrows,
ships, waves, rivers, cities
large and small, fairs, radishes,
all kinds of hammers, paper, saws,
streets, rooms - thus he continued counting

and counting – valleys, plains, mountains,
hills – and while the doctor and nurse
were bent over him in the hospital,
still his lips were moving
and counting – cauldrons, spoons, books,
drawers, jars – and the last words
which he uttered were: thumbtacks, buttons.

(translation by Assa Auerbach)

Figure 1.2: What is the world made of? Another point of view.

where $\alpha = e^2/\hbar c \approx 1/137$ is the fine structure constant[4].

There are two terms in (14.3) which involve the electron's spin:

$$\begin{aligned} \text{Zeeman interaction} \ : \ \hat{H}_{\rm Z} &= \frac{e\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{H} \\ \text{Spin-orbit interaction} \ : \ \hat{H}_{\rm SO} &= \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}V\times\left(\boldsymbol{p}+\tfrac{e}{c}\boldsymbol{A}\right) \quad . \end{aligned} \tag{1.5}$$

The numerical value for $\mu_{\rm B}$ is

$$\begin{aligned} \mu_{\rm B} &= \frac{e\hbar}{2mc} = 5.788\times10^{-9}\,{\rm eV/G} \\ \mu_{\rm B}/k_{\rm B} &= 6.717\times10^{-5}\,{\rm K/G} \quad . \end{aligned} \tag{1.6}$$

[4]Note that with $\mu_{\rm n} = e\hbar/2m_{\rm p}c$ for the nuclear magneton, $g_{\rm p} = 2.793$ and $g_{\rm n} = -1.913$. These results immediately suggest that there is composite structure to the nucleons, *i.e.* quarks.

So on the scale of electron volts, laboratory scale fields ($H \lesssim 10^6$ G) are rather small (and ~ 2000 times smaller for nucleons).

In condensed matter, the potential $V(\boldsymbol{r})$ for the electron is a sum of terms from individual ion cores, *i.e.* $V(\boldsymbol{r}) = \sum_a v_a(\boldsymbol{r} - \boldsymbol{R}_a)$ where $v_a(\boldsymbol{r})$ is the atomic potential and $\boldsymbol{R}_a$ is the location of the $a^{\rm th}$ nucleus. For Coulomb interactions, $v_a(\boldsymbol{r}) = -Z_a e^2/r$. The ion cores interact as well, with a potential $u_{ab}(\boldsymbol{r}) = Z_a Z_b e^2/r$. Putting it all together, we have[5]

$$\begin{aligned}\mathcal{H} = \sum_i \frac{\boldsymbol{\pi}_i^2}{2m} - \frac{e\hbar}{2mc}\boldsymbol{H}\cdot\sum_i \boldsymbol{\sigma}_i - \sum_{i<j}\frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{i,a} v_a(\boldsymbol{r}_i - \boldsymbol{R}_a) \\ + \frac{\hbar}{4m^2c^2}\sum_{i,a}\boldsymbol{\sigma}_i\cdot\boldsymbol{\nabla}_i v_a(\boldsymbol{r}_i - \boldsymbol{R}_a)\times\boldsymbol{\pi}_i + \sum_a \frac{\boldsymbol{\Pi}_a^2}{2M_a} + \sum_{a<b} u_{ab}(\boldsymbol{R}_a - \boldsymbol{R}_b) \quad ,\end{aligned} \tag{1.7}$$

with $\boldsymbol{\Pi}_a = \boldsymbol{P}_a - Z_a e A(\boldsymbol{R}_a)/c$, and subject to the commutation relations,

$$\left[p_i^\alpha, r_j^\beta\right] = -i\hbar\delta_{ij}\delta^{\alpha\beta} \quad , \quad \left[P_a^\alpha, R_b^\beta\right] = -i\hbar\delta_{ab}\delta^{\alpha\beta} \quad , \quad \left[\sigma_i^\mu, \sigma_j^\nu\right] = 2i\delta_{ij}\epsilon^{\mu\nu\lambda}\,\sigma_i^\lambda \quad . \tag{1.8}$$

We may call Eqn. 1.7 the *Mother of all Hamiltonians*.

In a crystal, we may write $\boldsymbol{R}_a = \boldsymbol{R}_a^0 + \delta\boldsymbol{R}_a$, where $\{\boldsymbol{R}_a^0\}$ are the mean locations of the ion cores. These form a regular lattice. The deviations $\delta\boldsymbol{R}_a$ describe quantum mechanical fluctuations about the fixed crystalline coordinates. The quantized vibrations of a crystalline lattice are called *phonons*. Thus, the mother of all Hamiltonians describes electrons, which possess quantized spin, interacting among themselves and with fluctuating ion cores.

1.2 Classical and Quantum Statistical Physics

While the Mother of all Hamiltonians is too complex for us to attack directly, progress can oftentimes be made by examining parts of it, such as the spectrum of lattice vibrations, or the electronic energy spectrum in the presence of a crystalline lattice potential, possibly including spin-orbit effects. For such noninteracting systems, where the Hamiltonian is quadratic in electron or phonon operators, we can often go quite far in obtaining complete solutions. The problem becomes complicated when we consider the effects of *quenched impurities*, such as vacancies or substitutional defects in the crystalline lattice, which are random in nature. We must then consider how observable quantities are distributed with respect to an *ensemble* of configurations of the randomness. This can give rise to interesting, robust, and fundamentally new possibilities, such as localization.

In order to account for phonon-phonon, electron-phonon, and/or electron-electron interactions, there are few useful techniques at our disposal. We can always resort to perturbation

[5] We drop the last two and all following terms in Eqn. 14.3.

theory in the interactions, but this can lead us astray if the true ground state of the interacting system lies on the other side of a phase boundary from the noninteracting ground state. For gapless systems, such as metals, the situation is particularly tricky and we generally have no good reason why perturbation theory should converge[6].

Suppose we have an idealized system described by a model Hamiltonian H which is simpler than $\mathcal{H}$ and where we might hope to make some progress. The thermodynamic properties of the system are calculable from the Helmholtz free energy $F = -k_{\rm B}T \ln \mathsf{Tr} \exp(-H/k_{\rm B}T)$ or one of its Legendre transforms, such as the grand potential[7] $\Omega = -k_{\rm B}T \ln \mathsf{Tr} \exp(-K/k_{\rm B}T)$, where $K = H - \mu N$. We write the Hamiltonian as

$$H = H_0 - \sum_\alpha h_\alpha \, Q_\alpha \quad , \tag{1.9}$$

where Q_α is a Hermitian operator and h_α is a conjugate field for each index α. Then the free energy F is a function of the fields, *i.e.* $F = F\big(\{h_\alpha\}\big)$ and we have that the thermodynamic average of Q_α is

$$\big\langle Q_\alpha \big\rangle = -\frac{\partial F}{\partial h_\alpha} \quad . \tag{1.10}$$

One can now define the *susceptibility*

$$\chi_{\alpha\beta} = \frac{\partial \langle Q_\alpha \rangle}{\partial h_\beta} = -\frac{\partial^2 F}{\partial h_\alpha \, \partial h_\beta} \quad . \tag{1.11}$$

Dynamical responses are defined for Hamiltonians of the form

$$H(t) = H_0 - \sum_\alpha h_\alpha(t) \, Q_\alpha \quad . \tag{1.12}$$

Time-dependent first order perturbation theory then yields

$$\big\langle Q_\alpha(t) \big\rangle = \int\limits_{-\infty}^{\infty} dt' \; \chi_{\alpha\beta}(t-t') \, h_\beta(t') + \mathcal{O}(h^2) \tag{1.13}$$

with

$$\chi_{\alpha\beta}(t-t') = \frac{i}{\hbar} \big\langle \big[Q_\alpha(t), Q_\beta(t') \big] \big\rangle \, \Theta(t-t') \quad . \tag{1.14}$$

[6]Oftentimes in quantum field theory, perturbation theory yields an *asymptotic series* in the coupling parameter(s). In this case the series expansions for physical quantities may be formally divergent, but provided the couplings are not too large, yield better and better approximations up until a particular order in the expansion. For example, in quantum electrodynamics, the relevant coupling is the fine structure constant $\alpha = e^2/\hbar c \approx 1/137$, and perturbation theory, typically evaluated using Feynman diagrams, is expected to start diverging at order α^{-1}.

[7]Also called the Landau free energy.

Note that the response is *causal*, *i.e.* the value of $\big\langle Q_\alpha(t)\big\rangle$ depends only on the values of $\big\{h_\beta(t')\big\}$ for $t' < t$. The *spectral representation* of the response function $\chi_{\alpha\beta}(t-t')$ is defined via its Fourier transform,

$$\begin{aligned}\hat\chi_{\alpha\beta}(\omega+i\epsilon) &\equiv \frac{i}{\hbar}\int\limits_0^\infty dt\,\big\langle[Q_\alpha(t),Q_\beta(0)]\big\rangle\,e^{i\omega t}\,e^{-\epsilon t} \\ &= \frac{1}{\hbar Z}\sum_{m,n} e^{-\beta E^0_m}\left\{\frac{\big\langle\, m\,\big|\,Q_\beta\,\big|\,n\,\big\rangle\big\langle\, n\,\big|\,Q_\alpha\,\big|\,m\,\big\rangle}{\omega+(E^0_n-E^0_m)/\hbar+i\epsilon} - \frac{\big\langle\, m\,\big|\,Q_\alpha\,\big|\,n\,\big\rangle\big\langle\, n\,\big|\,Q_\beta\,\big|\,m\,\big\rangle}{\omega-(E^0_n-E^0_m)/\hbar+i\epsilon}\right\}\quad ,\end{aligned} \tag{1.15}$$

where $Z = \mathsf{Tr}\,\exp(-H_0/k_{\rm B}T)$ and where $H_0\,|n\rangle = E^0_n\,|n\rangle$.

Finally, the indices α and β may be appended by spatial coordinates in the case of local operators $Q_\alpha(\boldsymbol{r})$, in which case we write

$$H(t) = H_0 - \sum_\alpha \int d^dr\; h_\alpha(\boldsymbol{r},t)\,Q_\alpha(\boldsymbol{r}) \quad . \tag{1.16}$$

For example, $Q_\alpha(\boldsymbol{r})$ could be taken to be the local density $n(\boldsymbol{r})$ or a component $S^\alpha(\boldsymbol{r})$ of the local spin density. One then defines the response function

$$\chi_{\alpha\beta}(\boldsymbol{r},t\,|\,\boldsymbol{r};,t') \equiv \frac{\delta\big\langle Q_\alpha(\boldsymbol{r},t)\big\rangle}{\delta h_\beta(\boldsymbol{r}',t')} \quad . \tag{1.17}$$

1.2.1 Correspondence between quantum and classical statistical mechanics

In classical statistical physics, in systems where spontaneous symmetry breaking results in an ordered phase in which the local order parameter field takes on a finite average, *i.e.* $\langle\phi(\boldsymbol{r})\rangle \neq 0$, the connected correlation function at large distances in the vicinity of the transition behaves as (for r fixed and $T \to T_{\rm c}$)

$$C(\boldsymbol{r},T) = \big\langle\phi(\boldsymbol{r})\,\phi(\boldsymbol{0})\big\rangle - \langle\phi(\boldsymbol{r})\rangle\,\langle\phi(\boldsymbol{0})\rangle \sim A\,r^{2-d}\,e^{-r/\xi} \quad , \tag{1.18}$$

where $\xi(T) \sim |T-T_{\rm c}|^{-\nu}$ is the spatial correlation length and ν the correlation length exponent[8]. Pretty much all classical phase transitions are described by a continuum field theory of sorts, with a free energy functional

$$F\big[\boldsymbol{\phi}(\boldsymbol{r})\big] = \int d^dx\; f(\boldsymbol{\phi},\boldsymbol{\nabla\phi}) \quad . \tag{1.19}$$

[8]Precisely at $T = T_{\rm c}$, the correlation function behaves as $\big\langle\phi(\boldsymbol{r})\,\phi(\boldsymbol{0})\big\rangle \sim r^{-(d-2+\eta)}$, where η is the so-called *anomalous dimension*.

Here $\boldsymbol{\phi}(\boldsymbol{x}) = \{\phi^1, \ldots, \phi^N\}$ is an N-component real field, and by $\boldsymbol{\nabla\phi}$ we mean the set of dN first derivatives $\partial\phi^a/\partial x^i$ with $a \in \{1, \ldots, N\}$ and $i \in \{1, \ldots, d\}$. The partition function is given by the functional integral

$$Z = \int D\boldsymbol{\phi}(\boldsymbol{r})\, e^{-F[\boldsymbol{\phi}(\boldsymbol{r})]} \quad . \tag{1.20}$$

For the Ising model, one generally has $N = 1$ and one may take

$$f_{\mathsf{Ising}}(\phi, \boldsymbol{\nabla}\phi) = \tfrac{1}{2}(\boldsymbol{\nabla}\phi)^2 + \tfrac{1}{2}a\phi^2 + \tfrac{1}{4}b\phi^4 \quad , \tag{1.21}$$

i.e. a scalar ϕ^4 theory in d dimensions.

For bosonic quantum systems, the path integral formulation results in the following prescription. The partition function is once again given by a functional integral,

$$Z = \int D\boldsymbol{\phi}(\boldsymbol{r}, \tau)\, e^{-S_{\mathsf{E}}[\boldsymbol{\phi}(\boldsymbol{r},\tau)]} \quad , \tag{1.22}$$

but now the field $\boldsymbol{\phi}(\boldsymbol{r}, \tau)$ is dependent both on spatial coordinates $\boldsymbol{r}$ as well as a "Euclidean time" coordinate $\tau \in \big[0, \hbar/k_{\mathrm{B}}T\big]$, where T is the temperature. $S_{\mathsf{E}}[\phi]$ is the *Euclidean action*,

$$S_{\mathsf{E}}[\boldsymbol{\phi}(\boldsymbol{r}, t)] = \int_0^{\hbar\beta} d\tau \int d^d x\, \mathcal{L}_{\mathsf{E}}(\boldsymbol{\phi}, \boldsymbol{\nabla\phi}, \partial_\tau\boldsymbol{\phi}) \quad , \tag{1.23}$$

where $\mathcal{L}_{\mathsf{E}}(\boldsymbol{\phi}, \boldsymbol{\nabla\phi}, \partial_\tau\boldsymbol{\phi})$ is the *Euclidean Lagrangian density*. Note that spacetime consists of a 'slice' of finite temporal width $\hbar/k_{\mathrm{B}}T$. If we define the coordinate $x^0 \equiv c\tau$, where c is a (possibly arbitrary) constant with dimensions of velocity, then we see that the action is given by a $(d+1)$-dimensional integral over a slab of finite width $L_\tau = c\tau/k_{\mathrm{B}}T$ in the x^0 dimension. Hence the oft-heard maxim, *quantum mechanics in d (space) dimensions is equivalent to classical statistical mechanics in $(d+1)$ dimensions*. While this is essentially true, it may be the case, as we emphasize below, that the resulting $(d+1)$-dimensional classical model derived from a given quantum theory will be *anisotropic* in that the x^0 direction is special, both due to its finite extent for $T > 0$ systems and also due to the form of the derivative terms in the Lagrangian density. Indeed, nonrelativistic quantum theories generally result in such anisotropic classical models, although an emergent low-energy relativistic symmetry is sometimes present for particular systems, such as quantum antiferromagnets or acoustic phonons, or perhaps in the vicinity of a quantum critical point, such as near the tips of the 'Mott lobes' of the Bose Hubbard model.

For bosonic systems, $\boldsymbol{\phi}(\boldsymbol{r}, \tau)$ may be taken to be an N-component real field, or a complexified version thereof[9]. As an example of a Euclidean Lagrangian density, consider the case

$$\mathcal{L}_{\mathsf{E}}(\boldsymbol{\phi}, \boldsymbol{\nabla\phi}, \partial_\tau\boldsymbol{\phi}) = \frac{1}{2}\left(\frac{\partial\boldsymbol{\phi}}{\partial\tau}\right)^2 + \tfrac{1}{2}K(\boldsymbol{\nabla\phi})^2 + V(\boldsymbol{\phi}) \quad . \tag{1.24}$$

[9]For example $(\phi^1, \phi^2) \to \phi \equiv \phi^1 + i\phi^2$.

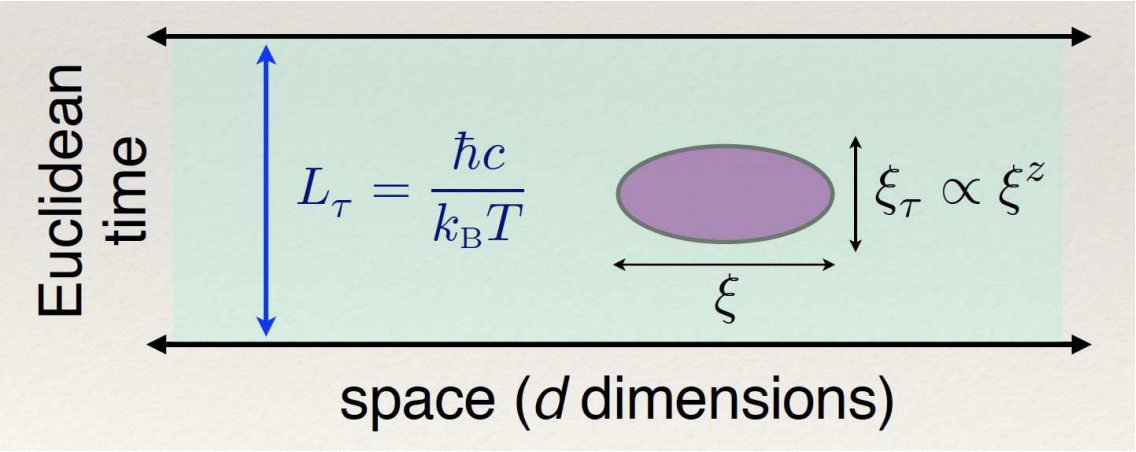


Figure 1.3: Spatial and temporal correlations in the vicinity of a quantum phase transition. (The parameter c carries dimensions of velocity and may be arbitrary.)

This model may possess various discrete or continuous symmetries. For example, if we let $g \in \mathsf{O}(N)$ be an arbitrary rotation in the internal space of the field $\boldsymbol{\phi}$, and if $V(g\boldsymbol{\phi}) = V(\boldsymbol{\phi})$ for all such g, then the model is said to possess an $\mathsf{O}(N)$ symmetry (which may be broken - see below. Another important symmetry of the model is the $\mathsf{O}(d+1)$ symmetry resulting from mixing space and time coordinates (with appropriate rescaling so as to render them of the same dimension). The *real time* correlations of this model will then exhibit a Lorentz symmetry. But not all QFTs have such a relativistic symmetry. Consider the case $N = 2$, for example, with

$$\mathcal{L}_{\mathsf{E}}(\boldsymbol{\phi}, \boldsymbol{\nabla}\boldsymbol{\phi}, \partial_\tau \boldsymbol{\phi}) = i\,\phi_1\,\partial_\tau\,\phi_2 - i\,\phi_2\,\partial_\tau\,\phi_1 + \tfrac{1}{2}K(\boldsymbol{\nabla}\boldsymbol{\phi})^2 + V(\boldsymbol{\phi}^2) \quad . \tag{1.25}$$

Here $\boldsymbol{\phi} = (\phi_1, \phi_2)$. The 'kinetic' term is only linear in time derivatives, while the spatial derivative terms appear quadratically. This model has no relativistic symmetry, but it does have a global $\mathsf{O}(2)$ symmetry.

For fermionic systems, one also has a path integral, but it is over anticommuting Grassmann fields $\psi_\sigma(\boldsymbol{r}, t)$ and their conjugates $\bar{\psi}_\sigma(\boldsymbol{r}, \tau)$, which we shall discuss in due time[10]. An example might be

$$\mathcal{L}_{\mathsf{E}}(\psi_\sigma, \bar{\psi}_\sigma, \boldsymbol{\nabla}\psi_\sigma, \boldsymbol{\nabla}\bar{\psi}_\sigma, \partial_\tau\psi_\sigma, \partial_\tau\bar{\psi}_\sigma) = \bar{\psi}_\sigma\Big(\hbar\,\partial_\tau - \mu - \frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2\Big)\psi_\sigma + U\bar{\psi}_\uparrow\bar{\psi}_\downarrow\psi_\downarrow\psi_\uparrow \tag{1.26}$$

Note that the Laplacian term may be spatially integrated by parts to yield a term proportional to $\boldsymbol{\nabla}\bar{\psi}_\sigma \cdot \boldsymbol{\nabla}\psi_\sigma$. This $\mathcal{L}_{\mathsf{E}}$ corresponds to the Hubbard model.

As we approach a quantum phase transition, which occurs at temperature $T = 0$ and a critical value $g = g_{\rm c}$ of some coupling, the spatial correlation length $\xi(g)$ grows as $\xi(g) \sim |g - g_{\rm c}|^{-\nu}$. For systems with relativistic invariance, the correlation length in the imaginary time direction

[10] Real time, rest assured.

$\xi_\tau(g)$ behaves in the same way, *i.e.* with the same correlation length exponent. But in general, the temporal exponent is different and may be written as $\nu_\tau = z\nu$, where z is the *dynamic critical exponent*. Thus, $\xi_\tau(g) \sim |g - g_c|^{-z\nu}$. As $g \to g_c$, both $\xi(g)$ and $\xi_\tau(g)$ will diverge. But at finite temperature, the width of the time slice is finite, given by $\hbar/k_B T$. Thus, when g is sufficiently close to g_c such that $\xi_\tau(g) \gg L_\tau = \hbar c/k_B T$, where c is an arbitrary measure of velocity, the fields will be 'locked' in the imaginary time direction. In this case, the expression for the partition function reverts to the classical result. What this tells us is that *quantum mechanics is irrelevant at any finite temperature T, i.e.* QM does not change the critical exponents at any finite temperature phase transitions. It should be emphasized that the above argument presumes the existence of a local order parameter field $\phi(\boldsymbol{r})$ whose correlations asymptotically decay on the scale of the correlation length $\xi(g, T)$. In the case of topological phases, where there is no such local order parameter, this argument may not apply.

1.3 Phases of matter

Condensed matter physicists are not interested in electrons, nuclei, and photons in isolation. Rather it is the collective properties of these interacting constituents which leads to the remarkably rich phenomenology of condensed matter. This richness is manifested by the bewildering array of *phases of matter* that can be conjured. Until relatively recently, phases of matter were described based on the *Landau paradigm*, in which an ordered phase is described by one or more nonzero *order parameters*, each of which describes a *broken global symmetry*.

1.3.1 Spontaneous symmetry breaking

In quantum mechanics, the eigenstates of a Hamiltonian H_0 which commutes with all the *generators* of a symmetry group G may be classified according to the *representations* of that group. Typically this entails the appearance of degeneracies in the eigenspectrum, with degenerate states transforming into each other under the group operations. Adding a perturbation V to the Hamiltonian which breaks G down to a subgroup H will accordingly split these degeneracies, and the new multiplets of $H = H_0 + V$ are characterized by representations of the lower symmetry group H.

In quantum field theory, or in the thermodynamic limit of a classical system, as a consequence of the infinite number of degrees of freedom, symmetries may be *spontaneously broken*. This means that even if the Hamiltonian H (or action S) for the field theory is invariant under a group G of symmetry transformations, the ground state or thermodynamic density matrix may not be invariant under the full symmetry group G. The presence or absence of spontaneous symmetry breaking (SSB), and its detailed manifestations, will in general depend on the couplings, or the temperature in the case of quantum statistical mechanics. SSB is usually associated with the presence of a local *order parameter* which transforms nontrivially under some

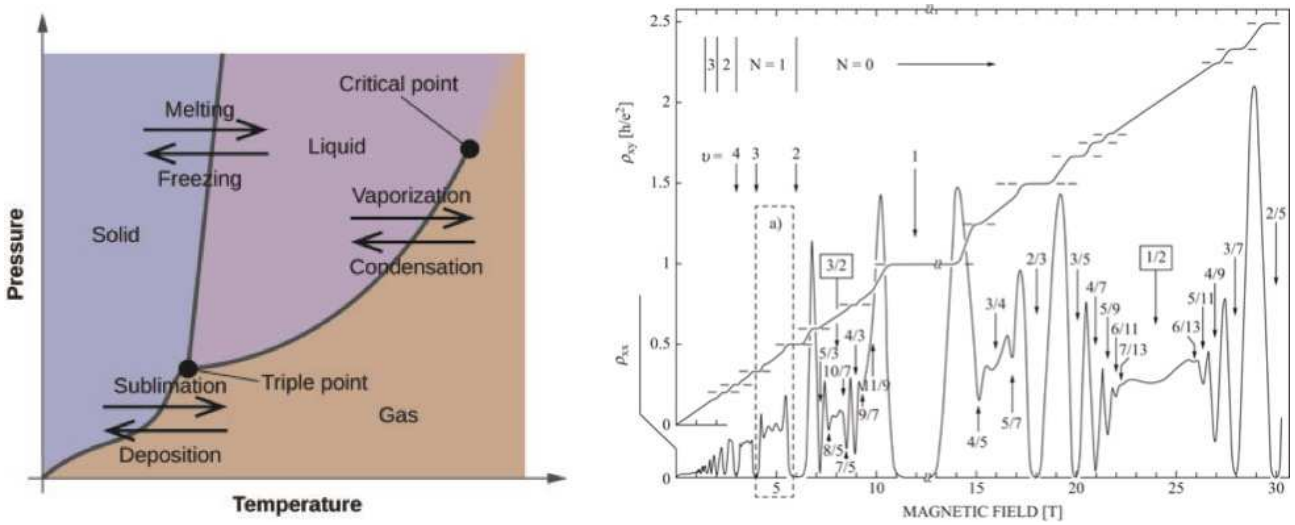


Figure 1.4: Left: Phase diagram of H_2O in the (T, p) plane. Right: Low temperature ($T \approx 150\,\mathrm{mK}$) longitudinal resistivity ρ_{xx} and Hall resistivity ρ_{xy} as a function of applied magnetic field in a two-dimensional electron gas system (GaAs/AlGaAs heterostructure), from R. Willett *et al.*, *Phys. Rev. Lett.* **59**, 1776 (1987). Each dip in ρ_{xx} and concomitant plateau in ρ_{xy} corresponds to a distinct phase of matter.

group operations, and whose whose quantum statistical average vanishes in a fully symmetric phase, but takes nonzero values in symmetry-broken phase[11]. The parade example is the Ising model, $H = -\sum_{i<j} J_{ij}\,\sigma_i\,\sigma_j$, where each $\sigma_i = \pm 1$, the subscript i indexes a physical location in space, such as a site $\boldsymbol{R}_i$ on a particular lattice. The model is explicitly $\mathbb{Z}_2$ symmetric under $\sigma_i \to \varepsilon\sigma_i$ for all i, where $\varepsilon \in \{+1, -1\}$, yet if the interaction matrix $J_{ij} = J(\boldsymbol{R}_i - \boldsymbol{R}_j)$ is short-ranged a nd the space dimension d is greater than one, there is a *critical temperature* $T_{\rm c}$ below which SSB sets in, and the system develops a spontaneous magnetization $\phi = \langle\sigma_i\rangle$. You know how in quantum mechanics, the eigenstates of a particle moving in one-dimensional double-well potential $V(x) = V(-x)$ can be classified by their parity eigenvalues P, and the lowest two energy states are respectively symmetric ($\mathsf{P} = +1$) and antisymmetric ($\mathsf{P} = -1$), and are delocalized among both wells. For a quantum field theory, however, with (Euclidean) Lagrangian density $\mathcal{L}_{\rm E} = \frac{1}{2}(\boldsymbol{\nabla}\phi)^2 + V(\phi)$, for $d > 1$ and $T < T_{\rm c}$, the system actually picks the left or the right well, so that $\langle\phi(\boldsymbol{r})\rangle \neq 0$. Another example is the spontaneously broken $\mathrm{O}(2)$ invariance of superfluids, where the boson annihilation operator $\psi(\boldsymbol{r})$ develops a spontaneous average $\big\langle\psi(\boldsymbol{r})\big\rangle = \sqrt{n_0}\,e^{i\theta}$, where n_0 is the condensate density and θ the condensate phase.

Truth be told, the above description is a bit of a swindle. In the ferromagnetic ($J_{ij} > 0$) Ising model, for example, at $T = 0$, there are still two ground states, $|\Uparrow\rangle \equiv |\uparrow\uparrow\uparrow\cdots\rangle$ and $|\Downarrow\rangle \equiv |\downarrow\downarrow\downarrow\cdots\rangle$. The (ergodic) zero temperature density matrix is $\rho_0 = \frac{1}{2}|\Uparrow\rangle\langle\Uparrow| + \frac{1}{2}|\Downarrow\rangle\langle\Downarrow|$, and

[11] While SSB is generally associated with the existence of a phase transitions, not all phase transitions involve SSB. Exceptions include the Kosterlitz-Thouless transition, and also those topological phases which have no local order parameter.

therefore $\langle\sigma_i\rangle = \mathsf{Tr}\left(\rho_0\,\sigma_i\right) = 0$. The order parameter apparently has vanished. WTF?! There are at least two compelling ways to resolve this seeming conundrum:

(a) First, rather than defining the order parameter of the Ising model, for example, to be the expected value $m = \langle\sigma_i\rangle$ of the local spin[12], consider instead the behavior of the *correlation function* $C_{ij} = \langle\sigma_i\,\sigma_j\rangle$ in the limit $d_{ij} = |\boldsymbol{R}_i - \boldsymbol{R}_j| \to \infty$. In a disordered phase, there is no correlation between infinitely far separated spins, hence $\lim_{d_{ij}\to\infty} C_{ij} = 0$. In the ordered phase, this is no longer true, and we define the *spontaneous magnetization* m from the long distance correlator: $m^2 \equiv \lim_{d_{ij}\to\infty}\langle\sigma_i\,\sigma_j\rangle$. In this formulation, SSB is associated with the emergence of *long-ranged order* in the correlators of operators which transform nontrivially under the symmetry group.

(ii) Second, we could impose an external field which *explicitly* breaks the symmetry, such as a Zeeman term $H' = -h\sum_i \sigma_i$ in the Ising model. We now compute the magnetization (per site) $m(T,h,V) = \langle\sigma_i\rangle$ as a function of temperature T, the external field h, *and* the volume V of our system. The order parameter $m(T)$ in zero field is then defined as

$$m(T) = \lim_{h\to 0}\lim_{V\to\infty} m(T,h,V) \quad . \tag{1.27}$$

The order of limits here is crucially important. The thermodynamic limit $V \to \infty$ is taken first, which means that the energy difference between $|\Uparrow\rangle$ and $|\Downarrow\rangle$, being proportional to hV, diverges, thus infinitely suppressing the $|\Downarrow\rangle$ state if $h > 0$ (and the $|\Uparrow\rangle$ state if $h < 0$). The magnitude of the order parameter will be independent on the way in which we take $h \to 0$, but its sign will depend on whether $h \to 0^+$ or $h \to 0^-$, with $\mathsf{sgn}(m) = \mathsf{sgn}(h)$. Physically, the direction in which a system orders can be decided by the presence of small stray fields or impurities. An illustration of how this works in the case of ideal Bose gas condensation is provided in the appendix §1.7 below.

Note that in both formulations, SSB is necessarily associated with the existence of a local operator $\mathcal{O}_i$ which is identified as the order parameter field. In (i) the correlations $\langle\mathcal{O}_i\,\mathcal{O}_j\rangle$ exhibit long-ranged order in the symmetry-broken phase. In (ii) $\mathcal{O}_i$ is the operator to which the external field h_i couples.

For a ferromagnet, the order parameter is the magnetization density, $\boldsymbol{m}$, and the broken symmetry is the group of rotations $\mathsf{O}(3)$ or $\mathsf{SU}(2)$, or possibly $\mathbb{Z}_2$. Under a group operation $g \in \mathsf{G}$, the order parameter transforms as $\boldsymbol{m} \to g\boldsymbol{m}$. For a crystal or charge density wave, the order parameters are the Fourier components $\varrho_{\boldsymbol{G}}$ of the density at a series of wavevectors which comprise the reciprocal lattice of the structure. The broken symmetry is that of continuous translation, *i.e.* the group $\mathbb{R}^d$ under addition, and under the group operation $t_{\boldsymbol{a}}$ corresponding to translation by $\boldsymbol{a} \in \mathbb{R}^d$, we have $\varrho_{\boldsymbol{G}} \to t_{\boldsymbol{a}}\,\varrho_{\boldsymbol{G}} = \varrho_{\boldsymbol{G}}\,e^{i\boldsymbol{G}\cdot\boldsymbol{a}}$. A given order parameter ϕ will in general depend on temperature T, pressure p, applied magnetic field $\boldsymbol{H}$, various coupling

[12] We assume translational invariance, which means $\langle\sigma_i\rangle$ is independent of the site index i.

parameters which enter the Hamiltonian $\{g_i\}$, *etc.* A multidimensional plot which labels the various phases of a system as a function of all these parameters is known as a *phase diagram*. Points, lines, or surfaces separating different phases are the loci of *phase transitions*, where the free energy and the order parameters of the system exhibit singularities. At a *first order transition*, certain properties change discontinuously. The canonical example of a first order transition is the freezing or boiling of liquid water to solid ice or gaseous vapor. These transitions involve a discontinuous change in the density as T or p is varied so as to cross a phase boundary[13]. At a *second order transition*, one or more order parameters vanish as a phase boundary is approached. Within the Landau paradigm, a nonzero order parameter is associated with the spontaneous breaking of a symmetry, such as spin rotation (in a magnet), space translation (in a crystal), *etc.* Furthermore, a second order transition between phases A and B is possible only if the symmetry groups $\mathsf{G}_{\rm A}$ of the A phase and $\mathsf{G}_{\rm B}$ of the B phase, both of which must be subgroups of the symmetry group G of the Hamiltonian, satisfy $\mathsf{G}_{\rm A} \subset \mathsf{G}_{\rm B}$ or $\mathsf{G}_{\rm B} \subset \mathsf{G}_{\rm A}$. A phase transition which takes place at $T = 0$ as a function of other parameters (p, H, *etc.*) is called a *quantum phase transition*.

1.3.2 Beyond the Landau paradigm

The classical $\mathsf{O}(2)$ model in $d = 2$ is precluded from achieving long-ranged order and spontaneous $\mathsf{O}(2)$ symmetry breaking at any finite temperature by the Hohenberg-Mermin-Wagner theorem, which we shall discuss below in §1.5.3. Nevertheless, the model does exhibit a second order phase transition, known as the Berezinskii-Kosterlitz-Thouless (BKT) transition[14]. The energy density in the continuum model is written as $\varepsilon(\boldsymbol{r}) = \frac{1}{2}\rho_{\rm s}(\boldsymbol{\nabla}\Theta)^2$, where $\rho_{\rm s}$ is the *stiffness parameter* and $\Theta(\boldsymbol{r})$ is a planar angle, with Θ and $\Theta + 2\pi n$ identified for all integers n. The model features point defects which are known as vortices, about which the winding number $\oint_{\mathcal{C}} d\boldsymbol{r}\cdot\boldsymbol{\nabla}\Theta$ along a small loop $\mathcal{C}$ enclosing the vortex is $2\pi q$, where $q \in \mathbb{Z}$ is the vorticity. Integrating the energy density for a single $q = 1$ vortex with $\Theta(\boldsymbol{r}) = \tan^{-1}(y/x)$ yields $E = \pi\rho_{\rm s}\ln(R/a)$, where R is the system radius and a an ultraviolet cutoff, which is naturally imposed in any lattice-based model. Since a vortex can be in any location, its entropy is $S = k_{\rm B}\ln(R^2/a^2)$, and we see that free vortices should proliferate when $F = E - TS = (\pi\rho_{\rm s} - 2k_{\rm B}T)\ln(R/a) < 0$, *i.e.* $T > T_{\rm c} = \pi\rho_{\rm s}/2k_{\rm B}$. This very crude derivation, which neglects interactions between the vortices, is essentially correct. The phase transition is associated with an *unbinding* of vortex-antivortex pairs. In the confined phase, where $T < T_{\rm c}$, the correlation function $C(\boldsymbol{r}, T) = \langle e^{i\Theta(\boldsymbol{r})}\, e^{-i\Theta(\boldsymbol{0})}\rangle$ decays as a power law, and there is no long-ranged order associated with a spontaneous breaking of $\mathsf{O}(2)$. In the plasma phase, where $T > T_{\rm c}$, free vortices and antivortices (*i.e.* defects with $q_i < 0$) are present, and $C(\boldsymbol{r}, T)$ decays exponentially with a finite correlation length ξ. We will discuss the

[13] The phase boundary between liquid and vapor terminates in a *critical point*, and it is therefore possible to continuously evolve from liquid to vapor without ever encountering a singularity. Thus, from the point of view of order parameters and broken symmetry, there is no essential distinction between liquid and vapor phases.

[14] V. L. Berezinskii, *Sov. Phys. JETP* **32**, 493 (1971); J. M. Kosterlitz and D. J. Thouless, *J. Phys. C: Solid State Phys.* **6**, 1181 (1973).

BKT transition and its analysis via the renormalization group later in this course.

Over the past 25 years or so, a new paradigm has emerged for certain phases of matter which have no order parameter in the conventional sense. These are called *topological phases*, and are exemplified by the phases of the quantum Hall effect. Typically topological phases exhibit a *bulk excitation gap*, meaning it requires a finite amount of energy in order to excite these systems in their bulk, and in this respect, they are akin to insulators. However, at the edges of finite systems, there exist *gapless edge states* which may carry current. The structure of the edge states is intimately related to the nature of the bulk phase, which despite being a condensate of sorts, has no local order parameter. Topological phases are of intense current interest, both theoretically and experimentally, but they remain somewhat exotic, as compared with, say, metals. Yet the metallic phase, which is an example of a *Landau Fermi liquid*, is difficult to characterize in terms of a conventional order parameter[15]. Another violation of the Landau paradigm which has been revealed in recent years is the (at this point purely theoretical) notion of *deconfined quantum criticality*. A deconfined quantum critical point violated the Landau requirement that the symmetry groups on either side of the transition have a subgroup relation.

Experimental probes

We will initially study two important phases of condensed matter in the solid state: metals and insulators[16]. Later we shall expand our horizons and consider magnets, superconductors, quantum Hall phases, and spin liquids. These various phases are revealed through various experimental probes, including *thermodynamic measurements* such as specific heat and magnetic susceptibility, *transport measurements* such as electrical resistivity and thermal conductivity, and *spectroscopies* such as Auger, photoemission and scanning probe measurements, and *optical properties* as measured in reflectivity, Kerr effect, dichroism, *etc.* The theoretical understanding of such probes typically involves the computation of correlation (or, equivalently, response) functions within the framework of quantum statistical physics.

When falsifiable theoretical models and their solutions are in harmonious agreement with experiment, and predict new observable phenomena, we can feel confident that we are indeed beginning to understand what the world is made of.

[15] The order parameter of a Fermi liquid is usually taken to be the *quasiparticle weight* Z, which, in isotropic systems, characterizes the discontinuous drop in the momentum occupation function $n(k)$ as one crosses the Fermi surface at $k = k_{\rm F}$.

[16] As we shall see, semiconductors form a sub-class of insulators.

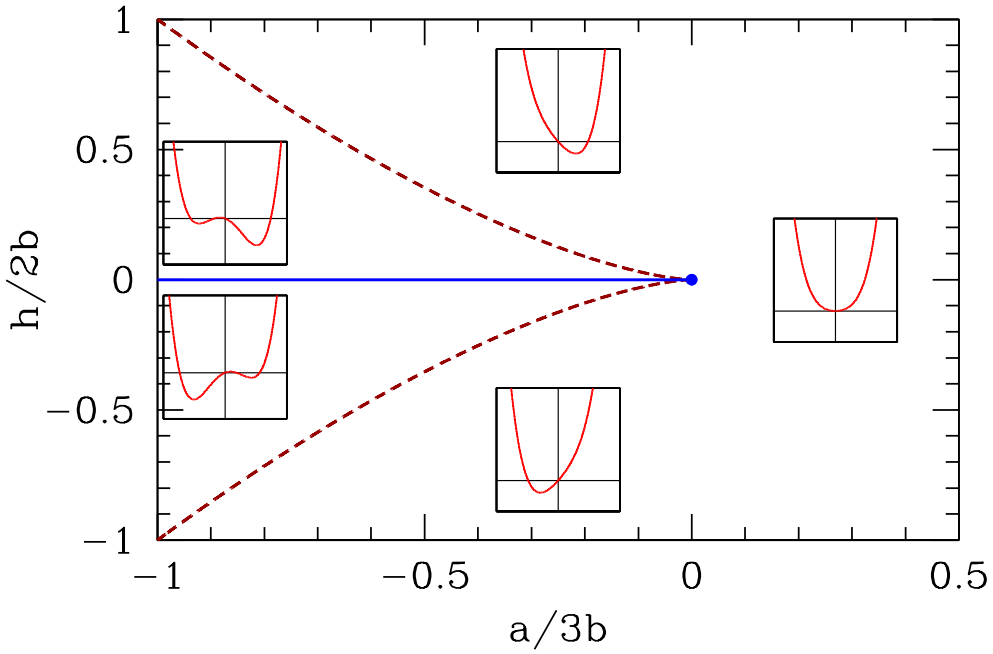


Figure 1.5: Phase diagram for the quartic Landau free energy $f = f_0 + \frac{1}{2}am^2 + \frac{1}{4}bm^4 - hm$, with $b > 0$. There is a first order line at $h = 0$ extending from $a = -\infty$ and terminating in a critical point at $a = 0$. For $|h| < h^*(a)$ (dashed red line) there are three solutions to the mean field equation, corresponding to one global minimum, one local minimum, and one local maximum. Insets show behavior of the free energy $f(m)$.

1.4 Landau Theory of Phase Transitions

Landau's theory of phase transitions is based on an expansion of the free energy of a thermodynamic system in terms of an *order parameter*, which is nonzero in an ordered phase and zero in a disordered phase. For example, the magnetization M of a ferromagnet in zero external field but at finite temperature typically vanishes for temperatures $T > T_{\rm c}$, where $T_{\rm c}$ is the *critical temperature*, also called the *Curie temperature* in a ferromagnet. A low order expansion in powers of the order parameter is appropriate sufficiently close to the phase transition, *i.e.* at temperatures such that the order parameter, if nonzero, is still small.

1.4.1 Quartic free energy with Ising symmetry

The simplest example is the quartic free energy,

$$f(m, h = 0, \theta) = f_0 + \tfrac{1}{2}am^2 + \tfrac{1}{4}bm^4 \ , \tag{1.28}$$

where $f_0 = f_0(\theta)$, $a = a(\theta)$, and $b = b(\theta)$. Here, θ is a dimensionless measure of the temperature. If for example the local exchange energy in the ferromagnet is J, then we might define $\theta =$

$k_{\rm B}T/zJ$, where z is the lattice coordination number. Let us assume $b > 0$, which is necessary if the free energy is to be bounded from below[17]. The equation of state ,

$$\frac{\partial f}{\partial m} = 0 = am + bm^3 \ , \tag{1.29}$$

has three solutions in the complex m plane: (i) $m = 0$, (ii) $m = \sqrt{-a/b}$, and (iii) $m = -\sqrt{-a/b}$. The latter two solutions lie along the (physical) real axis if $a < 0$. We assume that there exists a unique temperature $\theta_{\rm c}$ where $a(\theta_{\rm c}) = 0$. Minimizing f, we find

$$\begin{aligned} \theta < \theta_{\rm c} \quad &: \quad f(\theta) = f_0 - \frac{a^2}{4b} \\ \theta > \theta_{\rm c} \quad &: \quad f(\theta) = f_0 \ . \end{aligned} \tag{1.30}$$

The free energy is continuous at $\theta_{\rm c}$ since $a(\theta_{\rm c}) = 0$. The specific heat, however, is discontinuous across the transition, with

$$c\big(\theta_{\rm c}^+\big) - c\big(\theta_{\rm c}^-\big) = -\theta_{\rm c}\,\frac{\partial^2}{\partial\theta^2}\bigg|_{\theta=\theta_{\rm c}}\left(\frac{a^2}{4b}\right) = -\frac{\theta_{\rm c}\left[a'(\theta_{\rm c})\right]^2}{2b(\theta_{\rm c})} \ . \tag{1.31}$$

The presence of a magnetic field h breaks the $\mathbb{Z}_2$ symmetry of $m \to -m$. The free energy becomes

$$f(m, h, \theta) = f_0 + \tfrac{1}{2}am^2 + \tfrac{1}{4}bm^4 - hm \ , \tag{1.32}$$

and the mean field equation is

$$bm^3 + am - h = 0 \ . \tag{1.33}$$

This is a cubic equation for m with real coefficients, and as such it can either have three real solutions or one real solution and two complex solutions related by complex conjugation. Clearly we must have $a < 0$ in order to have three real roots, since $bm^3 + am$ is monotonically increasing otherwise. The boundary between these two classes of solution sets occurs when two roots coincide, which means $f''(m) = 0$ as well as $f'(m) = 0$. Simultaneously solving these two equations, we find

$$h^*(a) = \pm\frac{2}{3^{3/2}}\,\frac{(-a)^{3/2}}{b^{1/2}} \ , \tag{1.34}$$

or, equivalently,

$$a^*(h) = -\frac{3}{2^{2/3}}\,b^{1/3}\,|h|^{2/3}. \tag{1.35}$$

If, for fixed h, we have $a < a^*(h)$, then there will be three real solutions to the mean field equation $f'(m) = 0$, one of which is a global minimum (the one for which $m \cdot h > 0$). For

[17] It is always the case that f is bounded from below, on physical grounds. Were b negative, we'd have to consider higher order terms in the Landau expansion.

$a > a^*(h)$ there is only a single global minimum, at which m also has the same sign as h. If we solve the mean field equation perturbatively in h/a, we find

$$\begin{aligned} m(a,h) &= \frac{h}{a} - \frac{b}{a^4}\,h^3 + \mathcal{O}(h^5) \qquad\qquad (a>0) \\ &= \pm\frac{|a|^{1/2}}{b^{1/2}} + \frac{h}{2\,|a|} \pm \frac{3\,b^{1/2}}{8\,|a|^{5/2}}\,h^2 + \mathcal{O}(h^3) \qquad (a<0)\ . \end{aligned} \tag{1.36}$$

1.4.2 Cubic terms in Landau theory : first order transitions

Next, consider a free energy with a cubic term,

$$f = f_0 + \tfrac{1}{2}am^2 - \tfrac{1}{3}ym^3 + \tfrac{1}{4}bm^4\ , \tag{1.37}$$

with $b > 0$ for stability. Without loss of generality, we may assume $y > 0$ (else send $m \to -m$). Note that we no longer have $m \to -m$ (*i.e.* $\mathbb{Z}_2$) symmetry. The cubic term favors positive m. What is the phase diagram in the (a, y) plane?

Extremizing the free energy with respect to m, we obtain

$$\frac{\partial f}{\partial m} = 0 = am - ym^2 + bm^3\ . \tag{1.38}$$

This cubic equation factorizes into a linear and quadratic piece, and hence may be solved simply. The three solutions are $m = 0$ and

$$m = m_\pm \equiv \frac{y}{2b} \pm \sqrt{\Big(\frac{y}{2b}\Big)^2 - \frac{a}{b}}\ . \tag{1.39}$$

We now see that for $y^2 < 4ab$ there is only one real solution, at $m = 0$, while for $y^2 > 4ab$ there are three real solutions. Which solution has lowest free energy? To find out, we compare the energy $f(0)$ with $f(m_+)$[18]. Thus, we set

$$f(m) = f(0) \quad \Longrightarrow \quad \tfrac{1}{2}am^2 - \tfrac{1}{3}ym^3 + \tfrac{1}{4}bm^4 = 0\ , \tag{1.40}$$

and we now have two quadratic equations to solve simultaneously:

$$\begin{aligned} 0 &= a - ym + bm^2 \\ 0 &= \tfrac{1}{2}a - \tfrac{1}{3}ym + \tfrac{1}{4}bm^2 = 0\ . \end{aligned} \tag{1.41}$$

Eliminating the quadratic term gives $m = 3a/y$. Finally, substituting $m = m_+$ gives us a relation between a, b, and y:

$$y^2 = \tfrac{9}{2}\,ab\ . \tag{1.42}$$

[18] We needn't waste our time considering the $m = m_-$ solution, since the cubic term prefers positive m.

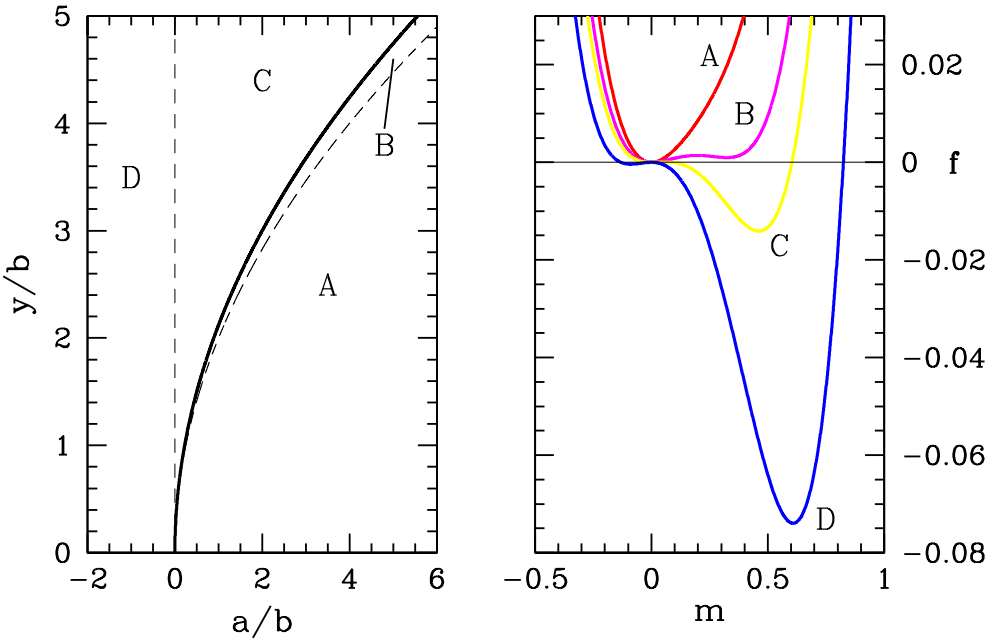


Figure 1.6: Behavior of the quartic free energy $f(m) = \frac{1}{2}am^2 - \frac{1}{3}ym^3 + \frac{1}{4}bm^4$. A: $y^2 < 4ab$; B: $4ab < y^2 < \frac{9}{2}ab$; C and D: $y^2 > \frac{9}{2}ab$. The thick black line denotes a line of first order transitions, where the order parameter is discontinuous across the transition.

Thus, we have the following:

$$\begin{aligned} a > \frac{y^2}{4b} \quad &: \quad \text{1 real root } m = 0 \\ \frac{y^2}{4b} > a > \frac{2y^2}{9b} \quad &: \quad \text{3 real roots; minimum at } m = 0 \\ \frac{2y^2}{9b} > a \quad &: \quad \text{3 real roots; minimum at } m = \frac{y}{2b} + \sqrt{\left(\frac{y}{2b}\right)^2 - \frac{a}{b}} \end{aligned} \tag{1.43}$$

The solution $m = 0$ lies at a local minimum of the free energy for $a > 0$ and at a local maximum for $a < 0$. Over the range $\frac{y^2}{4b} > a > \frac{2y^2}{9b}$, then, there is a global minimum at $m = 0$, a local minimum at $m = m_+$, and a local maximum at $m = m_-$, with $m_+ > m_- > 0$. For $\frac{2y^2}{9b} > a > 0$, there is a local minimum at $a = 0$, a global minimum at $m = m_+$, and a local maximum at $m = m_-$, again with $m_+ > m_- > 0$. For $a < 0$, there is a local maximum at $m = 0$, a local minimum at $m = m_-$, and a global minimum at $m = m_+$, with $m_+ > 0 > m_-$. See fig. 1.6.

With $y = 0$, we have a second order transition at $a = 0$. With $y \neq 0$, there is a discontinuous (first order) transition at $a_c = 2y^2/9b > 0$ and $m_c = 2y/3b$. This occurs before a reaches the value $a = 0$ where the curvature at $m = 0$ turns negative. If we write $a = \alpha(T - T_0)$, then the expected second order transition at $T = T_0$ is *preempted* by a first order transition at $T_c = T_0 + 2y^2/9\alpha b$.

1.4.3 Magnetization dynamics

Suppose we now impose some dynamics on the system, of the simple relaxational type

$$\frac{\partial m}{\partial t} = -\Gamma \, \frac{\partial f}{\partial m} \, , \tag{1.44}$$

where Γ is a phenomenological kinetic coefficient. Assuming $y > 0$ and $b > 0$, it is convenient to adimensionalize by writing

$$m \equiv \frac{y}{b} \cdot u \qquad , \qquad a \equiv \frac{y^2}{b} \cdot r \qquad , \qquad t \equiv \frac{b}{\Gamma y^2} \cdot s \, . \tag{1.45}$$

Then we obtain

$$\frac{\partial u}{\partial s} = -\frac{\partial \varphi}{\partial u} \, , \tag{1.46}$$

where the dimensionless free energy function is

$$\varphi(u) = \tfrac{1}{2} r u^2 - \tfrac{1}{3} u^3 + \tfrac{1}{4} u^4 \, . \tag{1.47}$$

We see that there is a single control parameter, r. The fixed points of the dynamics are then the stationary points of $\varphi(u)$, where $\varphi'(u) = 0$, with

$$\varphi'(u) = u \, (r - u + u^2) \, . \tag{1.48}$$

The solutions to $\varphi'(u) = 0$ are then given by

$$u^* = 0 \qquad , \qquad u^* = \tfrac{1}{2} \pm \sqrt{\tfrac{1}{4} - r} \, . \tag{1.49}$$

For $r > \frac{1}{4}$ there is one fixed point at $u = 0$, which is attractive under the dynamics $\dot{u} = -\varphi'(u)$ since $\varphi''(0) = r$. At $r = \frac{1}{4}$ there occurs a saddle-node bifurcation and a pair of fixed points is generated, one stable and one unstable. As we see from fig. 1.5, the interior fixed point is always unstable and the two exterior fixed points are always stable. At $r = 0$ there is a transcritical bifurcation where two fixed points of opposite stability collide and bounce off one another (metaphorically speaking).

At the saddle-node bifurcation, $r = \frac{1}{4}$ and $u = \frac{1}{2}$, and we find $\varphi(u = \frac{1}{2}; r = \frac{1}{4}) = \frac{1}{192}$, which is positive. Thus, the thermodynamic state of the system remains at $u = 0$ until the value of $\varphi(u_+)$ crosses zero. This occurs when $\varphi(u) = 0$ and $\varphi'(u) = 0$, the simultaneous solution of which yields $r = \frac{2}{9}$ and $u = \frac{2}{3}$.

Suppose we slowly ramp the control parameter r up and down as a function of the dimensionless time s. Under the dynamics of eqn. 1.46, $u(s)$ flows to the first stable fixed point encountered – this is always the case for a dynamical system with a one-dimensional phase space. Then as r is further varied, u follows the position of whatever locally stable fixed point

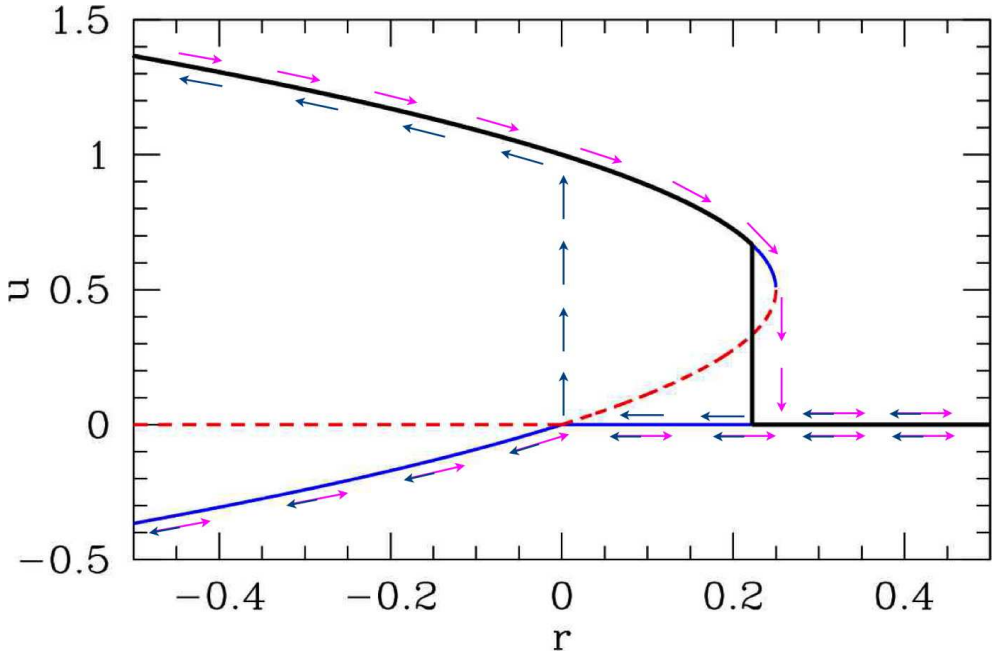


Figure 1.7: Fixed points for $\varphi(u) = \frac{1}{2}ru^2 - \frac{1}{3}u^3 + \frac{1}{4}u^4$ and flow under the dynamics $\dot{u} = -\varphi'(u)$. Solid curves represent stable fixed points and dashed curves unstable fixed points. Magenta arrows show behavior under slowly increasing control parameter r and dark blue arrows show behavior under slowly decreasing r. For $u > 0$ there is a hysteresis loop. The thick black curve shows the equilibrium thermodynamic value of $u(r)$, *i.e.* that value which minimizes the free energy $\varphi(u)$. There is a first order phase transition at $r = \frac{2}{9}$, where the thermodynamic value of u jumps from $u = 0$ to $u = \frac{2}{3}$.

it initially encountered. Thus, $u\big(r(s)\big)$ evolves smoothly until a bifurcation is encountered. The situation is depicted by the arrows in fig. 1.7. The equilibrium thermodynamic value for $u(r)$ is discontinuous; there is a first order phase transition at $r = \frac{2}{9}$, as we've already seen. As r is increased, $u(r)$ follows a trajectory indicated by the magenta arrows. For an negative initial value of u, the evolution as a function of r will be *reversible*. However, if $u(0)$ is initially positive, then the system exhibits *hysteresis*, as shown. Starting with a large positive value of r, $u(s)$ quickly evolves to $u = 0^+$, which means a positive infinitesimal value. Then as r is decreased, the system remains at $u = 0^+$ even through the first order transition, because $u = 0$ is an attractive fixed point. However, once r begins to go negative, the $u = 0$ fixed point becomes repulsive, and $u(s)$ quickly flows to the stable fixed point $u_+ = \frac{1}{2} + \sqrt{\frac{1}{4} - r}$. Further decreasing r, the system remains on this branch. If r is later increased, then $u(s)$ remains on the upper branch past $r = 0$, until the u_+ fixed point annihilates with the unstable fixed point at $u_- = \frac{1}{2} - \sqrt{\frac{1}{4} - r}$, at which time $u(s)$ quickly flows down to $u = 0^+$ again.

1.4.4 Sixth order Landau theory : tricritical point

Finally, consider a model with $\mathbb{Z}_2$ symmetry, with the Landau free energy

$$f = f_0 + \tfrac{1}{2}am^2 + \tfrac{1}{4}bm^4 + \tfrac{1}{6}cm^6 \ , \tag{1.50}$$

with $c > 0$ for stability. We seek the phase diagram in the (a, b) plane. Extremizing f with respect to m, we obtain

$$\frac{\partial f}{\partial m} = 0 = m\left(a + bm^2 + cm^4\right) , \tag{1.51}$$

which is a quintic with five solutions over the complex m plane. One solution is obviously $m = 0$. The other four are

$$m = \pm\sqrt{-\frac{b}{2c} \pm \sqrt{\left(\frac{b}{2c}\right)^2 - \frac{a}{c}}} \ . \tag{1.52}$$

For each $\pm$ symbol in the above equation, there are two options, hence four roots in all.

If $a > 0$ and $b > 0$, then four of the roots are imaginary and there is a unique minimum at $m = 0$.

For $a < 0$, there are only three solutions to $f'(m) = 0$ for real m, since the $-$ choice for the $\pm$ sign under the radical leads to imaginary roots. One of the solutions is $m = 0$. The other two are

$$m = \pm\sqrt{-\frac{b}{2c} + \sqrt{\left(\frac{b}{2c}\right)^2 - \frac{a}{c}}} \ . \tag{1.53}$$

The most interesting situation is $a > 0$ and $b < 0$. If $a > 0$ and $b < -2\sqrt{ac}$, all five roots are real. There must be three minima, separated by two local maxima. Clearly if m^* is a solution, then so is $-m^*$. Thus, the only question is whether the outer minima are of lower energy than the minimum at $m = 0$. We assess this by demanding $f(m^*) = f(0)$, where m^* is the position of the largest root (*i.e.* the rightmost minimum). This gives a second quadratic equation,

$$0 = \tfrac{1}{2}a + \tfrac{1}{4}bm^2 + \tfrac{1}{6}cm^4 \ , \tag{1.54}$$

which together with equation 1.51 gives

$$b = -\tfrac{4}{\sqrt{3}}\sqrt{ac} \ . \tag{1.55}$$

Thus, we have the following, for fixed $a > 0$:

$$\begin{aligned} b > -2\sqrt{ac} \quad &: \quad \text{1 real root } m = 0 \\ -2\sqrt{ac} > b > -\tfrac{4}{\sqrt{3}}\sqrt{ac} \quad &: \quad \text{5 real roots; minimum at } m = 0 \\ -\tfrac{4}{\sqrt{3}}\sqrt{ac} > b \quad &: \quad \text{5 real roots; minima at } m = \pm\sqrt{-\frac{b}{2c} + \sqrt{\left(\frac{b}{2c}\right)^2 - \frac{a}{c}}} \end{aligned} \tag{1.56}$$

The point $(a, b) = (0, 0)$, which lies at the confluence of a first order line and a second order line, is known as a *tricritical point*.

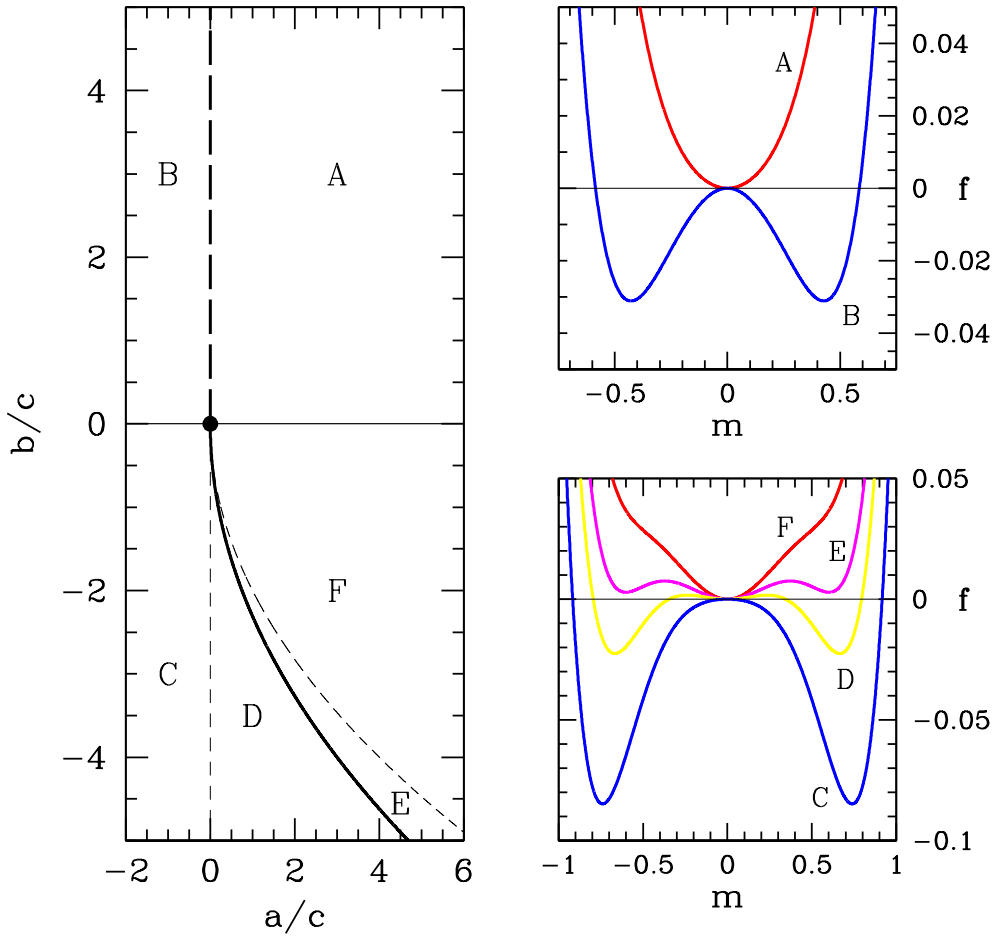


Figure 1.8: Behavior of the sextic free energy $f(m) = \frac{1}{2}am^2 + \frac{1}{4}bm^4 + \frac{1}{6}cm^6$. A: $a > 0$ and $b > 0$; B: $a < 0$ and $b > 0$; C: $a < 0$ and $b < 0$; D: $a > 0$ and $b < -\frac{4}{\sqrt{3}}\sqrt{ac}$; E: $a > 0$ and $-\frac{4}{\sqrt{3}}\sqrt{ac} < b < -2\sqrt{ac}$; F: $a > 0$ and $-2\sqrt{ac} < b < 0$. The thick dashed line is a line of second order transitions, which meets the thick solid line of first order transitions at the tricritical point, $(a, b) = (0, 0)$.

The case of barium titanate

The compound $BaTiO_3$ is a *ferroelectric* in which, at sufficiently low temperatures, a spontaneous electric polarization density P develops. The Landau free energy density may be written as

$$f(P, \varepsilon) = f_0 + \tfrac{1}{2}aP^2 + \tfrac{1}{4}bP^4 + \tfrac{1}{6}cP^6 - dEP + \varepsilon P^2 + \frac{\varepsilon^2}{2k} \quad , \tag{1.57}$$

where E is the electric field and ε is a component of the the strain field[19]. Note that the coupling between strain and polarization is linear in the former and quadratic in the latter, which is a

[19]Recall the strain tensor in a solid is given by $\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x^j} + \frac{\partial u_j}{\partial x^i}\right)$, where $\boldsymbol{u}(\boldsymbol{r})$ is the local displacement field.

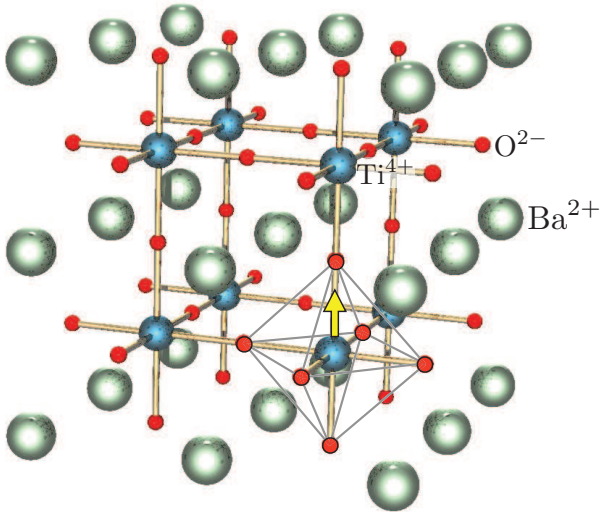


Figure 1.9: High temperature cubic perovskite crystal structure of $BaTiO_3$. Ba^{2+} sites are in green, Ti^{4+} in blue, and O^{2-} in red. The yellow arrow shows the direction in which the Ti^{4+} ion moves as the material is cooled below T_c within the displacive model. Image credit: Wikipedia.

consequence of symmetry associated with the displacive transition. In other materials such as KH_2PO_4, where the symmetry is lower, the coupling is of the form εP.

Setting $\partial f/\partial\varepsilon = 0$ we obtain $\varepsilon = -kdP^2$, resulting in the the effective free energy density

$$f_{\text{eff}}(P) = f_0 + \tfrac{1}{2}aP^2 + (\tfrac{1}{4}b - \tfrac{1}{2}kd^2)P^4 + \tfrac{1}{6}cP^6 - EP \quad . \tag{1.58}$$

If $b > 0$, and $dk^2 > \frac{1}{2}b$, the second order transition is driven to become first order due to the coupling to strain. However, it is oftentimes possible under experimental conditions to ensure that the strain is always zero, for example in the case of a thin epitaxial film whose lattice constants are perfectly matched to a substrate. In this case, stresses develop which constrain the strain to be $\varepsilon = 0$, and in the absence of an electric field E the transition is second order.

1.4.5 Hysteresis for the sextic potential

Once again, we consider the dissipative dynamics $\dot{m} = -\Gamma\, f'(m)$. We adimensionalize by writing

$$m \equiv \sqrt{\frac{|b|}{c}}\cdot u \qquad , \qquad a \equiv \frac{b^2}{c}\cdot r \qquad , \qquad t \equiv \frac{c}{\Gamma\, b^2}\cdot s\ . \tag{1.59}$$

Then we obtain once again the dimensionless equation

$$\frac{\partial u}{\partial s} = -\frac{\partial \varphi}{\partial u}\ , \tag{1.60}$$

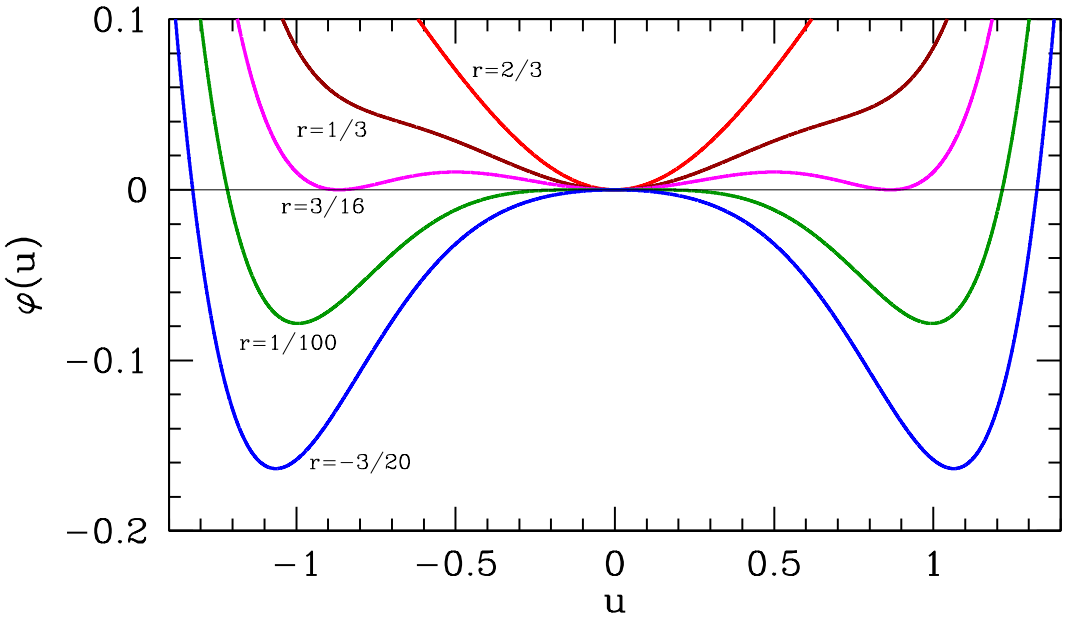


Figure 1.10: Free energy $\varphi(u) = \frac{1}{2}ru^2 - \frac{1}{4}u^4 + \frac{1}{6}u^6$ for several different values of the control parameter r.

where

$$\varphi(u) = \tfrac{1}{2}ru^2 \pm \tfrac{1}{4}u^4 + \tfrac{1}{6}u^6 \ . \tag{1.61}$$

In the above equation, the coefficient of the quartic term is positive if $b > 0$ and negative if $b < 0$. That is, the coefficient is $\mathsf{sgn}(b)$. When $b > 0$ we can ignore the sextic term for sufficiently small u, and we recover the quartic free energy studied earlier. There is then a second order transition at $r = 0$.

New and interesting behavior occurs for $b > 0$. The fixed points of the dynamics are obtained by setting $\varphi'(u) = 0$. We have

$$\begin{aligned} \varphi(u) &= \tfrac{1}{2}ru^2 - \tfrac{1}{4}u^4 + \tfrac{1}{6}u^6 \\ \varphi'(u) &= u\left(r - u^2 + u^4\right) . \end{aligned} \tag{1.62}$$

Thus, the equation $\varphi'(u) = 0$ factorizes into a linear factor u and a quartic factor $u^4 - u^2 + r$ which is quadratic in u^2. Thus, we can easily obtain the roots:

$$\begin{aligned} r < 0 \quad &: \quad u^* = 0 \ , \ u^* = \pm\sqrt{\tfrac{1}{2} + \sqrt{\tfrac{1}{4} - r}} \\ 0 < r < \tfrac{1}{4} \quad &: \quad u^* = 0 \ , \ u^* = \pm\sqrt{\tfrac{1}{2} + \sqrt{\tfrac{1}{4} - r}} \ , \ u^* = \pm\sqrt{\tfrac{1}{2} - \sqrt{\tfrac{1}{4} - r}} \\ r > \tfrac{1}{4} \quad &: \quad u^* = 0 \ . \end{aligned} \tag{1.63}$$

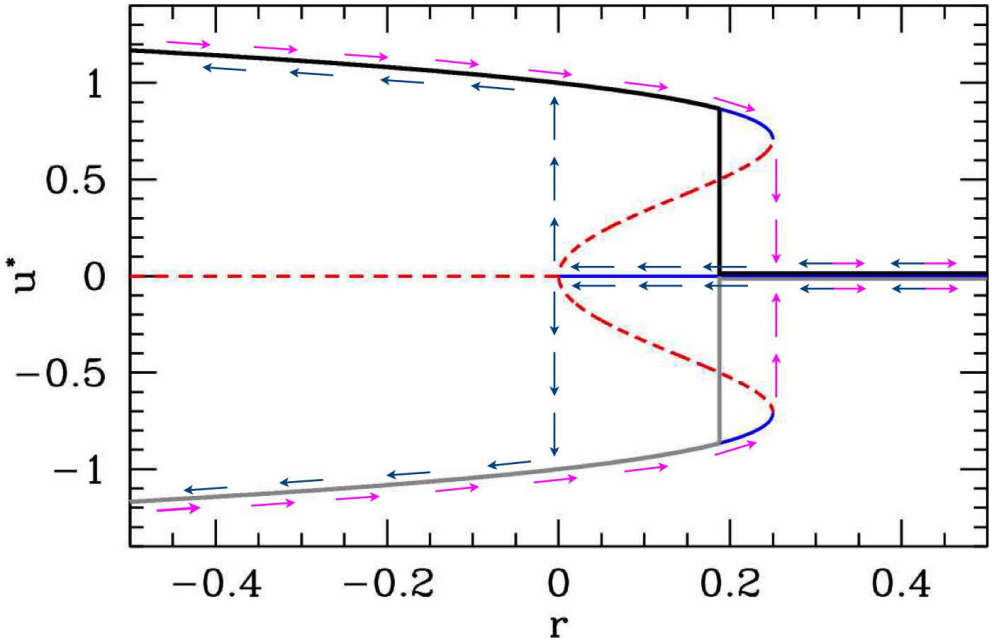


Figure 1.11: Fixed points $\varphi'(u^*) = 0$ for the sextic potential $\varphi(u) = \frac{1}{2}ru^2 - \frac{1}{4}u^4 + \frac{1}{6}u^6$, and corresponding dynamical flow (arrows) under $\dot{u} = -\varphi'(u)$. Solid curves show stable fixed points and dashed curves show unstable fixed points. The thick solid black and solid grey curves indicate the equilibrium thermodynamic values for u; note the overall $u \to -u$ symmetry. Within the region $r \in [0, \frac{1}{4}]$ the dynamics are irreversible and the system exhibits the phenomenon of hysteresis. There is a first order phase transition at $r = \frac{3}{16}$.

In fig. 1.11, we plot the fixed points and the hysteresis loops for this system. At $r = \frac{1}{4}$, there are two symmetrically located saddle-node bifurcations at $u = \pm\frac{1}{\sqrt{2}}$. We find $\varphi(u = \pm\frac{1}{\sqrt{2}}, r = \frac{1}{4}) = \frac{1}{48}$, which is positive, indicating that the stable fixed point $u^* = 0$ remains the thermodynamic minimum for the free energy $\varphi(u)$ as r is decreased through $r = \frac{1}{4}$. Setting $\varphi(u) = 0$ and $\varphi'(u) = 0$ simultaneously, we obtain $r = \frac{3}{16}$ and $u = \pm\frac{\sqrt{3}}{2}$. The thermodynamic value for u therefore jumps discontinuously from $u = 0$ to $u = \pm\frac{\sqrt{3}}{2}$ (either branch) at $r = \frac{3}{16}$; this is a first order transition.

Under the dissipative dynamics considered here, the system exhibits hysteresis, as indicated in the figure, where the arrows show the evolution of $u(s)$ for very slowly varying $r(s)$. When the control parameter r is large and positive, the flow is toward the sole fixed point at $u^* = 0$. At $r = \frac{1}{4}$, two simultaneous saddle-node bifurcations take place at $u^* = \pm\frac{1}{\sqrt{2}}$; the outer branch is stable and the inner branch unstable in both cases. At $r = 0$ there is a subcritical pitchfork bifurcation, and the fixed point at $u^* = 0$ becomes unstable.

Suppose one starts off with $r \gg \frac{1}{4}$ with some value $u > 0$. The flow $\dot{u} = -\varphi'(u)$ then rapidly results in $u \to 0^+$. This is the 'high temperature phase' in which there is no magnetization. Now let r increase slowly, using s as the dimensionless time variable. The scaled magnetization

$u(s) = u^*\big(r(s)\big)$ will remain pinned at the fixed point $u^* = 0^+$. As r passes through $r = \frac{1}{4}$, two new stable values of u^* appear, but our system remains at $u = 0^+$, since $u^* = 0$ is a stable fixed point. But after the subcritical pitchfork, $u^* = 0$ becomes unstable. The magnetization $u(s)$ then flows rapidly to the stable fixed point at $u^* = \frac{1}{\sqrt{2}}$, and follows the curve $u^*(r) = \big(\frac{1}{2}+(\frac{1}{4}-r)^{1/2}\big)^{1/2}$ for all $r < 0$.

Now suppose we start increasing r (*i.e.* increasing temperature). The magnetization follows the stable fixed point $u^*(r) = \big(\frac{1}{2} + (\frac{1}{4} - r)^{1/2}\big)^{1/2}$ past $r = 0$, beyond the first order phase transition point at $r = \frac{3}{16}$, and all the way up to $r = \frac{1}{4}$, at which point this fixed point is annihilated at a saddle-node bifurcation. The flow then rapidly takes $u \to u^* = 0^+$, where it remains as r continues to be increased further.

Within the region $r \in \big[0, \frac{1}{4}\big]$ of control parameter space, the dynamics are said to be *irreversible* and the behavior of $u(s)$ is said to be *hysteretic*.

1.4.6 Weak crystallization

That *weak crystallization*, meaning crystallization in a weakly first-order transition, should result in a triangular lattice in $d = 2$ was argued by Alexander and McTague[20] based on a Landau theory of the transition. The argument is as follows. Let $\varrho_{\boldsymbol{G}}$ be the amplitude of the Fourier component of the density $\varrho(\boldsymbol{r})$ with wavevector $\boldsymbol{G}$, which is a reciprocal lattice vector of the incipient crystalline phase. Then construct the free energy

$$F\big[\{\varrho_{\boldsymbol{G}}\}\big] = \frac{1}{2}\sum_{\boldsymbol{G}} \chi^{-1}(\boldsymbol{G})\,|\varrho_{\boldsymbol{G}}|^2 - \frac{1}{3}\,B\sum_{\boldsymbol{G}_1}\sum_{\boldsymbol{G}_2}\sum_{\boldsymbol{G}_3} \varrho_{\boldsymbol{G}_1}\,\varrho_{\boldsymbol{G}_2}\,\varrho_{\boldsymbol{G}_3}\,\delta_{\boldsymbol{G}_1+\boldsymbol{G}_2+\boldsymbol{G}_3,\boldsymbol{0}} + \frac{1}{4}\,C\sum_{\boldsymbol{G}_1}\sum_{\boldsymbol{G}_2}\sum_{\boldsymbol{G}_3}\sum_{\boldsymbol{G}_4} \varrho_{\boldsymbol{G}_1}\,\varrho_{\boldsymbol{G}_2}\,\varrho_{\boldsymbol{G}_3}\,\varrho_{\boldsymbol{G}_4}\,\delta_{\boldsymbol{G}_1+\boldsymbol{G}_2+\boldsymbol{G}_3+\boldsymbol{G}_4,\boldsymbol{0}} + \cdots \quad , \tag{1.64}$$

where

$$\chi^{-1}(\boldsymbol{k}) = r + b\,(\boldsymbol{k}^2 - G^2)^2 \tag{1.65}$$

is the inverse static susceptibility at wavevector $\boldsymbol{k}$, which for fixed r is minimized for $|\boldsymbol{k}| = G$. The quadratic term determines the magnitude of the preferred wavevectors at which condensation takes place at $r = r_{\rm c} = 0$, but this energy is degenerate over the circle (or sphere in $d = 3$) of radius G. For weak crystallization, then, the cubic term determines the crystal structure, and evidently prefers structures whose reciprocal lattices contain the maximum number of triangles, in order to satisfy the $\boldsymbol{G}_1 + \boldsymbol{G}_2 + \boldsymbol{G}_3 = 0$ condition. In $d = 2$ this prefers a reciprocal lattice which is triangular, hence the underlying direct Bravais lattice is also triangular (or honeycomb). In $d = 3$, this condition prefers the fcc structure among all regular lattices, and the

[20] S. Alexander and J. McTague, *Phys. Rev. Lett.* **41**, 702 (1978). See also E. I. Kats, V. V. Lebedev, and A. R. Muranov, *Phys. Rep.* **228**, 1 (1993).

corresponding direct lattice is thus bcc. It should be emphasized that the Alexander-McTague theory applies to the weak crystallization of a fluid, and really describes the formation of a charge density wave structure, rather than a Wigner crystal of point particles.

1.5 Four Vignettes

1.5.1 Lower critical dimension

Depending on whether the global symmetry group of a model is discrete or continuous, there exists a *lower critical dimension* d_ℓ at or below which no phase transition may take place at finite temperature. That is, for $d \leqslant d_\ell$, the critical temperature is $T_{\rm c} = 0$. Owing to its neglect of fluctuations, mean field theory generally *overestimates* the value of $T_{\rm c}$ because it overestimates the stability of the ordered phase[21]. Indeed, there are many examples where mean field theory predicts a finite $T_{\rm c}$ when the actual critical temperature is $T_{\rm c} = 0$. This happens for $d \leqslant d_\ell$.

Let's test the stability of the ordered (ferromagnetic) state of the one-dimensional Ising model at low temperatures. We consider order-destroying *domain wall* excitations which interpolate between regions of degenerate, symmetry-related ordered phase, *i.e.* $\uparrow\uparrow\uparrow\uparrow\uparrow$ and $\downarrow\downarrow\downarrow\downarrow\downarrow$. For a system with a discrete symmetry at low temperatures, the domain wall is abrupt, on the scale of a single lattice spacing. If the exchange energy is J, then the energy of a single domain wall is $2J$, since a link of energy $-J$ is replaced with one of energy $+J$. However, there are N possible locations for the domain wall, hence its entropy is $k_{\rm B} \ln N$. For a system with M domain walls, the free energy is

$$\begin{aligned} F &= 2MJ - k_{\rm B}T \ln \binom{N}{M} \\ &= N \cdot \left\{ 2Jx + k_{\rm B}T \Big[x \ln x + (1-x) \ln(1-x) \Big] \right\} , \end{aligned} \tag{1.66}$$

[21] One can concoct models for which the mean field transition temperature *underestimates* the actual critical temperature. Consider for example an Ising model with interaction $u(\sigma,\sigma') = -\epsilon^{-1} \ln(1 + \epsilon\sigma\sigma')$, where the spins take values $\sigma, \sigma' = \pm 1$, and where $0 < \epsilon < 1$. If we write $\sigma = \langle\sigma\rangle + \delta\sigma$ at each site and neglect terms quadratic in fluctuations, the resulting mean field Hamiltonian is equivalent to a set of decoupled spins in an external field $h = zm/(1+\epsilon m^2)$, where $m = \langle\sigma\rangle$. From the mean field equation $m = \tanh(h/T)$, one obtains the MF transition temperature is $T_{\rm c}^{\rm MF} = z$, the lattice coordination number, independent of ϵ. On the other hand, we may also write $u(\sigma,\sigma') = u_\epsilon - J_\epsilon\,\sigma\sigma'$, where $u_\epsilon = -\ln(1-\epsilon^2)/2\epsilon$ and $J_\epsilon = \epsilon^{-1}\tanh^{-1}(\epsilon)$. On the square lattice, where $z = 4$, one has the exact result $T_{\rm c}(\epsilon) = 2J_\epsilon/\sinh^{-1}(1) = 2.269\,J_\epsilon$, which diverges as $\epsilon \to 1$, while $T_{\rm c}^{\rm MF} = 4$ remains finite. For $\epsilon > 0.9265$, one has $T_{\rm c}(\epsilon) > T_{\rm c}^{\rm MF}$. Finally, if instead of deriving the MFT via the 'neglect of fluctuations' method one uses a local variational density matrix of the form $\varrho = \prod_i \Big[\frac{1}{2}(1+m_i)\,\delta_{\sigma_i,+1} + \frac{1}{2}(1-m_i)\,\delta_{\sigma_i,-1} \Big]$, with m as a variational parameter, from the variational free energy density $f = N^{-1}\big[\,{\rm Tr}(\varrho H) + T\,{\rm Tr}(\varrho\ln\varrho)\big]$, one obtains the MF equation $zJ_\epsilon m = T\tanh^{-1}(m)$, whence $T_{\rm c}^{\rm MF} = zJ_\epsilon$ which *does* exceed the exact result on the square lattice. So in this example, it depends on the method employed to derive the MF Hamiltonian.

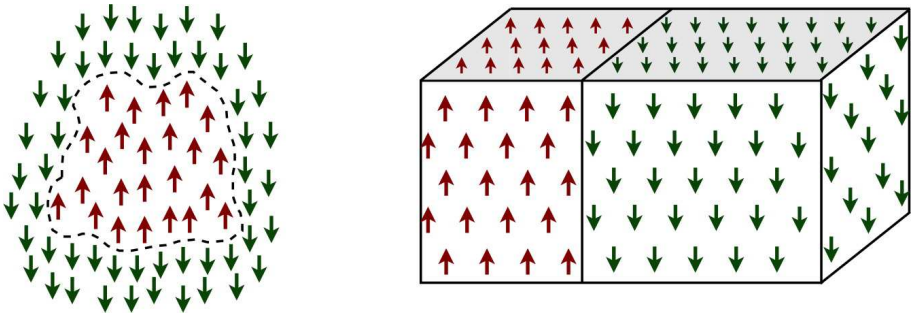


Figure 1.12: Domain walls in the two-dimensional (left) and three-dimensional (right) Ising model.

where $x = M/N$ is the density of domain walls, and where we have used Stirling's approximation for $k!$ when k is large. Extremizing with respect to x, we find

$$\frac{x}{1-x} = e^{-2J/k_{\rm B}T} \qquad \Longrightarrow \qquad x = \frac{1}{e^{2J/k_{\rm B}T}+1}\,. \tag{1.67}$$

The average distance between domain walls is x^{-1}, which is finite for finite T. Thus, the thermodynamic state of the system is *disordered*, with no net average magnetization.

Consider next an Ising domain wall in d dimensions. Let the linear dimension of the system be $L \cdot a$, where L is a real number and a is the lattice constant. Then the energy of a single domain wall which partitions the entire system is $2J \cdot L^{d-1}$. The domain wall entropy is difficult to compute, because the wall can fluctuate significantly, but for a single domain wall we have $S \gtrsim k_{\rm B} \ln L$. Thus, the free energy $F = 2JL^{d-1} - k_{\rm B}T \ln L$ is dominated by the energy term if $d > 1$, suggesting that the system *may* be ordered. We can do a slightly better job in $d = 2$ by writing

$$Z \approx \exp\left(L^d \sum_P N_P \, e^{-2PJ/k_{\rm B}T} \right)\,, \tag{1.68}$$

where the sum is over all closd loops of perimeter P, and N_P is the number of such loops. An example of such a loop circumscribing a domain is depicted in the left panel of fig. 1.12. It turns out that

$$N_P \simeq \kappa^P P^{-\theta} \cdot \Big\{ 1 + \mathcal{O}(P^{-1}) \Big\}\,, \tag{1.69}$$

where $\kappa = z - 1$ with z the lattice coordination number, and θ is some exponent. We can understand the κ^P factor in the following way. At each step along the perimeter of the loop, there are $\kappa = z-1$ possible directions to go (since one doesn't backtrack). The fact that the loop must avoid overlapping itself and must return to its original position to be closed leads to the power law term $P^{-\theta}$, which is subleading since $\kappa^P P^{-\theta} = \exp(P \ln \kappa - \theta \ln P)$ and $P \gg \ln P$ for $P \gg 1$. Thus,

$$F \approx -\frac{1}{\beta} L^d \sum_P P^{-\theta} \, e^{(\ln\kappa - 2\beta J)P}\,, \tag{1.70}$$

which diverges if $\ln\kappa > 2\beta J$, *i.e.* if $T > 2J/k_{\rm B}\ln(z-1)$. We identify this singularity with the phase transition. The high temperature phase involves a proliferation of such loops. The excluded volume effects between the loops, which we have not taken into account, then enter in an essential way so that the sum converges. Thus, we have the following picture:

$$\begin{aligned}\ln\kappa < 2\beta J \;&:\; \text{large loops suppressed ; ordered phase}\\ \ln\kappa > 2\beta J \;&:\; \text{large loops proliferate ; disordered phase .}\end{aligned}$$

On the square lattice, we obtain

$$k_{\rm B}T_{\rm c}^{\rm approx} = \frac{2J}{\ln 3} = 1.82\,J$$
$$k_{\rm B}T_{\rm c}^{\rm exact} = \frac{2J}{\sinh^{-1}(1)} = 2.27\,J\ .$$

The agreement is better than we should reasonably expect from such a crude argument.

Nota bene : Beware of arguments which allegedly prove the existence of an ordered phase. Generally speaking, any approximation will *underestimate* the entropy, and thus will overestimate the stability of the putative ordered phase.

Continuous symmetries

When the global symmetry group is continuous, the domain walls interpolate smoothly between ordered phases. Consider the classical continuum $\mathsf{O}(N)$ model

$$H = \tfrac{1}{2}\rho_{\rm s}\int d^dr\,(\partial_\mu n^a)^2\ , \tag{1.71}$$

where $\hat{\boldsymbol{n}}(\boldsymbol{x}) = (n^1,\ldots,n^N)$ with $\hat{\boldsymbol{n}}^2 = 1$. The quantity $\rho_{\rm s}$ is called the *stiffness parameter* and has dimensions of $[\rho_{\rm s}] = \mathsf{EL}^{2-d}$. Any ground state configuration, such as $\hat{\boldsymbol{n}}(\boldsymbol{x}) = \hat{\mathrm{e}}_1$, breaks the $\mathsf{O}(N)$ symmetry.

Consider now a domain wall configuration

$$\hat{\boldsymbol{n}}(\boldsymbol{x}) = \cos\big(\theta(\boldsymbol{x})\big)\,\hat{\mathrm{e}}_1 + \sin\big(\theta(\boldsymbol{x})\big)\,\hat{\mathrm{e}}_2 \quad , \tag{1.72}$$

where $\theta(\boldsymbol{x}) = 2\pi q x^1/L$, which describes a slow q-fold ($q\in\mathbb{Z}$) twist of the unit vector $\hat{\boldsymbol{n}}(\boldsymbol{x})$ in the $(\hat{\mathrm{e}}_1,\hat{\mathrm{e}}_2)$ plane with period L, which we take to be the linear dimension of the sample[22]. The domain wall then resembles the sketch in Fig. 1.13, and its energy is computed to be

$$E = \tfrac{1}{2}\rho_{\rm s}\,L^{d-1}\int_0^L dx^1\left(\frac{2\pi q}{L}\right)^2 = 2\pi^2q^2\rho_{\rm s}\,L^{d-2}\ . \tag{1.73}$$

[22]Periodic boundary conditions are presumed.

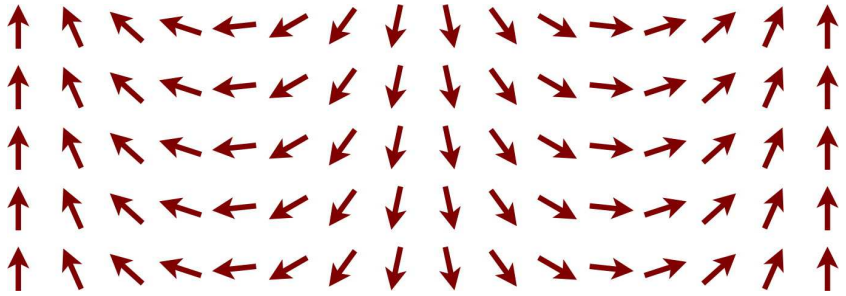


Figure 1.13: A domain wall in an XY ferromagnet.

Recall that in the case of discrete symmetry, the domain wall energy scaled as $E \propto L^{d-1}$. Thus, with $S \gtrsim k_{\rm B} \ln L$ for a single wall, we see that the entropy term dominates if $d \leqslant 2$, in which case there is no finite temperature phase transition. Thus, the lower critical dimension d_ℓ depends on whether the global symmetry is discrete or continuous, with

$$\begin{aligned} \text{discrete global symmetry} &\implies d_\ell = 1 \\ \text{continuous global symmetry} &\implies d_\ell = 2\,. \end{aligned}$$

Note that all along we have assumed local, *short-ranged* interactions. Long-ranged interactions can enhance order and thereby suppress d_ℓ.

Thus, we expect that for models with discrete symmetries, $d_\ell = 1$ and there is no finite temperature phase transition for $d \leqslant 1$. For models with continuous symmetries, $d_\ell = 2$, and we expect $T_{\rm c} = 0$ for $d \leqslant 2$. In this context we should emphasize that the two-dimensional XY model *does* exhibit a phase transition at finite temperature, called the *Kosterlitz-Thouless* transition. However, this phase transition is *not* associated with the breaking of the continuous global $\mathsf{O}(2)$ symmetry and rather has to do with the unbinding of vortices and antivortices. So there is still no true long-ranged order below the critical temperature $T_{\rm KT}$, even though there is a phase transition!

1.5.2 Random systems : Imry-Ma argument

In condensed matter systems, intrinsic randomness often exists due to quenched impurities, grain boundaries, immobile vacancies, *etc.* How does this quenched randomness affect a system's attempt to order at $T = 0$? This question was taken up in a beautiful and brief paper by J. Imry and S.-K. Ma, *Phys. Rev. Lett.* **35**, 1399 (1975). Imry and Ma considered models in which there are short-ranged interactions and a random local field coupling to the local order

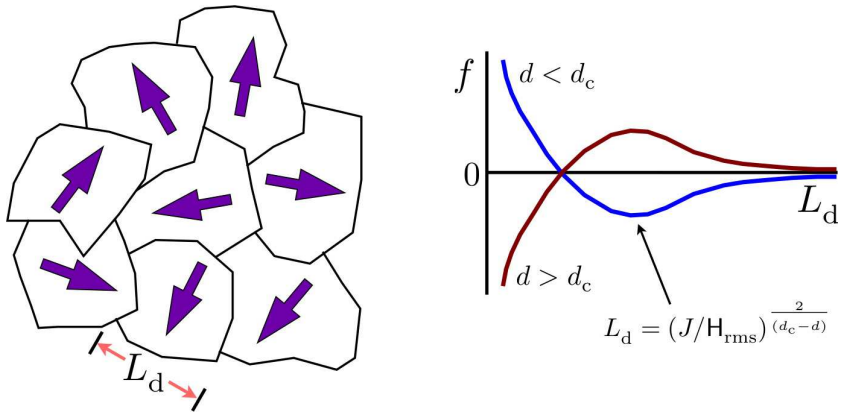


Figure 1.14: Left panel : Imry-Ma domains for an $\mathsf{O}(2)$ model. The arrows point in the direction of the local order parameter field $\langle \hat{\boldsymbol{n}}(\boldsymbol{r}) \rangle$. Right panel : free energy density as a function of domain size $L_{\rm d}$. Keep in mind that the minimum possible value for $L_{\rm d}$ is the lattice spacing a.

parameter:

$$H_{\rm RFI} = -J \sum_{\langle ij \rangle} \sigma_i \, \sigma_j - \sum_i h_i \, \sigma_i \tag{1.74}$$

$$H_{\rm RFO(}n{\rm)} = -J \sum_{\langle ij \rangle} \hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{n}}_j - \sum_i h_i^\alpha \, n_i^\alpha \ , \tag{1.75}$$

where

$$\langle\!\langle \, h_i^\alpha \, \rangle\!\rangle = 0 \qquad , \qquad \langle\!\langle \, h_i^\alpha \, h_j^\beta \, \rangle\!\rangle = \Gamma \, \delta^{\alpha\beta} \, \delta_{ij} \ , \tag{1.76}$$

where $\langle\!\langle \, \cdot \, \rangle\!\rangle$ denotes a configurational average over the disorder. Imry and Ma reasoned that a system could try to lower its free energy by forming *domains* in which the order parameter takes advantage of local fluctuations in the random field. The size of these domains is assumed to be $L_{\rm d}$, a length scale to be determined. See the sketch in the left panel of fig. 1.14.

There are two contributions to the energy of a given domain: bulk and surface terms. The bulk energy is

$$E_{\rm bulk} = -h_{\rm rms} \, (L_{\rm d}/a)^{d/2} \ , \tag{1.77}$$

where a is the lattice spacing. This is because when we add together $(L_{\rm d}/a)^d$ random fields, the magnitude of the result is proportional to the square root of the number of terms, *i.e.* to $(L_{\rm d}/a)^{d/2}$. The quantity $h_{\rm rms} = \sqrt{\Gamma}$ is the root-mean-square fluctuation in the random field at a given site. The surface energy is

$$E_{\rm surface} \propto \begin{cases} J \, (L_{\rm d}/a)^{d-1} & \text{(discrete symmetry)} \\ J \, (L_{\rm d}/a)^{d-2} & \text{(continuous symmetry)} \ . \end{cases} \tag{1.78}$$

We compute the critical dimension $d_{\rm c}$ by balancing the bulk and surface energies,

$$\begin{aligned} d-1 &= \tfrac{1}{2}d \qquad \Longrightarrow \qquad d_{\rm c} = 2 \qquad \text{(discrete)} \\ d-2 &= \tfrac{1}{2}d \qquad \Longrightarrow \qquad d_{\rm c} = 4 \qquad \text{(continuous)} \ . \end{aligned}$$

The total free energy is $F = (V/L_{\rm d}^d) \cdot \Delta E$, where $\Delta E = E_{\rm bulk} + E_{\rm surf}$. Thus, the free energy per unit cell is

$$f = \frac{F}{V/a^d} \approx J \left(\frac{a}{L_{\rm d}}\right)^{\frac{1}{2}d_{\rm c}} - h_{\rm rms} \left(\frac{a}{L_{\rm d}}\right)^{\frac{1}{2}d} . \tag{1.79}$$

If $d < d_{\rm c}$, the surface term dominates for small $L_{\rm d}$ and the bulk term dominates for large $L_{\rm d}$ There is global minimum at

$$\frac{L_{\rm d}}{a} = \left(\frac{d_{\rm c}}{d} \cdot \frac{J}{h_{\rm rms}}\right)^{\frac{2}{d_{\rm c}-d}} . \tag{1.80}$$

For $d > d_{\rm c}$, the relative dominance of the bulk and surface terms is reversed, and there is a global maximum at this value of $L_{\rm d}$.

Sketches of the free energy $f(L_{\rm d})$ in both cases are provided in the right panel of fig. 1.14. We must keep in mind that the domain size $L_{\rm d}$ cannot become smaller than the lattice spacing a. Hence we should draw a vertical line on the graph at $L_d = a$ and discard the portion $L_d < a$ as unphysical. For $d < d_{\rm c}$, we see that the state with $L_d = \infty$, *i.e.* the ordered state, is never the state of lowest free energy. *In dimensions $d < d_{\rm c}$, the ordered state is always unstable to domain formation in the presence of a random field.*

For $d > d_{\rm c}$, there are two possibilities, depending on the relative size of J and $h_{\rm rms}$. We can see this by evaluating $f(L_d = a) = J - h_{\rm rms}$ and $f(L_{\rm d} = \infty) = 0$. Thus, if $J > h_{\rm rms}$, the minimum energy state occurs for $L_{\rm d} = \infty$. In this case, the system has an ordered ground state, and we expect a finite temperature transition to a disordered state at some critical temperature $T_{\rm c} > 0$. If, on the other hand, $J < h_{\rm rms}$, then the fluctuations in h overwhelm the exchange energy at $T = 0$, and the ground state is disordered down to the very smallest length scale (*i.e.* the lattice spacing a).

Please read the essay, Memories of Shang-Keng Ma.

1.5.3 Hohenberg-Mermin-Wagner theorem

The Hohenberg-Mermin-Wagner (HMW) theorem is a powerful result which establishes the absence of broken continuous symmetry for systems at any finite temperature in dimensions $d \leqslant 2$. Extensions of the theorem have been applied to certain $d = 1$ quantum systems at $T = 0$, such as antiferromagnets, crystals, and Bose superfluids. As we shall see, it is related to Goldstone's theorem, which we shall discuss below in §1.5.4.

Before getting into the details, a plea for proper attribution. Historically, the HMW theorem has often been referred to as the "Mermin-Wagner theorem", but in fact it should rightly be

called the "Hohenberg-Mermin-Wagner theorem"[23]. For a latter day perspective, including a brief recounting of the history, see B. I. Halperin, *J. Stat. Phys.* **175**, 521 (2019).

For a poor man's derivation of the HMW theorem, we proffer the following argument. Expanding in quadratic fluctuations for a translationally invariant quantum system in the thermodynamic limit, consider the model Hamiltonian

$$H = \sum_{\boldsymbol{k}} \overbrace{\left\{ \frac{p_{\boldsymbol{k}}\, p_{-\boldsymbol{k}}}{2m_{\boldsymbol{k}}} + \tfrac{1}{2} m_{\boldsymbol{k}} \omega_{\boldsymbol{k}}^2\, x_{\boldsymbol{k}}\, x_{-\boldsymbol{k}} \right\}}^{H_{\boldsymbol{k}}} = \sum_{\boldsymbol{k}} (a_{\boldsymbol{k}}^\dagger a_{\boldsymbol{k}} + \tfrac{1}{2})\, \hbar\omega_{\boldsymbol{k}} \quad , \tag{1.81}$$

where $\left[p_{\boldsymbol{k}}, x_{\boldsymbol{k}'}\right] = -i\hbar\delta_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{0}}$ and $\left[a_{\boldsymbol{k}}, a_{\boldsymbol{k}'}^\dagger\right] = \delta_{\boldsymbol{k},\boldsymbol{k}'}$. The average energy per $\boldsymbol{k}$ mode is given by $\langle H_{\boldsymbol{k}} \rangle = (n_{\boldsymbol{k}} + \frac{1}{2})\hbar\omega_{\boldsymbol{k}}$, with $n_{\boldsymbol{k}} = 1/\big(\exp(\beta\hbar\omega_{\boldsymbol{k}}) - 1\big)$ the mode occupancy. Now let's compute the fluctuations in the local coordinate $x_{\boldsymbol{r}}$:

$$\langle x_{\boldsymbol{r}}^2 \rangle = \frac{1}{N} \sum_{\boldsymbol{k}} \langle x_{\boldsymbol{k}} x_{-\boldsymbol{k}} \rangle = \frac{1}{N} \sum_{\boldsymbol{k}} \frac{(n_{\boldsymbol{k}} + \frac{1}{2})\hbar\omega_{\boldsymbol{k}}}{m_{\boldsymbol{k}}\omega_{\boldsymbol{k}}^2} = \mathsf{vol}(\Omega) \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^2} \left(\frac{1}{\exp(\beta\hbar\omega_{\boldsymbol{k}}) - 1} + \frac{1}{2} \right) \frac{\hbar}{m_{\boldsymbol{k}}\omega_{\boldsymbol{k}}} \quad , \tag{1.82}$$

where N is the number of unit cells in the system, Ω is the unit cell, and $\hat{\Omega}$ is the Brillouin zone.

We presume the dispersion $\omega_{\boldsymbol{k}}$ behaves as $\omega_{\boldsymbol{k}} = c|\boldsymbol{k}|$ as $\boldsymbol{k} \to \boldsymbol{0}$, and further assume that $m_{\boldsymbol{k}\to 0}$ is finite. At finite temperature, the integrand therefore diverges as $T/\omega_{\boldsymbol{k}}^2 \sim T/\boldsymbol{k}^2$ as $\boldsymbol{k} \to \boldsymbol{0}$, and thus the integral is IR-divergent for $d \leqslant 2$. This means that we have *expanded about the wrong vacuum*. Note that the same considerations apply at $T = 0$ in $d = 1$, where $n_{\boldsymbol{k}} = 0$, due to the intrinsic quantum fluctuations.

Formal proof of the HMW theorem

We follow here the original Mermin-Wagner treatment of the HMW theorem. Consider the Heisenberg model,

$$H = -\sum_{\boldsymbol{R},\boldsymbol{R}'} J(\boldsymbol{R} - \boldsymbol{R}')\, \boldsymbol{S}_{\boldsymbol{R}} \cdot \boldsymbol{S}_{\boldsymbol{R}'} - \sum_{\boldsymbol{R}} h_{\boldsymbol{R}}\, S_{\boldsymbol{R}}^z \quad . \tag{1.83}$$

Here the positions $\{\boldsymbol{R}\}$ are sites on some Bravais lattice, the spin operators obey the $\mathsf{SU}(2)$ algebra[24]

$$\left[S_{\boldsymbol{R}}^\alpha, S_{\boldsymbol{R}'}^\beta\right] = i\, \delta_{\boldsymbol{R}\boldsymbol{R}'}\, \epsilon^{\alpha\beta\gamma}\, S_{\boldsymbol{R}}^\gamma \quad , \tag{1.84}$$

and the coupling $J(\boldsymbol{R} - \boldsymbol{R}') = J(\boldsymbol{R}' - \boldsymbol{R})$ is inversion symmetric, and satisfies $J(\boldsymbol{0}) = 0$. The Fourier components $\hat{h}_{\boldsymbol{k}}$ of the magnetic field satisfy $\hat{h}_{-\boldsymbol{k}} = \hat{h}_{\boldsymbol{k}}^*$, since the Hamiltonian must be

[23] P. C. Hohenberg, *Phys. Rev.* **158**, 383 (1967); N. D. Mermin and H. Wagner, *Phys. Rev. Lett.* **17**, 1133 (1966).
[24] We set $\hbar = 1$ for convenience.

Hermitian. We will also need to require that the interaction kernel $J(\boldsymbol{R}-\boldsymbol{R}')$ be sufficiently short-ranged, decaying faster than $|\boldsymbol{R}-\boldsymbol{R}'|^{-(d+2)}$, as shall become clear below.

The proof proceeds by finding a bound for the magnitude $|\hat{m}_{\boldsymbol{k}}|$, where $\hat{m}_{\boldsymbol{k}} \equiv \langle S^z_{\boldsymbol{k}}\rangle/N$, with

$$S^z_{\boldsymbol{k}} = \sum_{\boldsymbol{R}} S^z_{\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{1.85}$$

The quantity $m_{\boldsymbol{k}}$ is the order parameter at wavevector $\boldsymbol{k}$. For example, at $\boldsymbol{k}=0$ we have $\hat{m}_{\boldsymbol{0}} = N^{-1}\sum_{\boldsymbol{R}}\langle S^z_{\boldsymbol{R}}\rangle$, which is the moment per site. The goal is to show that in the limit $\hat{h}_{\boldsymbol{k}} \to 0$ of vanishing external field that the order parameter must vanish if the spatial dimension d is sufficiently small.

Bogoliubov's inequality

The proof utilizes the Bogoliubov inequality,

$$\big|\big\langle [A,C]\big\rangle\big|^2 \leqslant \tfrac{1}{2}\beta\,\big\langle\{A,A^\dagger\}\big\rangle\,\big\langle\big[C^\dagger,[H,C]\big]\big\rangle \quad , \tag{1.86}$$

for any operators A and C, where $\big\langle\mathcal{O}\big\rangle = \mathrm{Tr}\,\big(\mathcal{O}\,e^{-\beta H}\big)/\,\mathrm{Tr}\,e^{-\beta H}$ is the thermodynamic expectation value of the operator $\mathcal{O}$. To prove Eqn. 1.86, we start by defining the operator scalar product,

$$(A,B) \equiv \sum_{m\neq n} \langle\, n\,|\,A^\dagger\,|\,m\,\rangle\langle\, m\,|\,B\,|\,n\,\rangle\left(\frac{W_m - W_n}{E_n - E_m}\right) \quad , \tag{1.87}$$

where $W_m = e^{-\beta E_m}/\,\mathrm{Tr}\,e^{-\beta H}$ is the Boltzmann weight for the state $|m\rangle$. One can check that $(\bullet,\bullet)$ satisfies (i) $(A,B) = (B,A)^*$ (conjugation symmetry), (ii) $(A,\lambda_1 B_1 + \lambda_2 B_2) = \lambda_1(A,B_1) + \lambda_2(A,B_2)$ (linearity), and (iii) $(A,A) \geqslant 0$ (positive semidefiniteness). Under these conditions, the Schwarz inequality,

$$\big|(A,B)\big|^2 \leqslant (A,A)\,(B,B) \quad , \tag{1.88}$$

holds.

Let's evaluate the Schwarz inequality with $B = [C^\dagger, H]$. We have

$$\begin{aligned}(A,B) = \big(A,[C^\dagger,H]\big) &= \sum_{m\neq n}\langle\, n\,|\,A^\dagger\,|\,m\,\rangle\langle\, m\,|\,(C^\dagger H - HC^\dagger)\,|\,n\,\rangle\left(\frac{W_m - W_n}{E_n - E_m}\right) \\ &= \sum_{m,n}\langle\, n\,|\,A^\dagger\,|\,m\,\rangle\langle\, m\,|\,C^\dagger\,|\,n\,\rangle\,(W_m - W_n) = \big\langle[C^\dagger,A^\dagger]\big\rangle \quad .\end{aligned} \tag{1.89}$$

Substituting $A = B$ in this result, we have

$$(B,B) = \big\langle[C^\dagger,B^\dagger]\big\rangle = \big\langle\big[C^\dagger,[H,C]\big]\big\rangle \quad , \tag{1.90}$$

and thus the Schwarz inequality guarantees

$$\big|\big\langle [C^\dagger, A^\dagger]\big\rangle\big|^2 = \big|\big\langle [A, C]\big\rangle\big|^2 \leqslant (A, A)\,\big\langle \big[C^\dagger, [H, C]\big]\big\rangle \quad . \tag{1.91}$$

Next, note that

$$\begin{aligned}\frac{W_m - W_n}{E_n - E_m} &= \left(\frac{W_m + W_n}{E_n - E_m}\right)\left(\frac{W_m - W_n}{W_m + W_n}\right) \\ &= \left(\frac{W_m + W_n}{E_n - E_m}\right)\tanh\big(\tfrac{1}{2}\beta(E_n - E_m)\big) \leqslant \frac{\beta}{2}\,(W_m + W_n) \quad ,\end{aligned} \tag{1.92}$$

since $0 \leqslant x^{-1}\tanh x \leqslant 1$. Thus we have

$$(A, A) \leqslant \frac{\beta}{2}\sum_{m\neq n}\langle\, n\,|\,A^\dagger\,|\,m\,\rangle\langle\, m\,|\,A\,|\,n\,\rangle\,(W_m + W_n) \leqslant \frac{\beta}{2}\,\big\langle\{A, A^\dagger\}\big\rangle \quad , \tag{1.93}$$

thereby establishing the Bogoliubov inequality of Eqn. 1.86.

Application to quantum Heisenberg model

Consider now the quantum Heisenberg model of Eqn. 1.83. Define the Fourier variables

$$S^\alpha_{\boldsymbol{k}} = \sum_{\boldsymbol{R}} S^\alpha_{\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad \hat{J}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} J(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad \hat{h}_{\boldsymbol{k}} = \frac{1}{N}\sum_{\boldsymbol{R}} h_{\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \tag{1.94}$$

and their inverses

$$S^\alpha_{\boldsymbol{R}} = \frac{1}{N}\sum_{\boldsymbol{k}} S^\alpha_{\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad J(\boldsymbol{R}) = \frac{1}{N}\sum_{\boldsymbol{k}} \hat{J}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad\qquad h_{\boldsymbol{R}} = \sum_{\boldsymbol{k}} h_{\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{1.95}$$

Note the placement of the $1/N$ factors differs in the definitions of $\hat{J}(\boldsymbol{k})$ and $\hat{h}_{\boldsymbol{k}}$. The commutation relations among the Fourier spin components are

$$[S^+_{\boldsymbol{k}}\,,\, S^-_{\boldsymbol{k}'}] = 2S^z_{\boldsymbol{k}+\boldsymbol{k}'} \qquad , \qquad [S^z_{\boldsymbol{k}}\,,\, S^\pm_{\boldsymbol{k}'}] = \pm S^\pm_{\boldsymbol{k}+\boldsymbol{k}'} \quad . \tag{1.96}$$

We apply the Bogoliubov inequality of Eqn. 1.86 with

$$A = S^-_{-\boldsymbol{k}_1} \quad , \quad A^\dagger = S^+_{\boldsymbol{k}_1} \quad , \quad C = S^+_{\boldsymbol{k}_2} \quad , \quad C^\dagger = S^-_{-\boldsymbol{k}_2} \quad . \tag{1.97}$$

Thus, our version of the inequality may be written

$$\big|\big\langle [S^-_{-\boldsymbol{k}_1}, S^+_{\boldsymbol{k}_2}]\big\rangle\big|^2 \leqslant \tfrac{1}{2}\beta\,\big\langle\{S^-_{-\boldsymbol{k}_1}, S^+_{\boldsymbol{k}_1}\}\big\rangle\,\big\langle\big[S^-_{-\boldsymbol{k}_2}, [H, S^+_{\boldsymbol{k}_2}]\big]\big\rangle \quad , \tag{1.98}$$

Note that we may write $H = H_0 + H_1$ where

$$\begin{aligned} H_0 &= -\frac{1}{2N}\sum_{\boldsymbol{q}} \hat{J}(\boldsymbol{q}) \left\{ \tfrac{1}{2} S^+_{\boldsymbol{q}}\, S^-_{-\boldsymbol{q}} + \tfrac{1}{2} S^-_{\boldsymbol{q}}\, S^+_{-\boldsymbol{q}} + S^z_{\boldsymbol{q}}\, S^z_{-\boldsymbol{q}} \right\} \\ H_1 &= -\sum_{\boldsymbol{q}} \hat{h}_{-\boldsymbol{q}}\, S^z_{\boldsymbol{q}} \quad . \end{aligned} \tag{1.99}$$

We begin by computing the commutator

$$[H_0, S^+_{\boldsymbol{k}_2}] = \frac{1}{2N}\sum_{\boldsymbol{q}} \hat{J}(\boldsymbol{q}) \left\{ S^+_{\boldsymbol{q}}\, S^z_{\boldsymbol{k}_2-\boldsymbol{q}} + S^z_{\boldsymbol{k}_2+\boldsymbol{q}}\, S^+_{-\boldsymbol{q}} - S^+_{\boldsymbol{k}_2+\boldsymbol{q}}\, S^z_{-\boldsymbol{q}} - S^z_{\boldsymbol{q}}\, S^+_{\boldsymbol{k}_2-\boldsymbol{q}} \right\} \quad , \tag{1.100}$$

from which we obtain, with a little work,

$$\left[S^-_{-\boldsymbol{k}_2}, [H_0, S^+_{\boldsymbol{k}_2}]\right] = \frac{1}{2N}\sum_{\boldsymbol{q}} \Big(\hat{J}(\boldsymbol{q}) - \hat{J}(\boldsymbol{q}+\boldsymbol{k}_2) \Big) \Big(S^+_{\boldsymbol{q}}\, S^-_{-\boldsymbol{q}} + S^-_{\boldsymbol{q}}\, S^+_{-\boldsymbol{q}} + 4\, S^z_{\boldsymbol{q}}\, S^z_{-\boldsymbol{q}} \Big) \tag{1.101}$$

and where we have invoked $\sum_{\boldsymbol{k}} \hat{J}(\boldsymbol{k}) = 0$, which follows from the condition $J(\boldsymbol{0}) = 0$.

Second we compute $[H_1, S^+_{\boldsymbol{k}_2}] = -\sum_{\boldsymbol{q}} \hat{h}_{-\boldsymbol{q}}\, S^+_{\boldsymbol{q}+\boldsymbol{k}_2}$, and thus

$$\left[S^-_{-\boldsymbol{k}_2}, [H_1, S^+_{\boldsymbol{k}_2}]\right] = 2\sum_{\boldsymbol{q}} \hat{h}_{-\boldsymbol{q}}\, S^z_{\boldsymbol{q}} \quad . \tag{1.102}$$

Third, $[S^-_{-\boldsymbol{k}_1}, S^+_{\boldsymbol{k}_2}] = -S^z_{\boldsymbol{k}_2-\boldsymbol{k}_1}$. Finally, we have $\{S^-_{-\boldsymbol{k}_1}, S^+_{\boldsymbol{k}_1}\} = S^-_{-\boldsymbol{k}_1} S^+_{\boldsymbol{k}_1} + S^+_{\boldsymbol{k}_1} S^-_{-\boldsymbol{k}_1}$. Note that

$$\left| \left\langle [S^-_{-\boldsymbol{k}_1}, S^+_{\boldsymbol{k}_2}] \right\rangle \right|^2 = \left| \langle S^z_{\boldsymbol{k}_2-\boldsymbol{k}_1} \rangle \right|^2 = N^2\, |\hat{m}_{\boldsymbol{k}_2-\boldsymbol{k}_1}|^2 \quad . \tag{1.103}$$

We define the quantity $\Gamma(\boldsymbol{k}_2) \equiv \Gamma_0(\boldsymbol{k}_2) + \Gamma_1(\boldsymbol{k}_2)$, where

$$\begin{aligned} \Gamma_0(\boldsymbol{k}_2) &= \left[S^-_{-\boldsymbol{k}_2}, [H_0, S^+_{\boldsymbol{k}_2}]\right] \\ &= \frac{1}{2N}\sum_{\boldsymbol{q}} \Big(\hat{J}(\boldsymbol{q}) - \tfrac{1}{2}\hat{J}(\boldsymbol{q}+\boldsymbol{k}_2) - \tfrac{1}{2}\hat{J}(\boldsymbol{q}-\boldsymbol{k}_2) \Big) \left\langle 4 S^z_{\boldsymbol{q}}\, S^z_{-\boldsymbol{q}} + S^+_{\boldsymbol{q}}\, S^-_{-\boldsymbol{q}} + S^-_{\boldsymbol{q}}\, S^+_{-\boldsymbol{q}} \right\rangle \\ &= \sum_{\boldsymbol{R}_1, \boldsymbol{R}_2} J(\boldsymbol{R}_1) \left(1 - \cos(\boldsymbol{k}_2 \cdot \boldsymbol{R}_1)\right) \left\langle S^x_{\boldsymbol{R}_2} S^x_{\boldsymbol{R}_1+\boldsymbol{R}_2} + S^y_{\boldsymbol{R}_2} S^y_{\boldsymbol{R}_1+\boldsymbol{R}_2} + 2\, S^z_{\boldsymbol{R}_2} S^z_{\boldsymbol{R}_1+\boldsymbol{R}_2} \right\rangle \end{aligned} \tag{1.104}$$

and

$$\Gamma_1(\boldsymbol{k}_2) = \left\langle \left[S^-_{-\boldsymbol{k}_2}, [H_0, S^+_{\boldsymbol{k}_2}]\right] \right\rangle = 2N \sum_{\boldsymbol{q}} \hat{h}_{-\boldsymbol{q}}\, \hat{m}_{\boldsymbol{q}} \quad . \tag{1.105}$$

We may now easily derive the following bounds:

$$\begin{aligned} \Gamma_0(\boldsymbol{k}_2) &\leqslant NS(S+\tfrac{1}{2})\, \boldsymbol{k}_2^2 \sum_{\boldsymbol{R}} \boldsymbol{R}^2 J(\boldsymbol{R}) \equiv 4N\mathcal{J}\, \boldsymbol{k}_2^2\, a^2 \\ \Gamma_1(\boldsymbol{k}_2) &\leqslant 2N \sum_{\boldsymbol{q}} |\hat{h}_{-\boldsymbol{q}}| \cdot |\hat{m}_{\boldsymbol{q}}| \quad . \end{aligned} \tag{1.106}$$

with $\mathcal{J} \equiv \frac{1}{4}S(S+\frac{1}{2})\sum_{\boldsymbol{R}}(\boldsymbol{R}/a)^2\,J(\boldsymbol{R})$, where we presume a d-dimensional cubic lattice of lattice constant a.

We are almost ready to claim our prize. At this point we have the inequality

$$\left\langle S^-_{\boldsymbol{K}-\boldsymbol{k}_2}S^+_{\boldsymbol{k}_2-\boldsymbol{K}} + S^+_{\boldsymbol{k}_2-\boldsymbol{K}}\,S^-_{\boldsymbol{K}-\boldsymbol{k}_2}\right\rangle \geqslant \frac{N}{2\beta}\cdot\frac{|\hat{m}_{\boldsymbol{K}}|^2}{\mathcal{J}\boldsymbol{k}_2^2\,a^2 + \frac{1}{2}\sum_{\boldsymbol{q}}|\hat{h}_{-\boldsymbol{q}}|\cdot|\hat{m}_{\boldsymbol{q}}|} \quad , \tag{1.107}$$

with $\boldsymbol{k}_1 \equiv \boldsymbol{k}_2 - \boldsymbol{K}$. We now sum both sides of the equation over $\boldsymbol{k}_2$, which owing to presumed periodic boundary conditions takes a discrete set of values $\boldsymbol{k}_2 = \big(2\pi n_1/N_1a, 2\pi n_2/N_2a, 2\pi n_3/N_3a\big)$, with $N = N_1\,N_2\,N_3$ where $n_j \in \{1,\ldots,N_j\}$. The system is thus a cubic rectangle of dimensions $N_1 \times N_2 \times N_3$ cells, with each cell of volume a^3. On the LHS, we obtain

$$\sum_{\boldsymbol{k}_2}\left\langle S^-_{\boldsymbol{K}-\boldsymbol{k}_2}S^+_{\boldsymbol{k}_2-\boldsymbol{K}} + S^+_{\boldsymbol{k}_2-\boldsymbol{K}}\,S^-_{\boldsymbol{K}-\boldsymbol{k}_2}\right\rangle = 2N\sum_{\boldsymbol{R}}\left\langle (S^x_{\boldsymbol{R}})^2 + (S^y_{\boldsymbol{R}})^2\right\rangle < 2N^2S(S+1) \quad . \tag{1.108}$$

Thus,

$$|\hat{m}_{\boldsymbol{K}}|^2 < \frac{4S(S+1)}{k_{\rm B}T}\Bigg/ a^d\int\limits_{\hat{\Omega}}\frac{d^dk}{(2\pi)^d}\,\frac{1}{\mathcal{J}\boldsymbol{k}^2a^2 + \frac{1}{2}\sum_{\boldsymbol{q}}|\hat{h}_{-\boldsymbol{q}}|\cdot|\hat{m}_{\boldsymbol{q}}|} \tag{1.109}$$

Consider the case where $\hat{h}_{\boldsymbol{q}} = \hat{h}_{-\boldsymbol{K}}\,\delta_{\boldsymbol{q},-\boldsymbol{K}} + \hat{h}_{\boldsymbol{K}}\,\delta_{\boldsymbol{q},\boldsymbol{K}}$. The Brillouin zone is a d-dimensional cube of side length $2\pi/a$. We are free to *underestimate* the integral, since this *overestimates* the RHS of the above inequality, and to this end we integrate instead over a d-dimensional sphere of radius $b = \pi/a$, yielding

$$|\hat{m}_{\boldsymbol{K}}|^2 < \frac{4S(S+1)}{k_{\rm B}T}\Bigg/\frac{\Omega_d}{(2\pi)^d}\int\limits_0^\pi\frac{du\,u^{d-1}}{\mathcal{J}u^2 + |\hat{h}_{\boldsymbol{K}}|\cdot|\hat{m}_{\boldsymbol{K}}|} \quad , \tag{1.110}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the total solid angle in d space dimensions and $u = |\boldsymbol{k}|a$. We simplify notion by defining $h \equiv |\hat{h}_{\boldsymbol{K}}|$ and $m \equiv |\hat{m}_{\boldsymbol{K}}|$. We may now derive the following inequalities.

$$\begin{aligned} d = 1 \;:\quad & m < C_1\left(\frac{h\mathcal{J}}{(k_{\rm B}T)^2}\right)^{1/3} \\ d = 2 \;:\quad & m < C_2\,\frac{(\mathcal{J}/k_{\rm B}T)^{1/2}}{\ln^{1/2}(\mathcal{J}/hm)} \\ h = 0,\;d > 2 \;:\quad & m < C_d\,\mathcal{J}/k_{\rm B}T \quad , \end{aligned} \tag{1.111}$$

where the $C_{1,2,\ldots}$ are numerical constants. In the limit $h \to 0$, we see that m must vanish for $d = 1$ and $d = 2$. For $d > 2$, there is no such requirement.

1.5.4 Goldstone's theorem

Goldstone's theorem says that whenever the ground state $|\Psi_0\rangle$ of a thermodynamically large quantum system (or QFT) breaks a continuous symmetry of the Hamiltonian H, there is a branch of gapless excitations in the spectrum of H, known as *Goldstone bosons*. Let's illustrate this phenomenon with the explicit example of the spin-S quantum Heisenberg ferromagnet,

$$H = -\sum_{i<j} J_{ij}\, \boldsymbol{S}_i \cdot \boldsymbol{S}_i \quad . \tag{1.112}$$

We assume $J_{ij} = J(\boldsymbol{R}_i - \boldsymbol{R}_j) > 0$, *i.e.* the system is ferromagnetic, preferring all pairs of interacting spins to align. Each component of the total spin operator $\boldsymbol{\mathcal{S}} = \sum_i \boldsymbol{S}_i$ commutes with each term of H, *i.e.*

$$[\mathcal{S}^\alpha, \boldsymbol{S}_i \cdot \boldsymbol{S}_j] = 0 \tag{1.113}$$

because $\boldsymbol{S}_i \cdot \boldsymbol{S}_j$ is rotationally invariant and the $\mathcal{S}^\alpha$ are the generators of rotations. The global symmetry group is SU(2), which is continuous. It is clear that any state $|\Psi_0\rangle$ for which $\boldsymbol{S}_i \cdot \boldsymbol{S}_j\,|\Psi_0\rangle = S^2\,|\Psi_0\rangle$ is a ground state of H, since each individual term in the sum is separately minimized. The ground state energy is then $E_0 = -\frac{1}{2}N\hat{J}(\boldsymbol{0})S^2$, where $\hat{J}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} J(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$.

Consider then the state $|\Uparrow\rangle$ in which $S_i^z\,|\Uparrow\rangle = S\,|\Uparrow\rangle$, *i.e.* all spins maximally polarized in the $\hat{\boldsymbol{z}}$ direction. Clearly $|\Uparrow\rangle$ is a ground state. But so is $|\Downarrow\rangle$. Indeed, so is any state of the form $|\Psi_0(\hat{\boldsymbol{n}})\rangle = \otimes_i |\hat{\boldsymbol{n}}\rangle_i$ in which each spin is in a coherent state maximally polarized along the direction $\hat{\boldsymbol{n}}$, *i.e.* $\hat{\boldsymbol{n}}\cdot\boldsymbol{S}_i\,|\Psi_0(\hat{\boldsymbol{n}})\rangle = S\,|\Psi_0(\hat{\boldsymbol{n}})\rangle$[25]. Thus, we have an entire manifold of states, all with energy E_0, corresponding to total spin $\mathcal{S} = NS$. The degeneracy of this ground state sector is $\mathcal{S}(\mathcal{S}+1)$, and states with different values of $\mathcal{S}^z$ are of course orthogonal. In the coherent state basis, we have

$$\big|\langle\, \Psi_0(\hat{\boldsymbol{n}})\,|\,\Psi_0(\hat{\boldsymbol{n}}')\,\rangle\big|^2 = \left(\frac{1+\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}'}{2}\right)^N \quad , \tag{1.114}$$

where N is the total number of spins. In the thermodynamic limit, these states are *macroscopically distinct*

$$\big|\langle\, \Psi_0(\hat{\boldsymbol{n}})\,\big|\,\mathcal{O}_i\,\mathcal{O}_j\,\big|\,\Psi_0(\hat{\boldsymbol{n}}')\,\rangle\big|^2 = \left(\frac{1+\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{n}}'}{2}\right)^{N-2} \big|\langle\, \hat{\boldsymbol{n}},\hat{\boldsymbol{n}}\,|\,\mathcal{O}_i\,\mathcal{O}_j\,|\,\hat{\boldsymbol{n}},\hat{\boldsymbol{n}}\,\rangle\big|^2 \quad , \tag{1.115}$$

which vanishes in the thermodynamic limit.

Next, consider the operator $S^\alpha(\boldsymbol{k}) = N^{-1}\sum_i S_i^\alpha\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_i}$, which we met in Eqn. 1.95. It is straightforward to derive the result

$$[H, \mathcal{S}^-(\boldsymbol{k})] = \sum_{i,j} J(\boldsymbol{R}_i - \boldsymbol{R}_j)\left(1 - e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}\right) S_j^-\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_j}\, S_i^z \quad . \tag{1.116}$$

[25] In this notation, our previously defined state $|\Psi_0\rangle = |\Uparrow\rangle$ is expressed as $|\Psi_0\rangle = |\Psi_0(\hat{\boldsymbol{z}})\rangle$.

We assume that the diagonal elements of J_{ij} all vanish, *i.e.* $J(\mathbf{0}) = \sum_{\boldsymbol{k}} \hat{J}(\boldsymbol{k}) = 0$. The above identity, acting on the state $|\Uparrow\rangle$, then yields

$$[H, \mathcal{S}^-(\boldsymbol{k})]\,|\Uparrow\rangle = S\big[\hat{J}(\mathbf{0}) - \hat{J}(\boldsymbol{k})\big]\,\mathcal{S}^-(\boldsymbol{k})\,|\Uparrow\rangle \quad , \tag{1.117}$$

which establishes that $\mathcal{S}^-_{\boldsymbol{k}}\,|\Uparrow\rangle$ is an *eigenstate* of H, with eigenvalue $E_{\boldsymbol{k}} = E_0 + S\big[\hat{J}(\mathbf{0}) - \hat{J}(\boldsymbol{k})\big]$. Thus, the excitation spectrum of this branch, which has one state for each wavevector $\boldsymbol{k}$, is

$$\hbar\omega_{\boldsymbol{k}} = E_{\boldsymbol{k}} - E_0 = S\big[\hat{J}(\mathbf{0}) - \hat{J}(\boldsymbol{k})\big] \quad . \tag{1.118}$$

These excitations are called *spin waves*, and they are examples of Goldstone bosons. Note that the $\boldsymbol{k} = \mathbf{0}$ spin wave state,

$$\mathcal{S}^-(\mathbf{0})\,|\Uparrow\rangle = \sum_i S_i^-\,|\Uparrow\rangle \quad , \tag{1.119}$$

is among the ground state manifold, with total spin $\mathcal{S} = NS$ and polarization $\mathcal{S}^z = NS - 1$.

Note that in systems with a discrete symmetry, such as the Ising model, with symmetry group $\mathbb{Z}_2$, there are in general no Goldstone bosons. For the Ising ferromagnet $H = -\sum_{i<j} J_{ij}\,\sigma_i\,\sigma_j$, where each $\sigma_i = \pm 1$, there are two degenerate ground states, $|\Uparrow\rangle$ and $|\Downarrow\rangle$, each of which has total energy $E_0 = -\frac{1}{2}N\hat{J}(\mathbf{0})$. In dimensions $d > 1$, the lowest-lying excitation in either case is a single spin flip, with excitation energy $\Delta E = 2\hat{J}(\mathbf{0})$. (In $d = 1$, with nearest-neighbor interactions, the lowest-lying excitation is a domain wall, with energy $\hat{J}(\mathbf{0}) = 2J$.)

Field theories and Goldstone's theorem

Consider the $|\boldsymbol{\phi}|^4$ field theory for the real n-component field $\boldsymbol{\phi} = (\phi^1, \ldots, \phi^n)$, with relativistic Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu\boldsymbol{\phi})(\partial^\mu\boldsymbol{\phi}) - \tfrac{1}{2}\,m^2\,\boldsymbol{\phi}^2 - \tfrac{1}{4}\,\lambda\,(\boldsymbol{\phi}^2)^2 \quad . \tag{1.120}$$

The metric used to raise and lower indices is $g_{\mu\nu} = \mathsf{diag}(+, -, \cdots, -)$. The equations of motion are found to be

$$\partial_\mu\partial^\mu\boldsymbol{\phi} + m^2\boldsymbol{\phi} + \lambda\,\boldsymbol{\phi}^2\,\boldsymbol{\phi} = 0 \quad . \tag{1.121}$$

If $m^2 > 0$, we can obtain solutions when $|\boldsymbol{\phi}| \ll 1$ by dropping the cubic term. The solutions are plane waves: $\boldsymbol{\phi}(\boldsymbol{x}, t) = A\,e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)}$ where $\omega^2 = \boldsymbol{k}^2 + m^2$. The spectrum is massive.

What happens when $m^2 < 0$? In this case the potential $V(\boldsymbol{\phi}) = \frac{1}{2}m^2\boldsymbol{\phi}^2 + \frac{1}{4}\lambda(\boldsymbol{\phi}^2)^2$ is minimized when $|\boldsymbol{\phi}| = \phi_0 = \sqrt{-m^2/\lambda}$. Let's write $\boldsymbol{\phi} = \phi_0\,(1+\eta)^{1/2}\,\hat{\boldsymbol{\omega}}$, where $\hat{\boldsymbol{\omega}}$ is a unit vector. One then obtains

$$\mathcal{L} = \tfrac{1}{8}\,\phi_0^2\,\frac{(\partial_\mu\eta)(\partial^\mu\eta)}{1+\eta} + \tfrac{1}{2}\,\phi_0^2\,(1+\eta)\,(\partial_\mu\hat{\boldsymbol{\omega}})(\partial^\mu\hat{\boldsymbol{\omega}}) - \tfrac{1}{4}\,m^2\,\phi_0^2(1-\eta^2) \quad . \tag{1.122}$$

The linearized EL equations are then

$$\begin{aligned} \partial_\mu\partial^\mu\eta &= 2m^2\eta \\ \partial_\mu\partial^\mu\hat{\boldsymbol{\omega}} &= 0 \quad . \end{aligned} \tag{1.123}$$

The first equation results in a massive relativistic dispersion $\omega^2 = \boldsymbol{k}^2 + 2m^2$, while the second, as we have seen, yields gapless solutions with $\omega = |\boldsymbol{k}|$, which are the Goldstone bosons. Note that if $\boldsymbol{\phi}$ has $n = 1$ component, then there is no $\hat{\boldsymbol{\omega}}$ field and there are no Goldstone bosons.

Consider next the example of the nonrelativistic complex $|\psi|^4$ theory, otherwise known as the Gross-Pitaevskii model, with

$$\mathcal{L} = i\hbar\,\bar\psi\,\partial_t\,\psi - \frac{\hbar^2}{2m}\,|\boldsymbol{\nabla}\psi|^2 - \tfrac{1}{2}g\big(|\psi|^2 - n_0\big)^2 \tag{1.124}$$

We write $\psi = \sqrt{n_0}\,(1+\eta)\,e^{i\xi}$, which yields

$$\mathcal{L} = -\hbar n_0(1+\eta)^2\,\partial_t\xi - \frac{\hbar^2 n_0}{2m}\,(\boldsymbol{\nabla}\eta)^2 - \frac{\hbar^2 n_0}{2m}\,(1+\eta)^2\,(\boldsymbol{\nabla}\xi)^2 - \tfrac{1}{2}\,g n_0^2\,(4\eta^2 + 4\eta^3 + \eta^4) \quad . \tag{1.125}$$

The linearized equations of motion are then found to be

$$\begin{aligned} -2\hbar n_0\,\partial_t\xi &= -\frac{\hbar^2 n_0}{m}\,\nabla^2\eta + 4gn_0^2\,\eta \\ +2\hbar n_0\,\partial_t\eta &= -\frac{\hbar^2 n_0}{m}\,\nabla^2\xi \quad . \end{aligned} \tag{1.126}$$

We now obtain plane wave solutions of the form

$$\begin{pmatrix}\eta(\boldsymbol{x},t)\\ \xi(\boldsymbol{x},t)\end{pmatrix} = \begin{pmatrix}\hat\eta\\ \hat\xi\end{pmatrix} e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)} \quad , \tag{1.127}$$

which, when inserted into the EL equations, yields the pair

$$\begin{aligned} 2i\hbar\omega\,\hat\xi &= \left(\frac{\hbar^2\boldsymbol{k}^2}{m} + 4gn_0\right)\hat\eta \\ -2i\hbar\omega\,\hat\eta &= \frac{\hbar^2\boldsymbol{k}^2}{m}\hat\xi \quad . \end{aligned} \tag{1.128}$$

The solutions are

$$\omega = \pm\,c\,|\boldsymbol{k}|\sqrt{1 + \frac{\hbar^2\boldsymbol{k}^2}{4m^2c^2}} \quad . \tag{1.129}$$

Note that as $\boldsymbol{k} \to 0$ we obtain a massless relativistic dispersion $\omega = \pm c\,|\boldsymbol{k}|$, while for $|\boldsymbol{k}| \to \infty$ we recover the ballistic dispersion $\omega = \hbar\boldsymbol{k}^2/2m$. These massless excitations are Goldstone modes of the broken $\mathrm{U}(1)$ symmetry and correspond to phonons in a superfluid.

Finally, let's get crazy and consider the *gauged*, relativistic $|\psi|^4$ theory, with

$$\mathcal{L} = \frac{1}{2g}\,(\partial_\mu + ieA_\mu)\bar\psi\,(\partial^\mu + ieA^\mu)\psi - \frac{m^2}{2gn_0}\,(\bar\psi\psi - n_0)^2 - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad , \tag{1.130}$$

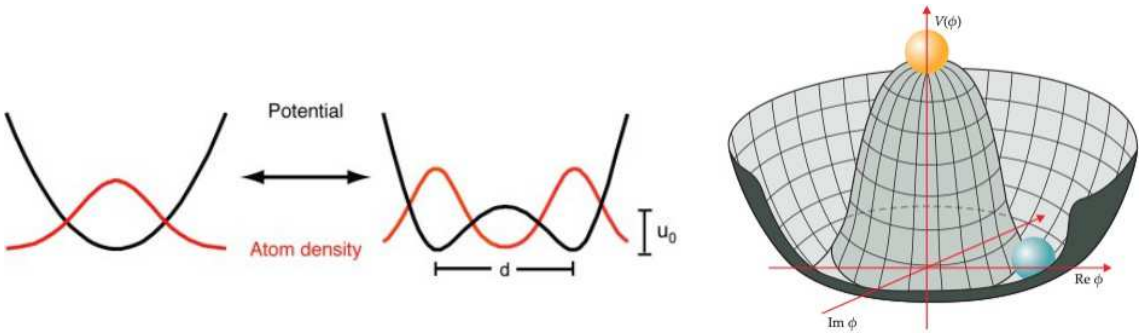


Figure 1.15: Left: double well potential $V(\phi)$ for the scalar field $\phi \in \mathbb{R}$. Right: Mexican hat potential for the complex scalar field $\phi \in \mathbb{C}$ with $V(\phi, \bar{\phi}) = \frac{1}{2}m^2\big(|\phi|^2 - n_0\big)^2$.

where the field strength tensor is written in terms of the gauge field as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. This theory has a *local gauge symmetry* whereby the combined operations

$$\psi \to \psi\, e^{i\alpha} \qquad , \qquad A_\mu \to A_\mu - e^{-1}\,\partial_\mu \alpha \quad , \tag{1.131}$$

where $\alpha(\boldsymbol{x}, t)$ is a field, leaves $\mathcal{L}$ invariant.

If we drop the gauge field, *i.e.* set $e = 0$, we may again expand about one of the local minima at $|\psi| = n_0^{1/2}$, writing

$$\psi = \sqrt{n_0}\,(1+\eta)\,e^{i\xi} \quad , \tag{1.132}$$

as above. Then

$$\mathcal{L} = \frac{n_0}{2g}\,(\partial_\mu\eta)(\partial^\mu\eta) + \frac{n_0}{2g}\,(1+\eta)^2\,(\partial_\mu\xi)(\partial^\mu\xi) - \frac{m^2 n_0}{2g}\left(4\eta^2 + 4\eta^3 + \eta^4\right) \quad . \tag{1.133}$$

The linearized equations of motion are then

$$\partial_\mu\partial^\mu\eta = (2m)^2\eta \qquad , \qquad \partial_\mu\partial^\mu\xi = 0 \quad , \tag{1.134}$$

yielding two modes. The η mode carries a massive relativistic dispersion $\omega^2 = \boldsymbol{k}^2 + (2m)^2$, while the ξ mode is the gapless Goldstone boson with $\omega^2 = \boldsymbol{k}^2$.

When $e \neq 0$, we may take advantage of the gauge invariance to define a shifted gauge field

$$\widetilde{A}_\mu \equiv A_\mu - e^{-1}\partial_\mu\xi \quad . \tag{1.135}$$

The transformed Lagrangian becomes

$$\mathcal{L} = \frac{n_0}{2g}\,(\partial_\mu\eta)(\partial^\mu\eta) + \frac{n_0\,e^2}{2g}\,(1+\eta)^2\,\widetilde{A}_\mu\widetilde{A}^\mu - \frac{m^2 n_0}{2g}\left(4\eta^2 + 4\eta^3 + \eta^4\right) - \tfrac{1}{4}\widetilde{F}_{\mu\nu}\widetilde{F}^{\mu\nu} \quad , \tag{1.136}$$

from which we can read off the following features:

- The photon has become a massive triplet $\widetilde{A}_\mu$, with $m_{\widetilde{A}} = e\sqrt{n_0/g}$.
- There is no ξ particle! It has been *eaten* by the photon field.
- The Higgs field η has a mass which is independent of e, with $m_\eta = 2m$.

To reiterate what we have learned in this section: Classically, a particle in a double well potential exhibits spontaneous symmetry breaking. It *must* choose which well in which to lie. But quantum mechanically, we know that the ground state is a symmetric superposition of left and right well states. A similar consideration holds in the case of continuous symmetries, for the Mexican hat potential. A classical particle will minimize its energy at any point along the trough of the potential, but the quantum ground state is a rotationally symmetric s-wave state with no nodes. In a field theory setting, there are an *infinite number* of degrees of freedom, and the double well or Mexican hat exists *at every point in space*. This kills tunneling between the different broken symmetry 'pure' states, *viz.* $\langle\, \mathsf{L} \,|\, \mathsf{R} \,\rangle \sim \langle\, \varphi_{\mathsf{L}} \,|\, \varphi_{\mathsf{R}} \,\rangle^V \to 0$ in the $V \to \infty$ limit of infinite system volume. This state of affairs is described in Fig. 1.15.

1.6 Appendix : The Foldy-Wouthuysen Transformation

1.6.1 The Dirac Hamiltonian

Let us write

$$\hat{H} = mc^2\,\gamma^0 + c\gamma^0\,\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + V\ , \tag{1.137}$$

where

$$\boldsymbol{\pi} = \boldsymbol{p} + \tfrac{e}{c}\boldsymbol{A} \tag{1.138}$$

is the dynamical momentum and where the γ^μ are the Dirac matrices,

$$\gamma^0 = \begin{pmatrix} 1_{2\times2} & 0_{2\times2} \\ 0_{2\times2} & -1_{2\times2} \end{pmatrix} \qquad , \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0_{2\times2} & \boldsymbol{\sigma}_{2\times2} \\ -\boldsymbol{\sigma}_{2\times2} & 0_{2\times2} \end{pmatrix}\ . \tag{1.139}$$

Here $\boldsymbol{\sigma}$ is the vector of Pauli matrices.

The idea behind the FW transformation is to unitarily transform to a different Hilbert space basis such that the coupling in $\hat{H}$ between the upper and lower components of the Dirac spinor vanishes. This may be done systematically as an expansion in inverse powers of the electron mass m. We begin by defining $K \equiv c\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + V$ so that $\hat{H} = mc^2\,\gamma^0 + K$. Note that K is of order m^0. We then write

$$\begin{aligned}\tilde{\hat{H}} &= e^{iS}\,\hat{H}\,e^{-iS} \\ &= \hat{H} + i\big[S,\hat{H}\big] + \frac{(i)^2}{2!}\,\big[S,[S,\hat{H}]\big] + \dots\ ,\end{aligned} \tag{1.140}$$

where S itself is written as a power series in $(mc^2)^{-1}$:

$$S = \frac{S_0}{mc^2} + \frac{S_1}{(mc^2)^2} + \dots \ . \tag{1.141}$$

The job now is to write $\tilde{\hat{H}}$ as a power series in m^{-1}. The first few terms are easy to find:

$$\tilde{\hat{H}} = mc^2\,\gamma^0 + K + i\big[S_0,\gamma^0\big] + \frac{1}{mc^2}\left(i\big[S_0,K\big] + i\big[S_1,\gamma^0\big] - \tfrac{1}{2}\big[S_0,[S_0,\gamma^0]\big]\right) + \dots \tag{1.142}$$

We choose the operators S_n so as to cancel, at each order in m^{-1}, the off-diagonal terms in $\tilde{\hat{H}}$ that couple the upper two components of Ψ to the lower two components of Ψ. To order m^0, we then demand

$$c\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + i\big[S_0,\gamma^0\big] = 0\ . \tag{1.143}$$

Note that we do not demand that $i\big[S_0,\gamma^0\big]$ completely cancel K – indeed it is impossible to find such an S_0, and one way to see this is to take the trace. The trace of any commutator must vanish, but $\mathrm{Tr}\,K = 4V$, which is in general nonzero. But this is of no concern to us, since we only need cancel the (traceless) *off-diagonal* part of K, which is to say $c\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\pi}$.

To solve for S_0, one can write it in terms of its four 2×2 subblocks, compute the commutator with γ^0, and then impose eqn. 14.341. One then finds $S_0 = -\frac{i}{2}c\,\boldsymbol{\gamma}\cdot\boldsymbol{\pi}$.

STUDENT EXERCISE: Derive the result $S_0 = -\frac{i}{2}c\,\boldsymbol{\gamma}\cdot\boldsymbol{\pi}$.

At the next level, we have to deal with the term in the round brackets in eqn. 14.340. Since we know S_0, we can compute the first and the third terms therein. In general, this will leave us with an off-diagonal term coupling upper and lower components of Ψ. We then choose S_1 so as to cancel this term. This calculation already is tedious, and we haven't even gotten to the spin-orbit interaction term yet, since it is of order m^{-2}.

1.6.2 Emergence of the spin-orbit and Zeeman interaction terms

Here's a simpler way to proceed to order m^{-2}. Let a,b be block indices and i,j be indices within each block. Thus, the component Ψ_{ai} is the $i^{\rm th}$ component of the $a^{\rm th}$ block; $\Psi_{a=1,i=2}$ is the lower component of the upper block, *i.e.* the second component of the four-vector Ψ.

Write the Hamiltonian as

$$\hat{H} = mc^2\,\tau^z + c\boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,\tau^x + V(\boldsymbol{r})\ , \tag{1.144}$$

where τ^μ are Pauli matrices with indices a,b and σ^ν are Pauli matrices with indices i,j. The σ and τ matrices commute because they act on different indices.

A very important result regarding Pauli matrices:

$$e^{i\theta\,\hat{\boldsymbol{n}}\cdot\boldsymbol{\tau}/2}\,\tau^\alpha\,e^{-i\theta\,\hat{\boldsymbol{n}}\cdot\boldsymbol{\tau}/2} = n^\alpha n^\beta\,\tau^\beta + \cos\theta\,(\delta^{\alpha\beta} - n^\alpha n^\beta)\,\tau^\beta + \sin\theta\,\epsilon^{\alpha\beta\gamma}\,n^\beta\,\tau^\gamma\ . \tag{1.145}$$

STUDENT EXERCISE: Verify and interpret the above result.

Using this result, we can write

$$A\,\tau^z + B\,\tau^x = \sqrt{A^2+B^2}\cdot e^{-i\tan^{-1}(B/A)\,\tau^y/2}\,\tau^z\,e^{i\tan^{-1}(B/A)\,\tau^y/2}\ , \tag{1.146}$$

and, for our specific purposes,

$$mc^2\,\tau^z + c\boldsymbol{\sigma}\cdot\boldsymbol{\pi}\,\tau^x = \sqrt{(mc^2)^2+(c\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}\cdot U\,\tau^z\,U^\dagger\ , \tag{1.147}$$

where

$$U = \exp\left\{-\frac{i}{2}\tan^{-1}\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right)\tau^y\right\}\ . \tag{1.148}$$

The fact that $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ is an operator is no obstacle here, since it commutes with the τ^μ matrices. We can give meaning to expressions like $\tan^{-1}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}/mc)$ in terms of their Taylor series expansions.

We therefore have the result,

$$U^\dagger\,\hat{H}\,U = \sqrt{(mc^2)^2+(c\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}\cdot\tau^z + U^\dagger\,V(\boldsymbol{r})\,U\ . \tag{1.149}$$

The first term is diagonal in the block indices. Expanding the square root, we have

$$\begin{aligned} mc^2\sqrt{1+\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right)^2} &= mc^2 + \frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}{2m} + \mathcal{O}(m^{-3}) \\ &= mc^2 + \frac{\boldsymbol{\pi}^2}{2m} + \frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma} + \mathcal{O}(m^{-3})\ , \end{aligned} \tag{1.150}$$

since

$$\begin{aligned} (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 &= \sigma^\mu\sigma^\nu\,\pi^\mu\pi^\nu = (\delta^{\mu\nu} + i\epsilon^{\mu\nu\lambda}\sigma^\lambda)\,\pi^\mu\pi^\nu \\ &= \boldsymbol{\pi}^2 + \tfrac{i}{2}\epsilon^{\mu\nu\lambda}\left[p^\mu + \tfrac{e}{c}A^\mu, p^\nu + \tfrac{e}{c}A^\nu\right] = \boldsymbol{\pi}^2 + \frac{e\hbar}{c}\,\boldsymbol{B}\cdot\boldsymbol{\sigma}\ . \end{aligned} \tag{1.151}$$

We next need to compute $U^\dagger\,V(\boldsymbol{r})\,U$ to order m^{-2}. To do this, first note that

$$U = 1 - \frac{i}{2}\,\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\,\tau^y - \frac{1}{8}\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right)^2 + \ldots\ , \tag{1.152}$$

Thus,

$$U^\dagger\,V\,U = V + \frac{i}{2mc}\big[\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, V\big]\,\tau^y - \frac{1}{8m^2c^2}\big[\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, [\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, V]\big] + \ldots\ . \tag{1.153}$$

Upon reflection, one realizes that, to this order, it suffices to take the first term in the Taylor expansion of $\tan^{-1}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}/mc)$ in eqn. 14.346, in which case one can then invoke eqn. 14.338 to

obtain the above result. The second term on the RHS of eqn. 14.351 is simply $\frac{\hbar}{2mc}\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y$. The third term is

$$\begin{aligned}\frac{i\hbar}{8m^2c^2}\big[\sigma^\mu\pi^\mu,\sigma^\nu\partial^\nu V\big] &= \frac{i\hbar}{8m^2c^2}\Big\{\sigma^\mu\,\big[\pi^\mu,\sigma^\nu\partial^\nu V\big]+\big[\sigma^\mu,\sigma^\nu\partial^\nu V\big]\,\pi^\mu\Big\}\\ &= \frac{i\hbar}{8m^2c^2}\left\{\frac{\hbar}{i}\partial^\mu\partial^\nu V\,\sigma^\mu\sigma^\nu+2i\epsilon^{\mu\nu\lambda}\sigma^\lambda\partial^\nu V\,\pi^\mu\right\}\\ &= \frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2V+\frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi}\ .\end{aligned}\tag{1.154}$$

Therefore,

$$\begin{aligned}U^\dagger\,\hat H\,U = &\left(mc^2+\frac{\boldsymbol{\pi}^2}{2m}+\frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma}\right)\tau^z+V+\frac{\hbar}{2mc}\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y\\ &+\frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2V+\frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi}+\mathcal{O}(m^{-3})\ .\end{aligned}\tag{1.155}$$

This is not block-diagonal, owing to the last term on the RHS of the top line. We can eliminate this term by effecting yet another unitary transformation. However, this will result in a contribution to the energy of order m^{-3}, so we can neglect it. To substantiate this last claim, drop all the block-diagonal terms except for the leading order one, $mc^2\,\tau^z$, and consider the Hamiltonian

$$\mathcal{K}=mc^2\,\tau^z+\frac{\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y\ .\tag{1.156}$$

We now know how to bring this to block-diagonal form. The result is

$$\tilde{\mathcal{K}}=mc^2\sqrt{1+\left(\frac{\hbar\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V}{2m^2c^3}\right)^2}\,\tau^z=\left(mc^2+\frac{\hbar^2(\boldsymbol{\nabla}\,V)^2}{8m^3c^4}+\ldots\right)\tau^z\ ,\tag{1.157}$$

and the correction is of order m^{-3}, as promised.

We now assume all the negative energy ($\tau^z=-1$) states are filled. The Hamiltonian for the electrons, valid to $\mathcal{O}(m^{-3})$, is then

$$\tilde{\hat H}=mc^2+V+\frac{\boldsymbol{\pi}^2}{2m}+\frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma}+\frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2V+\frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi}\ .\tag{1.158}$$

1.7 Appendix : Ideal Bose Gas Condensation

We begin with the grand canonical Hamiltonian $K=H-\mu N$ for the ideal Bose gas,

$$K=\sum_{\boldsymbol{k}}(\varepsilon_{\boldsymbol{k}}-\mu)\,b^\dagger_{\boldsymbol{k}}b_{\boldsymbol{k}}-\sqrt{N}\sum_{\boldsymbol{k}}\left(\nu_{\boldsymbol{k}}\,b^\dagger_{\boldsymbol{k}}+\bar\nu_{\boldsymbol{k}}\,b_{\boldsymbol{k}}\right)\quad .\tag{1.159}$$

Here $b^\dagger_{\boldsymbol{k}}$ is the creation operator for a boson in a state of wavevector $\boldsymbol{k}$, hence $\big[b_{\boldsymbol{k}}\,,\,b^\dagger_{\boldsymbol{k}'}\big] = \delta_{\boldsymbol{k}\boldsymbol{k}'}$. The dispersion relation is given by the function $\varepsilon_{\boldsymbol{k}}$, which is the energy of a particle with wavevector $\boldsymbol{k}$. We must have $\varepsilon_{\boldsymbol{k}} - \mu \geqslant 0$ for all $\boldsymbol{k}$, lest the spectrum of K be unbounded from below. The fields $\{\nu_{\boldsymbol{k}}, \bar\nu_{\boldsymbol{k}}\}$ break a global $\mathsf{O}(2)$ symmetry.

Students who have not taken a course in solid state physics can skip the following paragraph, and be aware that $N = V/v_0$ is the total volume of the system in units of a fundamental "unit cell" volume v_0. The thermodynamic limit is then $N \to \infty$. Note that N is not the boson particle number, which we'll call N_b.

Solid state physics boilerplate : We presume a setting in which the real space Hamiltonian is defined by some boson hopping on a Bravais lattice. The wavevectors $\boldsymbol{k}$ are then restricted to the first Brillouin zone, $\hat\Omega$, and assuming periodic boundary conditions are quantized according to the condition $\exp\big(iN_l\,\boldsymbol{k}\cdot\boldsymbol{a}_l\big) = 1$ for all $l \in \{1,\ldots,d\}$, where $\boldsymbol{a}_l$ is the $l^{\rm th}$ fundamental direct lattice vector and N_l is the size of the system in the $\boldsymbol{a}_l$ direction; d is the dimension of space. The total number of unit cells is $N \equiv \prod_l N_l$. Thus, quantization entails $\boldsymbol{k} = \sum_l (2\pi n_l/N_l)\,\boldsymbol{b}_l$, where $\boldsymbol{b}_l$ is the $l^{\rm th}$ elementary reciprocal lattice vector ($\boldsymbol{a}_l\cdot\boldsymbol{b}_{l'} = 2\pi\delta_{ll'}$) and n_l ranges over N_l distinct integers such that the allowed $\boldsymbol{k}$ points form a discrete approximation to $\hat\Omega$.

To solve, we first shift the boson creation and annihilation operators, writing

$$K = \sum_{\boldsymbol{k}} (\varepsilon_{\boldsymbol{k}} - \mu)\,\beta^\dagger_{\boldsymbol{k}}\beta_{\boldsymbol{k}} - N\sum_{\boldsymbol{k}} \frac{|\nu_{\boldsymbol{k}}|^2}{\varepsilon_{\boldsymbol{k}} - \mu} \quad , \tag{1.160}$$

where

$$\beta_{\boldsymbol{k}} = b_{\boldsymbol{k}} - \frac{\sqrt{N}\,\nu_{\boldsymbol{k}}}{\varepsilon_{\boldsymbol{k}} - \mu} \qquad , \qquad \beta^\dagger_{\boldsymbol{k}} = b^\dagger_{\boldsymbol{k}} - \frac{\sqrt{N}\,\bar\nu_{\boldsymbol{k}}}{\varepsilon_{\boldsymbol{k}} - \mu} \quad . \tag{1.161}$$

Note that $\big[\beta_{\boldsymbol{k}}\,,\,\beta^\dagger_{\boldsymbol{k}'}\big] = \delta_{\boldsymbol{k}\boldsymbol{k}'}$ so the above transformation is canonical. The Landau free energy $\Omega = -k_{\rm B}T\ln\Xi$, where $\Xi = \mathsf{Tr}\; e^{-K/k_{\rm B}T}$, is given by

$$\Omega = Nk_{\rm B}T\int\limits_{-\infty}^{\infty} d\varepsilon\; g(\varepsilon)\;\ln\big(1 - e^{(\mu-\varepsilon)/k_{\rm b}T}\big) - N\sum_{\boldsymbol{k}}\frac{|\nu_{\boldsymbol{k}}|^2}{\varepsilon_{\boldsymbol{k}} - \mu} \quad , \tag{1.162}$$

where $g(\varepsilon)$ is the density of energy states per unit cell,

$$g(\varepsilon) = \frac{1}{N}\sum_{\boldsymbol{k}}\delta\big(\varepsilon - \varepsilon_{\boldsymbol{k}}\big) \xrightarrow[N\to\infty]{} v_0\int\limits_{\hat\Omega}\frac{d^dk}{(2\pi)^d}\,\delta\big(\varepsilon - \varepsilon_{\boldsymbol{k}}\big) \quad . \tag{1.163}$$

Note that

$$\psi_{\boldsymbol{k}} \equiv \frac{1}{\sqrt{N}}\,\big\langle b_{\boldsymbol{k}}\big\rangle = -\frac{1}{N}\,\frac{\partial\Omega}{\partial\bar\nu_{\boldsymbol{k}}} = \frac{\nu_{\boldsymbol{k}}}{\varepsilon_{\boldsymbol{k}} - \mu} \quad . \tag{1.164}$$

In the condensed phase, $\psi_{\boldsymbol{k}}$ is nonzero.

The Landau free energy (grand potential) is a function $\Omega(T, N, \mu, \nu, \bar{\nu})$. We now make a Legendre transformation,

$$Y(T, N, \mu, \psi, \bar{\psi}) = \Omega(T, N, \mu, \nu, \bar{\nu}) + N \sum_{\boldsymbol{k}} \left(\nu_{\boldsymbol{k}} \bar{\psi}_{\boldsymbol{k}} + \bar{\nu}_{\boldsymbol{k}} \psi_{\boldsymbol{k}} \right) \quad . \tag{1.165}$$

Note that

$$\frac{\partial Y}{\partial \bar{\nu}_{\boldsymbol{k}}} = \frac{\partial \Omega}{\partial \bar{\nu}_{\boldsymbol{k}}} + N \psi_{\boldsymbol{k}} = 0 \quad , \tag{1.166}$$

by the definition of $\psi_{\boldsymbol{k}}$. Similarly, $\partial Y / \partial \nu_{\boldsymbol{k}} = 0$. We now have

$$Y(T, N, \mu, \psi, \bar{\psi}) = N k_{\rm B} T \int\limits_{-\infty}^{\infty} d\varepsilon \; g(\varepsilon) \; \ln \left(1 - e^{(\mu - \varepsilon)/k_{\rm b} T} \right) + N \sum_{\boldsymbol{k}} (\varepsilon_{\boldsymbol{k}} - \mu) \, |\psi_{\boldsymbol{k}}|^2 \quad . \tag{1.167}$$

Therefore, the boson particle number per unit cell is given by the *dimensionless density*,

$$n = \frac{N_b}{N} = -\frac{1}{N} \frac{\partial Y}{\partial \mu} = \sum_{\boldsymbol{k}} |\psi_{\boldsymbol{k}}|^2 + \int\limits_{-\infty}^{\infty} d\varepsilon \; \frac{g(\varepsilon)}{e^{(\varepsilon - \mu)/k_{\rm B} T} - 1} \quad , \tag{1.168}$$

and the relation between the condensate amplitude $\psi_{\boldsymbol{k}}$ and the field $\nu_{\boldsymbol{k}}$ is given by

$$\nu_{\boldsymbol{k}} = \frac{1}{N} \frac{\partial Y}{\partial \bar{\psi}_{\boldsymbol{k}}} = (\varepsilon_{\boldsymbol{k}} - \mu) \, \psi_{\boldsymbol{k}} \quad . \tag{1.169}$$

Recall that $\nu_{\boldsymbol{k}}$ acts as an external field. Let the dispersion $\varepsilon_{\boldsymbol{k}}$ be minimized at $\boldsymbol{k} = \boldsymbol{K}$. Without loss of generality, we may assume this minimum value is $\varepsilon_{\boldsymbol{K}} = 0$. We see that if $\nu_{\boldsymbol{k}} = 0$ then one of two must be true:

(i) $\psi_{\boldsymbol{k}} = 0$ for all $\boldsymbol{k}$

(ii) $\mu = \varepsilon_{\boldsymbol{K}}$, in which case $\psi_{\boldsymbol{K}}$ can be nonzero.

Thus, for $\nu = \bar{\nu} = 0$ and $\mu > 0$, we have the usual equation of state,

$$n(T, \mu) = \int\limits_{-\infty}^{\infty} d\varepsilon \; \frac{g(\varepsilon)}{e^{(\varepsilon - \mu)/k_{\rm B} T} - 1} \quad , \tag{1.170}$$

which relates the intensive variables n, T, and μ. When $\mu = 0$, the equation of state becomes

$$n(T, \mu = 0) = \overbrace{\sum_{\boldsymbol{K}} |\psi_{\boldsymbol{K}}|^2}^{n_0} + \overbrace{\int\limits_{-\infty}^{\infty} d\varepsilon \; \frac{g(\varepsilon)}{e^{\varepsilon/k_{\rm B} T} - 1}}^{n_>(T)} \quad , \tag{1.171}$$

where now the sum is over only those $\boldsymbol{K}$ for which $\varepsilon_{\boldsymbol{K}} = 0$. Typically this set has only one member, $\boldsymbol{K} = 0$, but it is quite possible, due to symmetry reasons, that there are more such $\boldsymbol{K}$ values. This last equation of state is one which relates the intensive variables n, T, and n_0, where

$$n_0 = \sum_{\boldsymbol{K}} |\psi_{\boldsymbol{K}}|^2 \tag{1.172}$$

is the dimensionless condensate density. If the integral $n_>(T)$ in Eqn. 1.171 is finite, then for $n > n_0(T)$ we must have $n_0 > 0$. Note that, for any T, $n_>(T)$ diverges logarithmically whenever $g(0)$ is finite. This means that Eqn. 1.170 can always be inverted to yield a finite $\mu(n, T)$, no matter how large the value of n, in which case there is no condensation and $n_0 = 0$. If $g(\varepsilon) \propto \varepsilon^\alpha$ with $\alpha > 0$, the integral converges and $n_>(T)$ is finite and monotonically increasing for all T. Thus, for fixed dimensionless number n, there will be a *critical temperature* $T_{\rm c}$ for which $n = n_>(T_{\rm c})$. For $T < T_{\rm c}$, Eqn. 1.170 has no solution for any μ and we must appeal to eqn. 1.171. The condensate density, given by $n_0(n, T) = n - n_>(T)$, is then finite for $T < T_{\rm c}$, and vanishes for $T \geqslant T_{\rm c}$.

In the condensed phase, the phase of the order parameter ψ inherits its phase from the external field ν, which is taken to zero, in the same way the magnetization in the symmetry-broken phase of an Ising ferromagnet inherits its direction from an applied field h which is taken to zero. The important feature is that in both cases the applied field is taken to zero *after* the approach to the thermodynamic limit.

1.8 Appendix : Asymptotic Series in a Zero-Dimensional Field theory

In this appendix we will solve numerically for a zero-dimensional field theory, *i.e.* an integral $F(\lambda)$, which depends parametrically on a dimensionless parameter λ, with $F(0) = 1$, and compare the results with expansions from diagrammatic perturbation theory, We will see that the perturbation expansion is *asymptotic*, meaning that it is formally divergent, *i.e.* with a vanishing radius of convergence. However, if the results of summing the first N terms results in a relative error $S_N(\lambda) = R_N(\lambda)/F(\lambda)$, where $R_N(\lambda)$ is the remainder after N terms, is minimized by setting $N = N^*(\lambda)$, where $N^*(\lambda) \to \infty$ as $\lambda \to 0$.

Problem : The normalized Gaussian distribution $P(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ has the $n^{\rm th}$ moment

$$\langle x^n \rangle = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} dx\; x^n\, e^{-x^2/2}$$

Clearly $\langle x^n \rangle = 0$ if n is a nonnegative odd integer. Next consider the *generating function*

$$Z(j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, e^{-x^2/2}\, e^{jx} = \exp\left(\tfrac{1}{2} j^2\right) .$$

(a) Show that $\langle x^n \rangle = (d^n Z/dj^n)\big|_{j=0}$ and provide an explicit result for $\langle x^{2k} \rangle$ where $k \in \mathbb{N}$.

(b) Now consider the following integral:

$$F(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, \exp\left(-\frac{1}{2}\, x^2 - \frac{\lambda}{4!}\, x^4 \right) .$$

Clearly $F(0) = 1$, but for general $\lambda > 0$ the integral has no known analytic form[26], but we may express the result as a power series in the parameter λ by Taylor expanding $\exp\left(-\frac{\lambda}{4!}\, x^4\right)$ and then using the result of part (a) for the moments $\langle x^{4k} \rangle$. Find the coefficients in the perturbation expansion,

$$F(\lambda) = \sum_{k=0}^{\infty} C_k\, \lambda^k .$$

(c) Define the *remainder after N terms* as

$$R_N(\lambda) = F(\lambda) - \sum_{k=0}^{N} C_k\, \lambda^k .$$

Compute $R_N(\lambda)$ by evaluating numerically the integral for $F(\lambda)$ (using `Mathematica` or some other numerical package) and subtracting the finite sum. Then define the ratio $S_N(\lambda) = R_N(\lambda)/F(\lambda)$, which is the relative error from the N term approximation and plot the absolute relative error $\big|S_N(\lambda)\big|$ *versus* N for several values of λ. (I suggest you plot the error on a log scale.) What do you find?? Try a few values of λ including $\lambda = 0.01$, $\lambda = 0.05$, $\lambda = 0.2$, $\lambda = 0.5$, $\lambda = 1$, $\lambda = 2$.

(d) Repeat the calculation for the integral

$$G(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\, \exp\left(-\frac{1}{2}\, x^2 - \frac{\lambda}{6!}\, x^6 \right) .$$

(e) Reflect meaningfully on the consequences for weakly and strongly coupled quantum field theories.

[26] In fact, it does. According to `Mathematica`, $F(\lambda) = \sqrt{\frac{2u}{\pi}}\, \exp(u)\, K_{1/4}(u)$, where $u = 3/4\lambda$ and $K_\nu(z)$ is the modified Bessel function. I am grateful to Prof. John McGreevy for pointing this out.

Solution

(a) Clearly $\langle x^n \rangle = (d^n Z/dj^n)\big|_{j=0}$, and so $\langle\, x^n \,\rangle = \left(d^n Z/dj^n\right)_{j=0}$. With $Z(j) = \exp\left(\frac{1}{2}j^2\right)$, only the $k^{\rm th}$ order term in j^2 in the Taylor series for $Z(j)$ contributes, and we obtain

$$\langle\, x^{2k} \,\rangle = \frac{d^{2k}}{dj^{2k}}\left(\frac{j^{2k}}{2^k\, k!}\right) = \frac{(2k)!}{2^k\, k!}\ .$$

(b) We have

$$F(\lambda) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{\lambda}{4!}\right)^n \langle\, x^{4n} \,\rangle = \sum_{n=0}^{\infty} \frac{(4n)!}{4^n\,(4!)^n\, n!\,(2n)!}\,(-\lambda)^n\ .$$

This series is *asymptotic*. It has the properties

$$\lim_{\lambda\to 0} \frac{R_N(\lambda)}{\lambda^N} = 0 \quad \text{(fixed } N\text{)} \qquad , \qquad \lim_{N\to\infty} \frac{R_N(\lambda)}{\lambda^N} = \infty \quad \text{(fixed } \lambda\text{)}\ ,$$

where $R_N(\lambda)$ is the remainder after N terms, defined in part (c). The radius of convergence is zero. To see this, note that if we reverse the sign of λ, then the integrand of $F(\lambda)$ diverges badly as $x \to \pm\infty$. So $F(\lambda)$ is infinite for $\lambda < 0$, which means that there is no disk of any finite radius of convergence which encloses the point $\lambda = 0$. Note that by Stirling's rule,

$$C_n \equiv \frac{(-1)^n\,(4n)!}{4^n\,(4!)^n\, n!\,(2n)!} \sim (-1)^n\, n^n \cdot \left(\tfrac{2}{3}\right)^n e^{-n} \cdot (\pi n)^{-1/2}\ ,$$

and we conclude that the magnitude of the summand reaches a minimum value when $n = n^*(\lambda)$, with $n^*(\lambda) \approx 3/2\lambda$ for small values of λ. For large n, the magnitude of the coefficient C_n grows as $|C_n| \sim e^{n\ln n + \mathcal{O}(n)}$, which dominates the λ^n term, no matter how small λ is.

(c) Results are plotted in fig. 1.16.

It is worth pointing out that the series for $F(\lambda)$ and for $\ln F(\lambda)$ have diagrammatic interpretations. For a Gaussian integral, one has

$$\langle\, x^{2k} \,\rangle = \langle\, x^2 \,\rangle^k \cdot A_{2k}$$

where A_{2k} is the *number of contractions*. For our integral, $\langle\, x^2 \,\rangle = 1$. The number of contractions A_{2k} is computed in the following way. For each of the $2k$ powers of x, we assign an index running from 1 to $2k$. The indices are *contracted*, *i.e.* paired, with each other. How many pairings are there? Suppose we start with any from among the $2k$ indices. Then there are $(2k-1)$ choices for its mate. We then choose another index arbitrarily. There are now $(2k-3)$ choices for *its* mate. Carrying this out to its completion, we find that the number of contractions is

$$A_{2k} = (2k-1)(2k-3)\cdots 3\cdot 1 = \frac{(2k)!}{2^k\, k!}\ ,$$

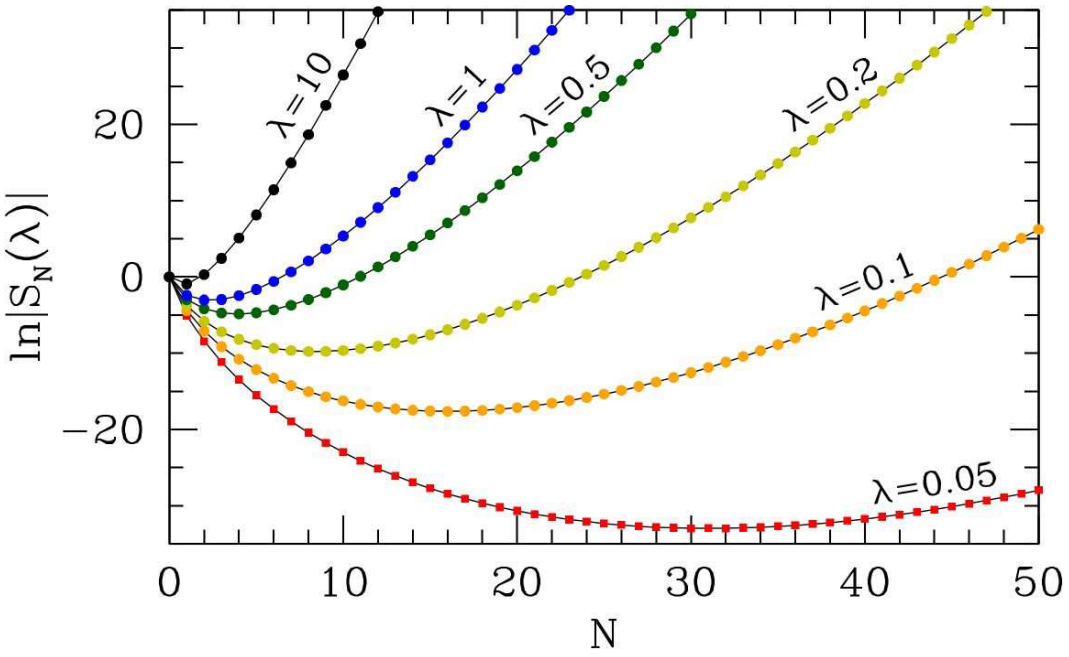


Figure 1.16: Relative error *versus* number of terms kept for the asymptotic series for $F(\lambda)$ (quartic theory). Note that the optimal number of terms to sum is $N^*(\lambda) \approx \frac{3}{2\lambda}$.

exactly as we found in part (a). Now consider the integral $F(\lambda)$. If we expand the quartic term in a power series, then each power of λ brings an additional four powers of x. It is therefore convenient to represent each such quartet with the symbol $\times$. At order N of the series expansion, we have N $\times$'s and $4N$ indices to contract. Each full contraction of the indices may be represented as a labeled diagram, which is in general composed of several disjoint connected subdiagrams. Let us label these subdiagrams, which we will call clusters, by an index γ. Now suppose we have a diagram consisting of m_γ subdiagrams of type γ, for each γ. If the cluster γ contains n_γ vertices ($\times$), then we must have

$$N = \sum_\gamma m_\gamma \, n_\gamma \ .$$

How many ways are there of assigning the labels to such a diagram? One might think $(4!)^N \cdot N!$, since for each vertex $\times$ there are $4!$ permutations of its four labels, and there are $N!$ ways to permute all the vertices. However, this overcounts diagrams which are *invariant* under one or more of these permutations. We define the *symmetry factor* s_γ of the (unlabeled) cluster γ as the number of permutations of the indices of a corresponding labeled cluster which result in the same contraction. We can also permute the m_γ identical disjoint clusters of type γ.

Examples of clusters and their corresponding symmetry factors are provided in fig. 1.17, for all diagrams with $n_\gamma \leqslant 3$. There is only one diagram with $n_\gamma = 1$, resembling ⊗. To obtain $s_\gamma = 8$, note that each of the circles can be separately rotated by an angle π about the long symmetry axis. In addition, the figure can undergo a planar rotation by π about an axis which runs

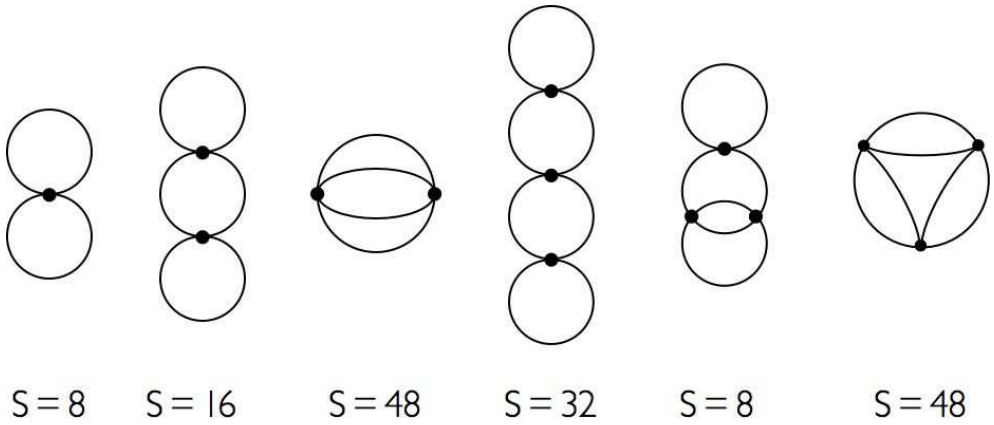


Figure 1.17: Cluster symmetry factors for the quartic theory. A vertex is represented as a black dot (•) with four 'legs'.

through the sole vertex and is normal to the plane of the diagram. This results in $s_\gamma = 2{\cdot}2{\cdot}2 = 8$. For the cluster ◯•◯, there is one extra circle, so $s_\gamma = 2^4 = 16$. The third diagram in figure shows two vertices connected by four lines. Any of the $4!$ permutations of these lines results in the same diagram. In addition, we may reflect about the vertical symmetry axis, interchanging the vertices, to obtain another symmetry operation. Thus $s_\gamma = 2 \cdot 4! = 48$. One might ask why we don't also count the planar rotation by π as a symmetry operation. The answer is that it is equivalent to a combination of a reflection and a permutation, so it is not in fact a distinct symmetry operation. (If it were distinct, then s_γ would be 96.) Finally, consider the last diagram in the figure, which resembles a sausage with three links joined at the ends into a circle. If we keep the vertices fixed, there are 8 symmetry operations associated with the freedom to exchange the two lines associated with each of the three sausages. There are an additional 6 symmetry operations associated with permuting the three vertices, which can be classified as three in-plane rotations by 0, $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$, each of which can also be combined with a reflection about the y-axis (this is known as the group C_{3v}). Thus, $s_\gamma = 8 \cdot 6 = 48$.

Now let us compute an expression for $F(\gamma)$ in terms of the clusters. We sum over all possible numbers of clusters at each order:

$$F(\gamma) = \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{\{m_\gamma\}} \frac{(4!)^N N!}{\prod_\gamma s_\gamma^{m_\gamma}\, m_\gamma!} \left(-\frac{\lambda}{4!}\right)^N \delta_{N,\sum_\gamma m_\gamma n_\gamma} = \exp\left(\sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma}\right) . \tag{1.173}$$

Thus,

$$\ln F(\gamma) = \sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma} ,$$

and *the logarithm of the sum over all diagrams is a sum over connected clusters*. It is instructive to work this out to order λ^2. We have, from the results of part (b),

$$F(\lambda) = 1 - \tfrac{1}{8}\,\lambda + \tfrac{35}{384}\,\lambda^2 + \mathcal{O}(\lambda^3) \quad \Longrightarrow \quad \ln F(\lambda) = -\tfrac{1}{8}\,\lambda + \tfrac{1}{12}\,\lambda^2 + \mathcal{O}(\lambda^3) .$$

λ	10	2	0.5	0.2	0.1	0.05	0.02
F	0.92344230	0.97298847	0.99119383	0.996153156	0.99800488	0.99898172	0.99958723
n^*	0.68	1.3	2.6	4.1	5.8	8.2	13

Table 1.1: $F(\lambda)$ and $n^*(\lambda)$ for part d.

Note that there is one diagram with $N = 1$ vertex, with symmetry factor $s = 8$. For $N = 2$ vertices, there are two diagrams, one with $s = 16$ and one with $s = 48$ (see fig. 1.17). Since $\frac{1}{16} + \frac{1}{48} = \frac{1}{12}$, the diagrammatic expansion is verified to order λ^2.

(d) We now have[27]

$$G(\lambda) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} dx \ \exp\left(-\frac{1}{2}\,x^2 - \frac{\lambda}{6!}\,x^6\right) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{\lambda}{6!}\right)^n \langle x^{6n}\rangle \equiv \sum_{n=0}^{\infty} C_n\,\lambda^n\ , \tag{1.174}$$

with

$$C_n = \frac{(-1)^n\,(6n)!}{(6!)^n\,n!\,2^{3n}\,(3n)!}\ .$$

Invoking Stirling's approximation, we find $\ln|C_n| \sim 2n\,\ln n - \left(2 + \ln\frac{5}{3}\right) n$. Thus we see that the magnitude of the contribution of the $n^{\rm th}$ term in the perturbation series is given by

$$C_n\,\lambda^n = (-1)^n \exp\left(2n\ln n - \left(2 + \ln\tfrac{10}{3}\right) n + n\ln\lambda\right)\ .$$

Differentiating, we find that this contribution is minimized for $n = n^*(\lambda) = (10/3\lambda)^{1/2}$. Via numerical integration using FORTRAN subroutines from QUADPACK, one obtains the results in Fig. 1.18 and Tab. 1.1.

The series for $G(\lambda)$ and for $\ln G(\lambda)$ again have diagrammatic interpretations. If we expand the sextic term in a power series, each power of λ brings an additional six powers of x. It is natural to represent each such sextet with as a vertex with six legs. At order N of the series expansion, we have N such vertices and $6N$ legs to contract. As before, each full contraction of the leg indices may be represented as a labeled diagram, which is in general composed of several disjoint connected clusters. If the cluster γ contains n_γ vertices, then for any diagram we again must have $N = \sum_\gamma m_\gamma n_\gamma$, where m_γ is the number of times the cluster γ appears. As with the quartic example, the number of ways of assigning labels to a given diagram is given by the total number of possible permutations $(6!)^N \cdot N!$ divided by a correction factor $\prod_\gamma s_\gamma^{m_\gamma} m_\gamma!$, where s_γ is the symmetry factor of the cluster γ, and the $m_\gamma!$ term accounts for the possibility of permuting among different labeled clusters of the same type γ.

[27] According to `Mathematica`, the $G(\lambda)$ has the analytic form $G(\lambda) = \pi\sqrt{u}\,\Big[\mathsf{Ai}^2(u) + \mathsf{Bi}^2(u)\Big]$, where $u = (15/2\lambda)^{1/3}$ and $\mathsf{Ai}(z)$ and $\mathsf{Bi}(z)$ are Airy functions. The definitions and properties of the Airy functions are discussed in §9.2 of the NIST *Handbook of Mathematical Functions*.

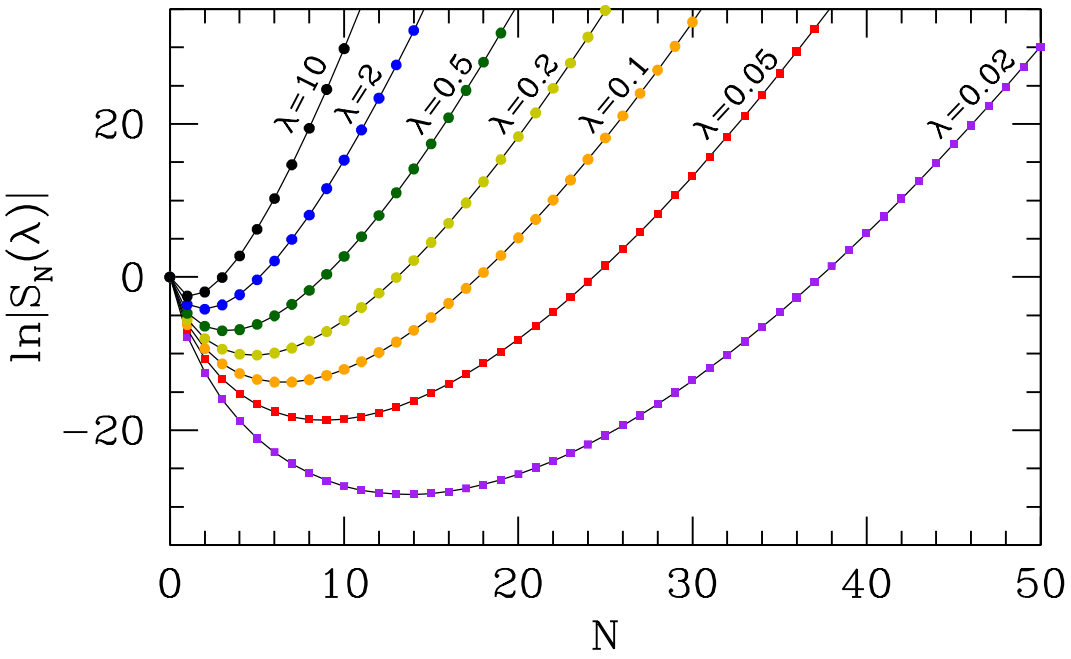
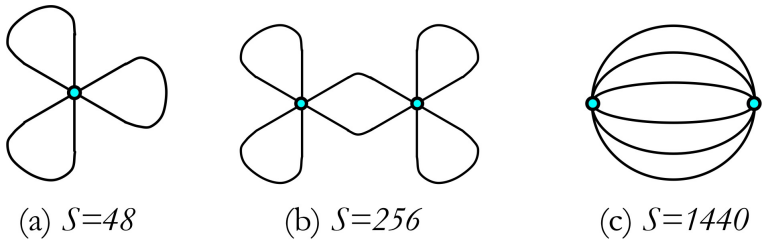


Figure 1.18: Logarithm of ratio of remainder after N terms $R_N(\lambda)$ to the value of the integral $G(\lambda)$ for the sextic theory, for various values of λ.

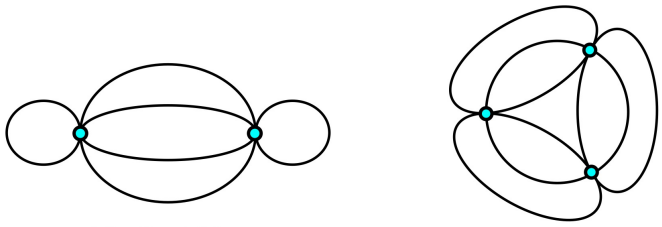
Examples of clusters and their corresponding symmetry factors are provided in Fig. 1.19. There is only one diagram with $n_\gamma = 1$, shown panel (a), resembling a three-petaled flower. To obtain $s_\gamma = 48$, note that each of the petals can be rotated by 180° about an axis bisecting the petal, yielding a factor of 2^3. The three petals can then be permuted, yielding an additional factor of $3!$. Hence the total symmetry factor is $s_\gamma = 2^3 \cdot 3! = 48$. Now we can see how dividing by the symmetry factor saves us from overcounting. In this case, we get $6!/s_\gamma = 720/48 = 15 = 5 \cdot 3 \cdot 1$, which is the correct number of contractions. For the diagram in panel (b), the four petals and the central loop can each be rotated about a symmetry axis, yielding a factor 2^5. The two left petals can be permuted, as can the two right petals. Finally, the two vertices can themselves be permuted. Thus, the symmetry factor is $s_\gamma = 2^5 \cdot 2^2 \cdot 2 = 2^8 = 256$. In panel (c), the six lines can be permuted ($6!$) and the vertices can be exchanged (2), hence $s_\gamma = 6! \cdot 2 = 1440$. In panel (d), the two outer loops each can be twisted by 180°, the central four lines can be permuted, and the vertices can be permuted, hence $s_\gamma = 2^2 \cdot 4! \cdot 2 = 192$. Finally, in panel (e), each pair of vertices is connected by three lines which can be permuted, and the vertices themselves can be permuted, so $s_\gamma = (3!)^3 \cdot 3! = 1296$.

Now let us compute an expression for $F(\gamma)$ in terms of the clusters. We sum over all possible numbers of clusters at each order:

$$G(\gamma) = \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{\{m_\gamma\}} \frac{(6!)^N N!}{\prod_\gamma s_\gamma^{m_\gamma} m_\gamma!} \left(-\frac{\lambda}{6!}\right)^N \delta_{N,\sum_\gamma m_\gamma n_\gamma} = \exp\left(\sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma}\right) . \tag{1.175}$$



(a) S=48 (b) S=256 (c) S=1440



(d) S=192 (e) S=1296

Figure 1.19: Diagrams and their symmetry factors for the $\frac{1}{6!}\lambda\, x^6$ zero-dimensional field theory.

Thus,

$$\ln G(\gamma) = \sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma}\ ,$$

and *the logarithm of the sum over all diagrams is a sum over connected clusters*. It is instructive to work this out to order λ^2. We have, from the results of part (a),

$$G(\lambda) = 1 - \frac{\lambda}{2^6\cdot 3} + \frac{7\cdot 11\cdot\lambda^2}{2^9\cdot 3\cdot 5} + \mathcal{O}(\lambda^3) \quad\Longrightarrow\quad \ln G(\lambda) = -\frac{\lambda}{2^6\cdot 3} + \frac{113\cdot\lambda^2}{2^8\cdot 3^2\cdot 5} + \mathcal{O}(\lambda^3)\ .$$

Note that there is one diagram with $N = 1$ vertex, with symmetry factor $s = 48$. For $N = 2$ vertices, there are three diagrams, one with $s = 256$, one with $s = 1440$, and one with $s = 192$ (see Fig. 1.19). Since $\frac{1}{256} + \frac{1}{1440} + \frac{1}{192} = \frac{113}{2^8 3^2 5}$, the diagrammatic expansion is verified to order λ^2.

(e) In quantum field theory (QFT), the vertices themselves carry space-time labels, and the contractions, *i.e.* the lines connecting the legs of the vertices, are *propagators* $G(x_i^\mu - x_j^\mu)$, where x_i^μ is the space-time label associated with vertex i. It is convenient to work in momentum-frequency space, in which case we work with the Fourier transform $\hat{G}(p^\mu)$ of the space-time propagators. Integrating over the space-time coordinates of each vertex then enforces total 4-momentum conservation at each vertex. We then must integrate over all the internal 4-momenta to obtain the numerical value for a given diagram. The diagrams, as you know, are associated with Feynman's approach to QFT and are known as Feynman diagrams. Our example here is equivalent to a $(0+0)$-dimensional field theory, *i.e.* zero space dimensions and zero time dimensions. There are then no internal 4-momenta to integrate over, and each propagator is simply a number rather than a function. The discussion above of symmetry factors s_γ carries over to the more general QFT case.

There is an important lesson to be learned here about the behavior of asymptotic series. As we have seen, if λ is sufficiently small, summing more and more terms in the perturbation series results in better and better results, until one reaches an optimal order when the error is minimized. Beyond this point, summing additional terms makes the result *worse*, and indeed the perturbation series diverges badly as $N \to \infty$. Typically the optimal order of perturbation theory is inversely proportional to the coupling constant. For quantum electrodynamics (QED), where the coupling constant is the fine structure constant $\alpha = e^2/\hbar c \approx \frac{1}{137}$, we lose the ability to calculate in a reasonable time long before we get to 137 loops, so practically speaking no problems arise from the lack of convergence. In quantum chromodynamics (QCD), however, the effective coupling constant is about two orders of magnitude larger, and perturbation theory is a much more subtle affair.

1.9 Appendix : Derivation of Ginzburg-Landau Functional

1.9.1 Discrete symmetry : $\mathbb{Z}_2$

We start with the Ising Hamiltonian,

$$H = -\tfrac{1}{2}\sum_{i,j} J_{ij}\,\sigma_i\,\sigma_j - \sum_i h_i\,\sigma_i + \tfrac{1}{2}\sum_i J_{ii} \quad , \tag{1.176}$$

where $\sigma_i \in \{+1, -1\}$ for all i. By subtracting the last term, we eliminate the constant contribution from the diagonal elements of the coupling matrix J_{ij}. You might well wonder why we simply don't specify that $J_{ii} = 0$ for all i from the outset. The reason for this will be made apparent at an important moment, and the suspense should motivate you to read further!

We define $K_{ij} \equiv J_{ij}/k_{\rm B}T$ and $g_i \equiv h_i/k_{\rm B}T$, so the density matrix is[28]

$$\varrho(\boldsymbol{\sigma}) = \exp\Big(\tfrac{1}{2}K_{ij}\,\sigma_i\,\sigma_j + g_i\,\sigma_i - \tfrac{1}{2}\,\mathsf{Tr}\,K\Big) \quad . \tag{1.177}$$

We now invoke the following identity:

$$\int_{-\infty}^{\infty} dm_1 \cdots \int_{-\infty}^{\infty} dm_N\ \exp\Big[-\tfrac{1}{2}K^{-1}_{ij} m_i\,m_j + b_i\,m_i\Big] = (2\pi\,\det K)^{1/2}\exp\Big[\tfrac{1}{2}K_{ij}\,b_i\,b_j\Big] \quad . \tag{1.178}$$

This licenses us to write

$$\varrho(\boldsymbol{\sigma}) = (2\pi\,\det K)^{-1/2}\,e^{-\frac{1}{2}\,\mathsf{Tr}\,K}\int d^N m\,\exp\Big[-\tfrac{1}{2}K^{-1}_{ij}\,m_i\,m_j + (m_i + g_i)\,\sigma_j\Big] \quad . \tag{1.179}$$

[28] In fact what we call $\varrho(\boldsymbol{\sigma})$ is really the diagonal element $\varrho(\boldsymbol{\sigma}\,|\,\boldsymbol{\sigma}) = \langle\,\boldsymbol{\sigma}\,|\,\varrho\,|\,\boldsymbol{\sigma}\,\rangle$ of the density matrix, whose off-diagonal matrix elements all vanish since all operators σ_i commute for our noninteracting Hamiltonian.

Now trace over the $\{\sigma_i\}$ to obtain the partition function:

$$Z = \int d^N m \, \exp\bigl[- \Phi(\boldsymbol{m}) \bigr] \quad , \tag{1.180}$$

where

$$\Phi(\boldsymbol{m}) = \Phi_0 + \tfrac{1}{2} \sum_{i,j} K^{-1}_{ij} \, m_i \, m_j - \sum_i \ln\bigl[\cosh(m_i + g_i) \bigr] \quad , \tag{1.181}$$

where $\Phi_0 = \frac{1}{2} \ln(2\pi \, \det K) + \frac{1}{2} \, \mathsf{Tr} \, K - N \ln 2$. Note also that $\langle \sigma_i \rangle = \bigl\langle \tanh(m_i + g_i) \bigr\rangle$.

We consider the situation where we are close to a second order phase transition and we apply a very weak field. In this case, we may expand $\Phi(\boldsymbol{m})$ in powers of the m_i and g_i to obtain

$$\Phi(\boldsymbol{m}) = \Phi_0 + \tfrac{1}{2} K^{-1}_{ij} \, m_i \, m_j + \sum_i \Bigl(- g_i \, m_i - \tfrac{1}{2} \, m_i^2 + \tfrac{1}{12} \, m_i^4 + \ldots \Bigr) \quad , \tag{1.182}$$

valid to fourth order in the $\{\phi_i\}$ and first order in the $\{g_i\}$. Now define the Fourier transforms,

$$m_i = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \hat{m}_{\boldsymbol{k}} \, e^{-\boldsymbol{k} \cdot \boldsymbol{R}_i} \qquad , \qquad \hat{m}_{\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_i m_i \, e^{-i \boldsymbol{k} \cdot \boldsymbol{R}_i} \tag{1.183}$$

as well as $\hat{K}(\boldsymbol{k}) \equiv \sum_{\boldsymbol{R}} K(\boldsymbol{R}) \, e^{-i \boldsymbol{k} \cdot \boldsymbol{R}}$, with $K_{ij} = K(\boldsymbol{R}_j - \boldsymbol{R}_i)$. We obtain

$$\Phi(\boldsymbol{m}) = \Phi_0 + \tfrac{1}{2} \sum_{\boldsymbol{k}} \bigl(\hat{K}^{-1}(\boldsymbol{k}) - 1 \bigr) \, |\hat{m}_{\boldsymbol{k}}|^2 + \tfrac{1}{12} \sum_i m_i^4 - \sum_i g_i \, m_i \quad , \tag{1.184}$$

where $\hat{K}^{-1}(\boldsymbol{k}) = 1/\hat{K}(\boldsymbol{k})$. On a d-dimensional cubic lattice with lattice constant a, assuming J_{ij} has isotropic nearest neighbor interactions only, we have

$$\hat{K}(\boldsymbol{k}) = 2K \sum_{\mu=1}^{d} \cos(k_\mu a) = 2dK - K \boldsymbol{k}^2 a^2 + \mathcal{O}(k_\mu^4) \quad , \tag{1.185}$$

where $K = J/k_{\rm B}T$ and J is the nearest neighbor coupling. One might think we could simply expand about the global maximum of $\hat{K}(\boldsymbol{k})$ lying at $\boldsymbol{k} = \boldsymbol{0}$ and then invert, to obtain

$$\hat{K}(\boldsymbol{k}) \approx \frac{1}{2dK} + \frac{1}{4d^2 K} \, \boldsymbol{k}^2 a^2 + \ldots \tag{1.186}$$

In the absence of an external field, the coefficient of the quadratic term in each m_i would then be $\frac{1}{2}\alpha$ with Landau parameter $\alpha = (2dK)^{-1} - 1$. This would predict an ordered phase for $a < 0$, or $k_{\rm B}T < zJ$, where $z = 2d$ is the lattice coordination number, which is indeed the classic mean field theory result.

However, there is a disturbing aspect to our derivation, which thus far has been exact. The determinant of K is given by $\det K = \prod_{\boldsymbol{k}} \hat{K}(\boldsymbol{k})$, and this vanishes when any of the eigenvalues

$\boldsymbol{K}(\boldsymbol{k})$ vanishes. It is easy to conjure up a vanishing eigenvalue, for example by taking $k_\mu = \pi/2a$ for all $\mu \in \{1, \ldots, d\}$. Indeed, the eigenspectrum of K extends over the interval $\hat{K}(\boldsymbol{k}) \in [-2dK, +2dK]$, hence there are negative eigenvalues as well. We can escape this problem by noting that our formulation allows us to include *arbitrary* diagonal elements in the matrix K. They are guaranteed not to affect our final result – if it is obtained exactly – since each diagonal element J_{ii} cancels in the expression for H in Eqn. 1.201. Thus, we may take

$$J_{ij} = J_0 + J_1 \, \delta_{|\boldsymbol{R}_i - \boldsymbol{R}_j|, a} \quad , \tag{1.187}$$

which yields

$$\hat{K}(\boldsymbol{k}) = K_0 + 2K_1 \sum_{\mu=1}^{d} \cos(k_\mu a) = K_0 + 2dK_1 - K_1 \boldsymbol{k}^2 a^2 + \ldots \quad . \tag{1.188}$$

Provided $K_0 > 2dK_1$, all the eigenvalues $\hat{K}(\boldsymbol{k})$ are positive, and we may safely expand about $\boldsymbol{k} = \boldsymbol{0}$, obtaining

$$\hat{K}^{-1}(\boldsymbol{k}) = \frac{k_{\rm B} T}{J_0 + zJ_1} \left(1 + \frac{J_1}{J_0 + zJ_1} \, \boldsymbol{k}^2 a^2 + \ldots \right) \quad . \tag{1.189}$$

The low temperature phase occurs for $\hat{K}(\boldsymbol{0}) < 1$, which means $k_{\rm B} T < J_0 + zJ_1$. Thus we have $k_{\rm B} T_{\rm c} = J_0 + zJ_1$, which unfortunately depends on the (almost[29]) arbitrary constant J_0. This is perhaps unsatisfying, since adding J_0 increases $T_{\rm c}$, and mean field theory already tends to overestimate the critical temperature. However, we shouldn't be too alarmed by any of this. We have no right to trust in low order expansions when it comes to predicting $T_{\rm c}$. However, from the standpoint of the renormalization group, we may expect the corresponding continuum field theory, with free energy

$$F[m] = \int d^d x \left(\tfrac{1}{2} a m^2 + \tfrac{1}{4} b m^4 + \tfrac{1}{2} \kappa \, |\boldsymbol{\nabla} m|^2 - hm \right) \quad , \tag{1.190}$$

to yield the correct critical exponents.

The Euler-Lagrange equations derived from the free energy in Eqn. 1.190 are

$$a\, m + b\, m^3 - \kappa \, \nabla^2 m = h \quad . \tag{1.191}$$

Consider the case $a < 0$ and $h = 0$, in which case the lowest free energy solution corresponds to $m(\boldsymbol{x}) = \pm m_0$ with $m_0 = \sqrt{|a|/b}$, and write $m(\boldsymbol{x}) = m_0 \, \mu(\boldsymbol{x})$. Let's solve for the case of a domain wall, where $\mu(\boldsymbol{x}) = \mu(x_1)$ is a function of only one coordinate and interpolates between the two vacuua, with $\mu(-\infty) = -1$ and $\mu(+\infty) = +1$. We then have

$$\xi^2 \, \frac{d^2 \mu}{dx_1^2} = -\mu + \mu^3 = -\frac{dU}{d\mu} \tag{1.192}$$

[29] Recall $J_0 > 2dJ_1$ must be imposed in order to avoid a vanishing of the determinant $\det K$.

where $\xi = \sqrt{\kappa/|a|}$ is the correlation length, and

$$U(\mu) = -\tfrac{1}{4}\big(1-\mu^2\big)^2 \quad . \tag{1.193}$$

Thus, we have

$$\sqrt{2}\,\xi\,\frac{d\mu}{dx_1} = 1-\mu^2 \quad , \tag{1.194}$$

with solution

$$\mu(x_1) = \tanh\big(x_1/\sqrt{2}\,\xi\big) \quad , \tag{1.195}$$

which satisfies the boundary conditions $\mu(\pm\infty) = \pm 1$. Thus, our domain wall profile is given by the expression $m(\boldsymbol{x}) = m_0\tanh\big(x_1/\sqrt{2}\,\xi\big)$, and the total domain wall energy is found to be

$$\Delta F = \frac{a^2}{4b}\,L^{d-1}\int\limits_{-\infty}^{\infty} dx_1\,\left\{\big(1-\mu^2\big)^2 + \left(\xi\,\frac{d\mu}{dx_1}\right)^2\right\} = \frac{\sqrt{8}\,a^2}{b}\,\xi\,L^{d-1} \quad , \tag{1.196}$$

which has the same L dependence as what we found in §1.5.1 for the Ising case.

1.9.2 Continuous symmetry : O(n)

A general n-dimensional unit vector $\boldsymbol{\phi} = (\phi^1,\ldots,\phi^n)$ may be expressed in terms of $n-2$ polar angles $\{\theta_1,\ldots,\theta_{n-2}$ and one azimuthal angle φ, *viz.*

$$\begin{aligned}
\phi^1 &= \sin\theta_1\cdots\sin\theta_{n-3}\sin\theta_{n-2}\cos\varphi \\
\phi^2 &= \sin\theta_1\cdots\sin\theta_{n-3}\sin\theta_{n-2}\sin\varphi \\
\phi^3 &= \sin\theta_1\cdots\sin\theta_{n-3}\cos\theta_{n-2} \\
&\;\;\vdots \\
\phi^{n-1} &= \sin\theta_1\cos\theta_2 \\
\phi^n &= \cos\theta_1 \quad .
\end{aligned} \tag{1.197}$$

Here $\theta_j \in [0,\pi]$ with $j \in \{1,\ldots,n-2\}$ and $\varphi \in [0,2\pi]$. The total solid angle in n dimensions is then

$$\Omega_n = 2\pi\int\limits_0^\pi d\theta_1\cdots\int\limits_0^\pi d\theta_{n-2}\;\sin^{n-2}\theta_1\cdots\sin\theta_{n-2} \tag{1.198}$$

Now

$$\int\limits_0^\pi d\theta\;\sin^p\theta = \frac{\sqrt{\pi}\,\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p+2}{2})} \quad , \tag{1.199}$$

and therefore

$$\Omega_n = 2\pi \cdot \frac{\sqrt{\pi}\,\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \cdot \frac{\sqrt{\pi}\,\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \cdots \frac{\sqrt{\pi}\,\Gamma(1)}{\Gamma(\frac{3}{2})} = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})} \quad . \tag{1.200}$$

This agrees with the following method of calculation:

$$\pi^{n/2} = \left(\int_{-\infty}^{\infty} dx\, e^{-x^2} \right)^n = \int d^n r\, e^{-\boldsymbol{r}^2} = \Omega_n \int_0^\infty dr\, r^{n-1}\, e^{-r^2} = \tfrac{1}{2}\,\Omega_n \int_0^\infty du\, u^{\frac{n}{2}-1}\, e^{-u} = \tfrac{1}{2}\,\Omega_n\,\Gamma(\tfrac{n}{2}) \quad .$$

Consider now the O(n) model Hamiltonian,

$$H = -\tfrac{1}{2} \sum_{i,j} J_{ij}\, \hat{\boldsymbol{\phi}}_i \cdot \hat{\boldsymbol{\phi}}_j - \sum_i \boldsymbol{h}_i \cdot \hat{\boldsymbol{\phi}}_i + \tfrac{1}{2} \sum_i J_{ii} \quad , \tag{1.201}$$

Using the Gaussian integral identity of Eqn. 1.178 n times, we obtain

$$Z = \int d^n m_1 \cdots \int d^n m_N\ \exp\big[-\Phi(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_N) \big] \tag{1.202}$$

where each $\boldsymbol{m}_j = (m_j^1, \ldots, m_j^n)$ is a real valued n-component vector, and

$$\Phi(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n) = \Phi_0 + \tfrac{1}{2} \sum_{i,j} K_{ij}\, \boldsymbol{m}_i \cdot \boldsymbol{m}_j - \sum_i \ln Y_n\big(|\boldsymbol{m}_i + \boldsymbol{g}_i|\big) \quad , \tag{1.203}$$

with $\Phi_0 = \frac{n}{2} \ln(2\pi\, \det K) + \frac{1}{2}\, \mathsf{Tr}\, K$, $\boldsymbol{g}_i = \boldsymbol{h}_i / k_{\rm B} T$, and

$$Y_n(y) = \ln \left(\int_0^\pi d\theta\, \big[\sin\theta\big]^{n-2}\, e^{y\cos\theta} \Big/ \int_0^\pi d\theta\, \big[\sin\theta\big]^{n-2} \right) = \frac{y^2}{2n} - \frac{y^4}{8n^2(n+1)} + \mathcal{O}(y^6) \quad . \tag{1.204}$$

Expanding to linear order in the fields $\{\boldsymbol{g}_i\}$ and quadratic order in the local order parameter values $\{\boldsymbol{m}_i\}$, we obtain

$$\Phi(\boldsymbol{m}_1, \ldots, \boldsymbol{m}_N) = \Phi_0 + \tfrac{1}{2} \sum_{\boldsymbol{k}} \big(\hat{K}^{-1}(\boldsymbol{k}) - \frac{1}{n} \big)\, |\hat{\boldsymbol{m}}_{\boldsymbol{k}}|^2 + \sum_i \frac{|\boldsymbol{m}_i|^4}{8n^2(n+1)} - \frac{1}{N} \sum_i \boldsymbol{g}_i \cdot \boldsymbol{m}_i + \ldots \tag{1.205}$$

which, in the continuum limit, corresponds to the dimensionless free energy density

$$\beta f(\boldsymbol{m}, \boldsymbol{\nabla}\boldsymbol{m}) = f_0 + \tfrac{1}{2}\, a(T)\, \boldsymbol{m}^2 + \tfrac{1}{4}\, b\, (\boldsymbol{m}^2)^2 + \tfrac{1}{2}\, \kappa\, (\boldsymbol{\nabla}\boldsymbol{m})^2 - \boldsymbol{j} \cdot \boldsymbol{m} + \ldots \quad . \tag{1.206}$$

Assuming $a(T) < 0$, we find the equilibrium value $|\boldsymbol{m}| = m_0 = \sqrt{|a|/b}$. Writing $\boldsymbol{m} = m_0 \hat{\boldsymbol{\omega}}$, where $\hat{\boldsymbol{\omega}}(\boldsymbol{x})$ is a unit vector field, and setting the external field to $\boldsymbol{j} = 0$, we find the free energy density for twists of the direction vector $\hat{\boldsymbol{\omega}}$ to be

$$f_{\rm twist}(\hat{\boldsymbol{\omega}}) = \tfrac{1}{2}\, \rho_{\rm s}\, (\boldsymbol{\nabla}\hat{\boldsymbol{\omega}})^2 \quad , \tag{1.207}$$

where the *spin stiffness* is $\rho_{\rm s} = \kappa\, m_0^2\, k_{\rm B} T$, which has dimensions EL^{2-d}. As shown in §1.5.1, a full twist in $\hat{\boldsymbol{\omega}}(\boldsymbol{x})$ in a single plane (ω^1, ω^2) across the width L of the system yields an energy which is proportional to L^{d-2}.

Chapter 2

Crystal Math

2.1 Classification of Crystalline Structures

Crystallography is the classification of spatially periodic structures according to their translational and rotational symmetries. It is a mature field[1], and the possible crystalline symmetries of two and three dimensional structures have been exhaustively classified. We shall not endeavor to prove, *e.g.*, that there are precisely 230 three-dimensional space groups. Rather, our proximate goal is to economically describe the most relevant aspects of the classification scheme, so that we may apply methods of group theory to analyze experimentally relevant physical processes in crystals.

2.1.1 Bravais Lattices

The notion of a *Bravais lattice* was discussed in §4.1.1. To review, a Bravais lattice $\mathcal{L}$ in d space dimensions is defined by a set of linearly independent vectors $\{\boldsymbol{a}_j\}$ with $j \in \{1, \ldots, d\}$ which define a *unit cell*. A general point $\boldsymbol{R}$ in the Bravais lattice is written as $\boldsymbol{R} = \sum_j n_j \boldsymbol{a}_j$, where each $n_j \in \mathbb{Z}$. The unit cell volume is given by

$$\Omega = \epsilon_{\mu_1 \cdots \mu_d} a_1^{\mu_1} \cdots a_d^{\mu_d} \quad , \tag{2.1}$$

and is by definition positive[2]. The choice of the vectors $\{\boldsymbol{a}_j\}$ is not unique, for one can always replace $\boldsymbol{a}_i$ with $\boldsymbol{a}_i + \boldsymbol{a}_j$ for any $j \neq i$, and, due to the antisymmetry of the determinant, Ω is unchanged. It is then conventional to choose the $\{\boldsymbol{a}_j\}$ so that they have the shortest possible

[1] Crystallography has enjoyed something of a resurgence in its relevance to recent theories of topological classification of electronic band structures. The interplay between symmetry and topology leads to a new classification for materials known as crystalline topological insulators, for example.

[2] One can always reorder the $\boldsymbol{a}_j$ so that $\Omega > 0$.

length, though even this prescription is not necessarily unique. The lattice of points $\{\boldsymbol{R}\}$ is called the *direct lattice*, and the $\{\boldsymbol{a}_j\}$ are the *elementary* (or *primitive*) *direct lattice vectors*.

One can then define the elementary (primitive) *reciprical lattice vectors*,

$$b_k^\nu \equiv \frac{2\pi}{\Omega}\,\epsilon_{\mu_1\cdots\mu_{k-1}\,\nu\,\mu_{k+1}\cdots\mu_d} a_1^{\mu_1}\cdots a_{k-1}^{\mu_{k-1}} a_{k+1}^{\mu_{k+1}}\cdots a_d^{\mu_d} \quad , \tag{2.2}$$

which satisfy $\boldsymbol{a}_i\cdot\boldsymbol{b}_j = 2\pi\,\delta_{ij}$. Indeed, we must have

$$\sum_{\mu=1}^{d} a_i^\mu\, b_j^\mu = 2\pi\,\delta_{ij} \qquad , \qquad \sum_{j=1}^{d} a_j^\mu\, b_j^\nu = 2\pi\,\delta_{\mu\nu} \quad , \tag{2.3}$$

because if the square matrices $A_{j,\mu}\equiv a_j^\mu$ and $B^{\mathsf{T}}_{\mu,j}\equiv b_j^\mu$ are inverses, they are each other's right as well as left inverse. For example, with $d=3$ we have $\Omega = \boldsymbol{a}_1\cdot\boldsymbol{a}_2\times\boldsymbol{a}_3$ and

$$\boldsymbol{b}_1 = \frac{2\pi}{\Omega}\,\boldsymbol{a}_2\times\boldsymbol{a}_3 \quad , \quad \boldsymbol{b}_2 = \frac{2\pi}{\Omega}\,\boldsymbol{a}_3\times\boldsymbol{a}_1 \quad , \quad \boldsymbol{b}_3 = \frac{2\pi}{\Omega}\,\boldsymbol{a}_1\times\boldsymbol{a}_2 \quad . \tag{2.4}$$

The set of vectors $\boldsymbol{K} = \sum_{j=1}^d m_j\,\boldsymbol{b}_j$, with each $m_i\in\mathbb{Z}$, is called the *reciprocal lattice*, $\hat{\mathcal{L}}$. The reciprocal lattice is therefore also a Bravais lattice, though not necessarily the same Bravais lattice as the direct lattice. For example, while the reciprocal lattice of a simple cubic lattice is also simple cubic, the reciprocal lattice of a body-centered cubic lattice is face-centered cubic. Constructing the reciprocal lattice of the reciprocal lattice, one arrives back at the original direct lattice. The unit cell volume of the reciprocal lattice is

$$\hat{\Omega} = \epsilon_{\mu_1\cdots\mu_d}\, b_1^{\mu_1}\cdots b_d^{\mu_d} = \frac{(2\pi)^d}{\Omega} \quad . \tag{2.5}$$

The repeating unit cells in the direct and reciprocal lattices may be written as the collection of points $\boldsymbol{r}$ and $\boldsymbol{k}$, respectively, where

$$\boldsymbol{r} = \sum_{j=1}^{d} x_j\,\boldsymbol{a}_j \qquad , \qquad \boldsymbol{k} = \sum_{j=1}^{d} y_j\,\boldsymbol{b}_j \quad , \tag{2.6}$$

where each $x_j, y_j\in[0,1]$. The symmetries of the direct and reciprocal lattices are more fully elicited by shifting each $\boldsymbol{r}$ and $\boldsymbol{k}$ point by a direct or reciprocal lattice vector so that it is as close as possible to the origin. Equivalently, sketch all the nonzero shortest direct/reciprocal lattice vectors emanating from the origin[3], and bisect each such vector with a perpendicular plane. The collection of points lying within all the planes will form the first *Wigner-Seitz cell* of the direct lattice, and the first *Brillouin zone* of the reciprocal lattice.

[3]There may be more than d shortest direct/reciprocal lattice vectors. For example, the triangular lattice is two-dimensional, but it has six nonzero shortest direct/reciprocal lattice vectors.

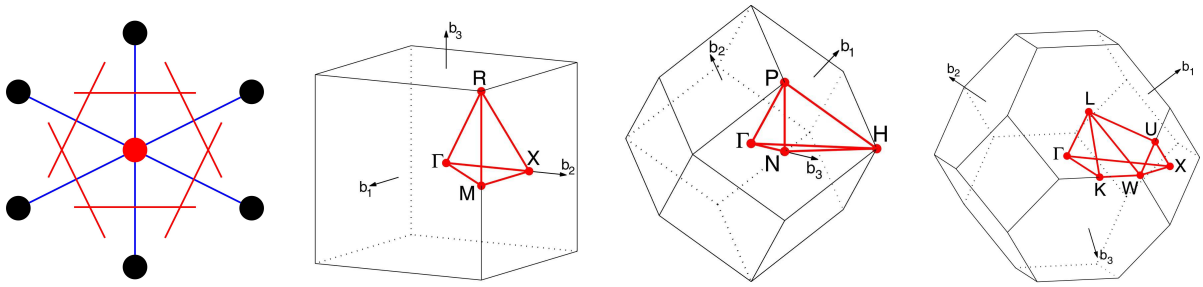


Figure 2.1: First panel shows construction of the first Wigner-Seitz cell or first Brillouin zone for the triangular lattice. Second, third, and fourth panels show first Brillouin zones for the simple cubic, body-centered cubic, and face-centered cubic direct lattices, respectively, with high symmetry points identified. Image credit: *Wikipedia* and Setyawan and Curtarolo, DOI: 10.1016/j.commatsci.2010.05.010.

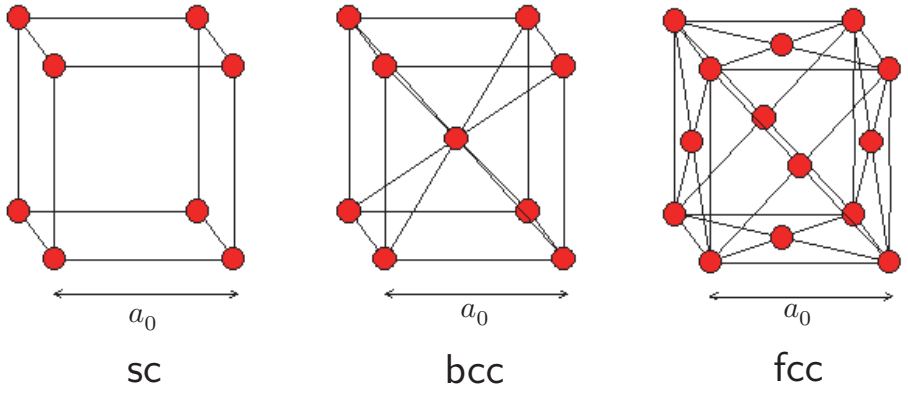


Figure 2.2: Simple cubic (sc), body-centered cubic (bcc), and face-centered cubic (fcc) lattices.

Finally, we cannot resist mentioning the beautiful and extremely important application of the Poisson summation formula to Bravais lattice systems:

$$\sum_{\boldsymbol{K}} e^{i\boldsymbol{K}\cdot\boldsymbol{r}} = \Omega \sum_{\boldsymbol{R}} \delta(\boldsymbol{r}-\boldsymbol{R}) \quad , \quad \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} = \hat{\Omega} \sum_{\boldsymbol{K}} \delta(\boldsymbol{k}-\boldsymbol{K}) \quad . \tag{2.7}$$

Example: fcc and bcc lattices

The primitive direct lattice vectors for the fcc structure may be taken as

$$\boldsymbol{a}_1 = \frac{a}{\sqrt{2}}(0,1,1) \quad , \quad \boldsymbol{a}_2 = \frac{a}{\sqrt{2}}(1,0,1) \quad , \quad \boldsymbol{a}_3 = \frac{a}{\sqrt{2}}(1,1,0) \quad . \tag{2.8}$$

The unit cell volume is $\Omega = \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3 = 2a^3$. Note that $|\boldsymbol{a}_j| = a$. Each FCC lattice point has twelve nearest neighbors, located at $\pm\boldsymbol{a}_1$, $\pm\boldsymbol{a}_2$, $\pm\boldsymbol{a}_3$, $\pm(\boldsymbol{a}_1-\boldsymbol{a}_2)$, $\pm(\boldsymbol{a}_2-\boldsymbol{a}_3)$, and $\pm(\boldsymbol{a}_3-\boldsymbol{a}_1)$.

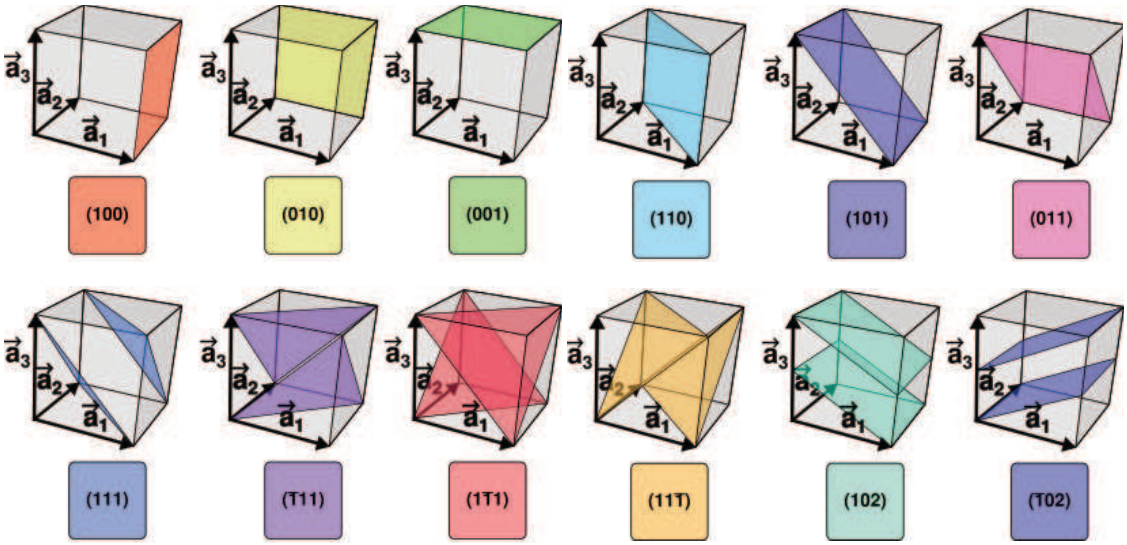


Figure 2.3: Examples of Miller planes. Image credit: Wikipedia.

The corresponding primitive reciprocal lattice vectors are

$$\boldsymbol{b}_1 = \frac{b}{\sqrt{3}}\,(-1,1,1) \quad , \quad \boldsymbol{b}_2 = \frac{b}{\sqrt{3}}\,(1,-1,1) \quad , \quad \boldsymbol{b}_3 = \frac{b}{\sqrt{3}}\,(1,1,-1)\,, \tag{2.9}$$

with $b = \sqrt{6}\,\pi/a$. These primitive vectors form a bcc structure, in which each site has eight nearest neighbors, located at $\pm\boldsymbol{b}_1$, $\pm\boldsymbol{b}_2$, $\pm\boldsymbol{b}_3$, and $\pm(\boldsymbol{b}_1+\boldsymbol{b}_2+\boldsymbol{b}_3)$. The simple, body-centered, and face-centered cubic structures are depicted in Fig. 2.2.

Be forewarned that in some texts, distances are given in terms of the side length of the cube. In the fcc lattice, if the cube has side length a_0, then the fcc lattice constant, *i.e.* the distance between nearest neighbor sites, is $a = a_0/\sqrt{2}$. Similarly, for the bcc case, if the cube has side length b_0, the corresponding bcc lattice constant is $b = \sqrt{3}\,b_0/2$.

In Fig. 2.1, the two rightmost panels show the first Brillouin zones corresponding to the bcc and fcc direct lattices, respectively. It follows that the same shapes describe the first Wigner-Seitz cells for the fcc and bcc lattices, respectively.

2.1.2 Miller indices

This eponymous notation system, first introduced by the British minerologist William H. Miller in 1839, provides a convenient way of indexing both directions and planes of points in a Bravais lattice. Briefly,

- $[\,h\,k\,l\,]$ represents a direction in the direct lattice given by the vector $h\boldsymbol{a}_1 + k\boldsymbol{a}_2 + l\boldsymbol{a}_3$. For negative numbers, one writes, *e.g.*, $\bar{2}$ instead of -2. Thus, $[\,1\,\bar{2}\,0\,]$ is the direction parallel

to $\boldsymbol{a}_1 - 2\boldsymbol{a}_2$. Only integers are used, so the direction parallel to $\frac{1}{4}\boldsymbol{a}_1 + \frac{1}{2}\boldsymbol{a}_2 - \frac{1}{3}\boldsymbol{a}_3$ is written as $[\,3\,6\,\bar{4}\,]$.

- $\langle\, h\,k\,l\,\rangle$ denotes the set of all directions which are related to $[\,h\,k\,l\,]$ by a rotational symmetry.
- $(\,h\,k\,l\,)$ represents a set of *lattice planes* which lie perpendicular to the vector $h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3$. Again, only integers are used, and any negative numbers are written with bars rather than minus signs.
- $\{\,h\,k\,l\,\}$ denotes all families of lattice planes related to $(\,h\,k\,l\,)$ by a rotational symmetry.

We can think of the Miller planes in terms of plane waves, *i.e.* as sets of points of constant phase $\phi(\boldsymbol{r}) = \boldsymbol{K}_{hkl} \cdot \boldsymbol{r}$, where $\boldsymbol{K}_{hkl} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3$ is a reciprocal lattice vector. If we write $\boldsymbol{R} = r\boldsymbol{a}_1 + s\boldsymbol{a}_2 + t\boldsymbol{a}_3$, we have $\phi(r,s,t) = 2\pi(hr + ks + lt) \equiv 2\pi N$, and thus the intersection of this plane with the $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ axes, which in general are not mutually orthogonal, lie at $N\boldsymbol{a}_1/h$, $N\boldsymbol{a}_2/k$, and $N\boldsymbol{a}_3/l$, respectively. In this way, one can identify the Miller indices of any lattice plane by taking the inverses of the respective coefficients and inverting them, then multiplying by the least common denominator if the results turn out to be fractional. From the formula $\exp(i\boldsymbol{K}\cdot\boldsymbol{r}) = 1$, we also see that the distance between consecutive Miller planes is $2\pi/|\boldsymbol{K}|$.

Cubic and hexagonal systems

For cubic systems, it is conventional to index the lattice planes based on the underlying simple cubic Bravais lattice. The bcc lattice is then viewed as a simple cubic lattice with a two element basis (see §2.1.5 below), and the fcc lattice as simple cubic with a four element basis. In hexagonal systems, typically one chooses the primary direct lattice vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ to subtend an angle of 120°, in which case $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ subtend an angle of 60°. Then defining $\boldsymbol{b}_0 \equiv \boldsymbol{b}_2 - \boldsymbol{b}_1$, we have that $\boldsymbol{b}_0$ is rotationally equivalent to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. Thus, if we define $i \equiv -(h+k)$, then we have the following rotations:

$$\begin{aligned} h\,\boldsymbol{b}_1 + k\,\boldsymbol{b}_2 &= R_{120^\circ}\big(k\,\boldsymbol{b}_1 + i\,\boldsymbol{b}_2\big) = R_{240^\circ}\big(i\,\boldsymbol{b}_1 + h\,\boldsymbol{b}_2\big) \\ &= R_{60^\circ}\big(\bar{i}\,\boldsymbol{b}_1 + \bar{h}\,\boldsymbol{b}_2\big) = R_{180^\circ}\big(\bar{h}\,\boldsymbol{b}_1 + \bar{k}\,\boldsymbol{b}_2\big) = R_{300^\circ}\big(\bar{k}\,\boldsymbol{b}_1 + \bar{i}\,\boldsymbol{b}_2\big) \quad . \end{aligned} \tag{2.10}$$

To reveal this rotational symmetry, the redundant fourth index i is used, and the Miller indices are reported as $(\,h\,k\,i\,l\,)$. The fourth index is always along the c-axis. The virtue of this four index notation is that it makes clear the relations between, *e.g.*, $(\,1\,1\,\bar{2}\,0\,) \equiv (\,1\,1\,0\,)$ and $(\,1\,\bar{2}\,1\,0\,) \equiv (\,1\,\bar{2}\,0\,)$, and in general

$$(\,h\,k\,i\,l\,) \;\to\; (\,\bar{i}\,\bar{h}\,\bar{k}\,l\,) \to\; (\,k\,i\,h\,l\,) \;\to\; (\,\bar{h}\,\bar{k}\,\bar{i}\,l\,) \;\to\; (\,i\,h\,k\,l\,) \;\to\; (\,\bar{k}\,\bar{i}\,\bar{h}\,l\,) \to (\,h\,k\,i\,l\,) \tag{2.11}$$

gives the full sixfold cycle.

2.1.3 Crystallographic restriction theorem

Consider a Bravais lattice and select one point as the origin. Now consider a general rotation $R \in \mathsf{SO}(3)$ and ask how the primary direct lattice vectors transform. If the Bravais lattice is symmetric under the operation R, then each $\boldsymbol{a}_j$ must transform into another Bravais lattice vector, *i.e.*

$$R_{\mu\nu}\, a_i^\nu = K_{ij}\, a_j^\mu \quad , \tag{2.12}$$

where K_{ij} is a matrix composed of integers. Now multiply both sides of the above equation by b_i^ρ and sum on the index i. From Eqn. 2.3, we have $a_i^\nu\, b_i^\rho = 2\pi\delta_{\nu\rho}$, hence $2\pi R_{\mu\rho} = K_{ij}\, a_j^\mu\, b_i^\rho$. Now take the trace over the indices μ and ρ, again invoking Eqn. 2.3, to get $\mathsf{Tr}\ R = \mathsf{Tr}\ K$. Now the trace of any matrix is invariant under similarity transformation, and in $d = 3$ dimensions, and if $R = R(\xi, \hat{\boldsymbol{n}})$ we can always rotate $\hat{\boldsymbol{n}}$ so that it lies along $\hat{\boldsymbol{z}}$, in which case

$$S\, R(\xi, \hat{\boldsymbol{n}})\, S^{-1} = \begin{pmatrix} \cos\xi & -\sin\xi & 0 \\ \sin\xi & \cos\xi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \tag{2.13}$$

in which case $\mathsf{Tr}\ R = 2\cos\xi + 1$. In $d = 2$ we have $\mathsf{Tr}\ R = 2\cos\xi$ for proper rotations. Thus, $\mathsf{Tr}\ R \in \mathbb{Z}$ is possible only for $\xi = 2\pi/n$ where $n = 1$, 2, 3, 4, or 6. Fivefold, sevenfold, *etc.* symmetries are forbidden! Note that it is perfectly possible to have a fivefold symmetric molecule, such as $\mathrm{C_{20}H_{10}}$, also known as corannulene. But when we insist on having both rotational as well as translational symmetries, the former are strongly restricted. Remarkably, there exists a family of three-dimensional structures, called *quasicrystals*, which exhibit forbidden fivefold or tenfold rotational symmetries. They elude the restriction theorem by virtue of not being true crystals, *i.e.* they are *quasi*periodic structures. See Fig. 2.4.

The result $\mathsf{Tr}\ R = \mathsf{Tr}\ K \in \mathbb{Z}$ is valid in all dimensions and does impose restrictions on the possible rotational symmetries. However, rotations in higher dimensions are in general not planar. Consider that it takes $d-1$ angles to specify an axis in d dimensions, but the dimension of $\mathsf{SO}(d)$ is $\frac{1}{2}d(d-1)$, so an additional $\frac{1}{2}(d-1)(d-2)$ parameters in addition to specifying an axis are required to fix an element of $\mathsf{SO}(d)$. For example, the four-dimensional F_4 lattice is a generalization of the three-dimensional bcc structure, consisting of two interpenetrating four-dimensional hypercubic lattices, and exhibits 12-fold rotational symmetries.

2.1.4 Enumeration of two and three-dimensional Bravais lattices

The complete classification of two and three Bravais lattices is as follows[4]. In two dimensions, there are four *lattice systems*: square, oblique, hexagonal, and rectangular. Of these, the rectangular system supports a subvariety called center rectangular, resulting in a total of five distinct two-dimensional Bravais lattices, shown in Fig. 2.5.

[4]To reinforce one's memory, there is even a song: https://ww3.haverford.edu/physics/songs/bravais.htm.

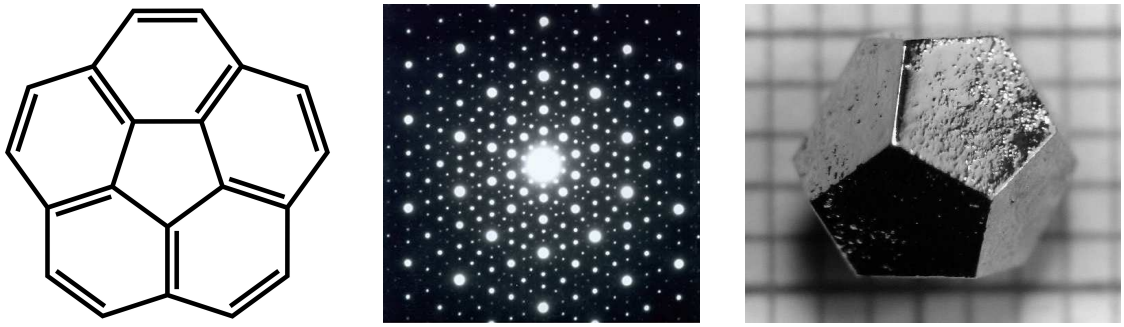


Figure 2.4: Left: Molecular structure of corannulene, $C_{20}H_{10}$. Center: Tenfold-symmetric diffraction pattern from a quasicrystalline alloy of aluminum, copper, and iron. Right: A Ho-Mg-Zn icosahedral quasicrystal forms a beautiful pentagonal dodecahedron (20 sites, 12 pentagonal faces, 30 edges, 3-fold coordinated), a structure dual to the icosahedron (12 sites, 20 triangular faces, 30 edges, 5-fold coordinated). Image credits: Wikipedia.

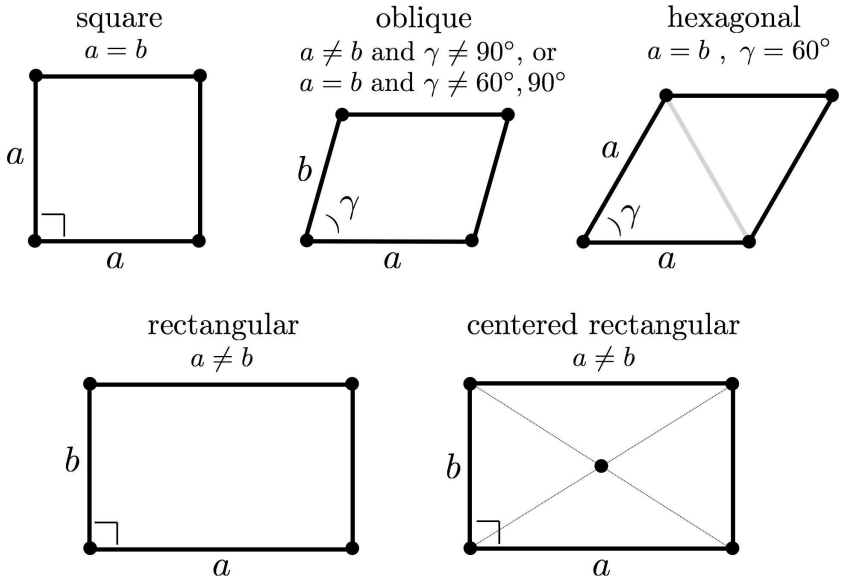


Figure 2.5: The five two-dimensional Bravais lattices.

In three dimensions, there are seven lattice systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic[5]. Of these, monoclinic supports two subvarieties or *types* (simple and base-centered), orthorhombic four subvarieties (simple, base-centered, body-centered, and face-centered), and cubic three subvarieties (simple, face-centered, and body-centered),

[5]The systematic enumeration of three-dimensional lattices based on symmetry was first done by M. L. Frankenheim in 1842. Frankenheim correctly found there were 32 distinct crystal classes, corresponding to the 32 distinct three-dimensional point groups, but he erred in counting 15 rather than 14 distinct lattices. A. L. Bravais, in 1845, was the first to get to 14, and for this he was immortalized. The identity of Frankenheim's spurious 15^{th} lattice remains unclear.

amounting to a grand total of 14 three-dimensional Bravais lattices:

(i) *Cubic* : this system is the most symmetric, with symmetry group $O_h \cong S_4 \times \mathbb{Z}_2$, which has 48 elements[6]. The $\mathbb{Z}_2$ factor arises from the *inversion* symmetry exhibited by all Bravais lattices. Recall inversion takes (x, y, z) to $(-x, -y, -z)$. The three cubic subvarieties (simple, body-centered, and face-centered) are depicted in the first three panels of 2.6.

(ii) *Tetragonal* : Lowering the cubic symmetry by stretching or compressing along one of the axes, one arrives at the tetragonal system, whose unit cell is a cubic rectangle with side lengths $a = b \neq c$. There are two sub-varieties: simple and body-centered, depicted in the left two panels of Fig. 2.8. Why is there not a face-centered subvariety as well? Because it is equivalent to the body-centered case[7]. The symmetry group is $D_{4h} \cong \mathbb{Z}_4 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

(iii) *Orthorhombic* : Further lowering the symmetry by stretching or compressing in along a second axis, we obtain the orthorhombic system. The only rotational symmetries are the three perpendicular mirror planes bisecting each of the unit cell sides, resulting in a $D_{2h} = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry. There are four subvarieties, depicted in Fig. 2.7: simple, base-centered, body-centered, and face-centered.

(iv) *Monoclinic* : Take an orthorhombic lattice and shear it so that the c-axis is no longer along $\hat{\boldsymbol{z}}$, but lies in the (y, z) plane at an angle $\beta \neq 90^\circ$ with respect to the horizontal. There are two distinct subvarieties, simple and base-centered, which are shown in the third and fourth panels of Fig. 2.8. The only remaining symmetries are reflection in the (y, z) plane and inversion, hence the symmetry group is $\mathbb{Z}_2 \times \mathbb{Z}_2$.

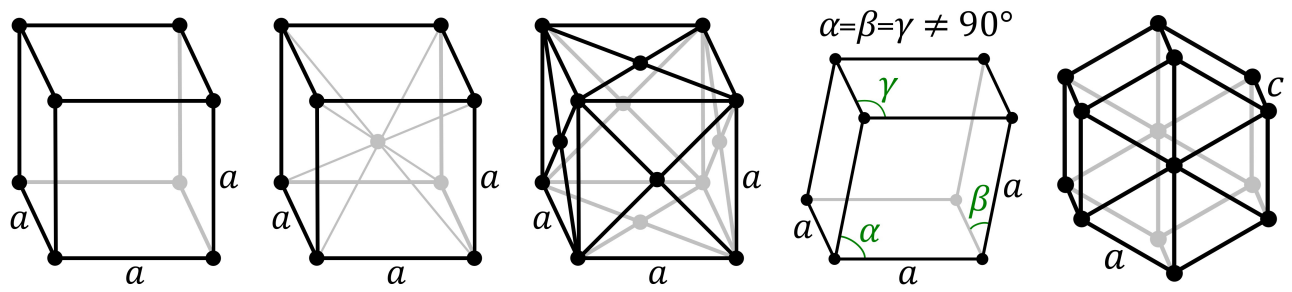
(v) *Triclinic* : Shearing in a second direction, one obtains the triclinic system, depicted in the right panel of Fig. 2.8. At least two of the angles $\vartheta_{ij} = \cos^{-1}\big(\hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_j\big)$ are not 90°, and all the axes are of unequal lengths. The only remaining symmetry is inversion, so the symmetry group is $\mathbb{Z}_2$.

(vi) *Trigonal* : Starting with the cubic system, rather than squashing it along one of its three orthogonal axes, imagine stretching it along the cube's diagonal. The resulting Bravais lattice is generated by three nonorthogonal primitive vectors which make the same angle with respect to one another, as depicted in the fourth panel of Fig. 2.6. The stretched cube diagonal becomes a threefold axis, and the symmetry group is D_{3d} , which is of order 12.

(vii) *Hexagonal* : Finally, we come to the hexagonal system, which is unrelated to the cube. The simple hexagonal lattice, depicted in the last panel of Fig. 2.6, is its only representative. Two of the primitive direct lattice vectors are of equal length a and subtend a relative angle of 60° or 120°. The third lies perpendicular to the plane defined by the first two, with an independent length c . The symmetry group is D_{6h} , which has 24 elements.

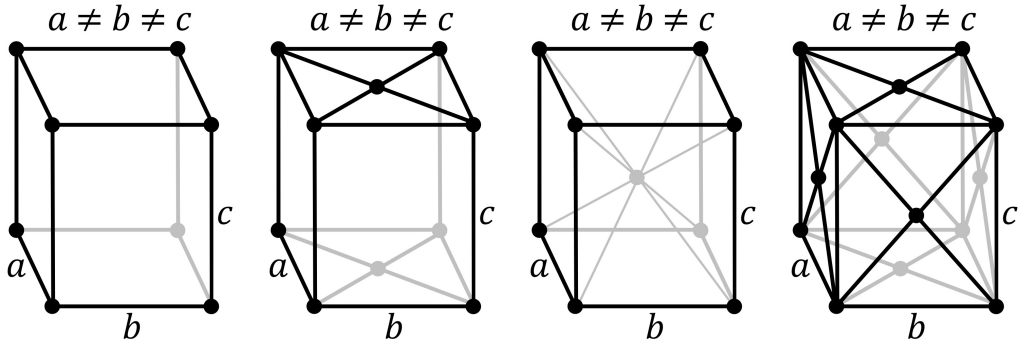
[6] Why is the symmetry group of the cube called O (or O_h with inversion)? Because the cube and the octahedron have the same symmetries. Hence O is sometimes called the octahedral group.

[7] See Ashcroft and Mermin, *Solid State Physics*, pp. 116-118.



(a) sc (b) bcc (c) fcc (d) trigonal (e) hex

Figure 2.6: Simple cubic, body-centered cubic, face-centered cubic, trigonal, and hexagonal Bravais lattices. Image credits: Wikipedia.



(a) so (b) baco (c) bco (d) fco

Figure 2.7: (Simple orthorhombic, base-centered orthorhombic, body-centered orthorhombic, and face-centered orthorhombic Bravais lattices. Image credits: Wikipedia.

2.1.5 Crystals

A Bravais lattice is a tiling of space with empty unit cells. We are in the position of a painter staring at a beautifully symmetric but otherwise empty canvas. The art with which we fill our canvas is the *crystalline unit cell*, and it consists of a number r of atoms or ions, where ions of species j are located at positions $\boldsymbol{\delta}_j$ relative to any given direct lattice point $\boldsymbol{R}$, with $j \in \{1, \ldots, r\}$. If the direct lattice points $\boldsymbol{R}$ themselves represent the positions of a class of ion, we write $\boldsymbol{\delta}_1 \equiv 0$. The set of vectors $\{\boldsymbol{\delta}_j\}$ is called a *basis*, and without loss of generality, we restrict the basis vectors so they do not lie outside the unit cell.

- ⋆ *In a crystal, ions of species j are located at positions $\boldsymbol{R} + \boldsymbol{\delta}_j$, where $\boldsymbol{R}$ is a Bravais lattice vector and $\boldsymbol{\delta}_j$ is a basis vector. All basis vectors are taken to lie within a single unit cell of the Bravais lattice.*

Obviously the existence of a basis, unless it is one of spherical symmetry with respect to each

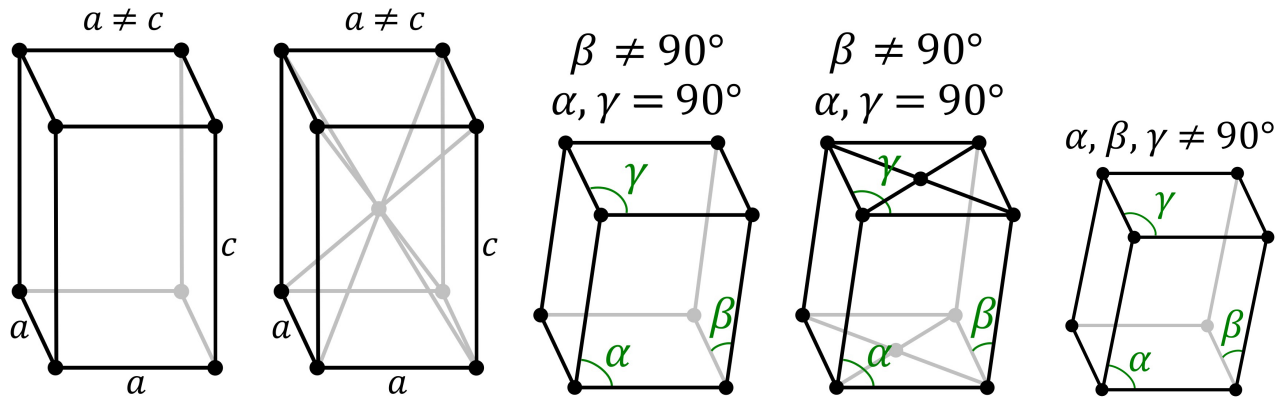


Figure 2.8: Simple tetragonal, body-centered tetragonal, monoclinic, base-centered monoclinic, and triclinic Bravais lattices. Image credits: Wikipedia.

Bravais lattice point, will have consequences for the allowed rotational symmetries of the crystal, in general reducing them to a subgroup of the symmetry group of the Bravais lattice itself. A vivid illustration of this is provided in Fig. 2.9 for the cubic lattice. When our canvas is completely blank, the cube is entirely white, and the symmetry group is O_h, with 48 elements, as shown in the middle bottom panel of the figure. If one of the reflection generators is broken, but all other generators are preserved, the symmetry is reduced from O_h to O, which has 24 elements. By breaking different symmetry operations, O_h can be reduced to the tetrahedral groups T_d and T_h, which also have 24 elements. Finally, each of O, T_d, and T_h may be broken down to the 12 element tetrahedral group T, depicted in the upper left panel. It all depends on how we paint the canvas.

As an example of a filled canvas, consider Fig. 2.10, which shows the unit cells of four high temperature cuprate superconductors. It is a good exercise to verify the stoichiometry in at least one example. Consider the unit cell for LSCO. The blue Cu ions at the top and bottom of the cell are each shared by eight of these cubic rectangular cells, so the eight Cu ions at the corners amount to one per cell. The Cu ion in the center belongs completely to this cell, so we have a total of two Cu per cell. Each of the eight green La/Sr ions lying along the vertical columns at the cell edges is shared by four cells, so they amount to a total of two per cell. The two La/Sr ions within the cell toward the top and bottom each count as one, for a total of four La/Sr per cell. Lastly, we come to the oxygen ions, shown in red. Each of the O ions along any of the 12 edges of the cell is shared by four cells. There are 16 such O sites, thus accounting for four O per cell. If you think about the periodic repetition of the cell, you should realize that each Cu ion is surrounded by six O ions arranged in an octahedron. There is also such an octahedron in the center of the cell, on which we now focus. Two of its O ions are displaced vertically with respect to the central Cu ion, and are therefore wholly part of our cell. The other four each lie in the center of a face, and are each shared by two cells. Thus, this central octahedron accounts for an additional four O ions, for a grand total of 8 per cell. Our final tally: two Cu, four La/Sr,

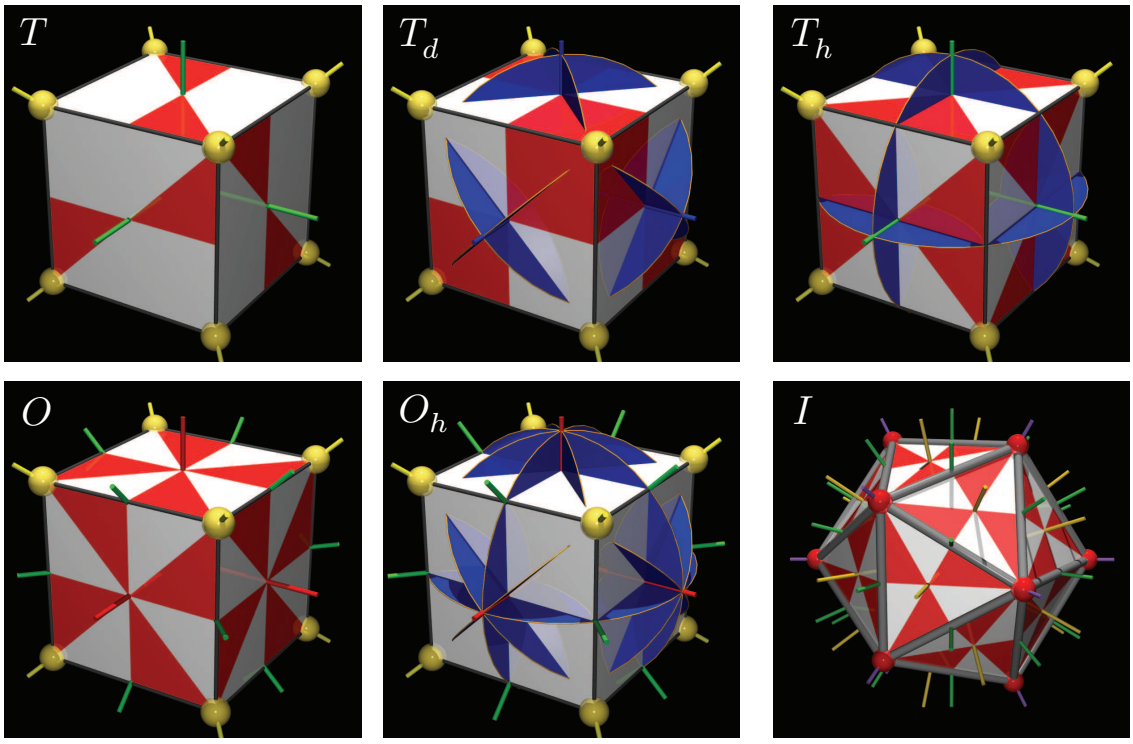


Figure 2.9: Tetrahedral, cubic, and icosahedral group symmetry operations. Twofold, threefold, fourfold, and sixfold axes are shown. The blue fins extend into discs, slicing the cube in two along reflection planes. Image credits: http://azufre.quimica.uniovi.es/d-MolSym/.

and eight O per cell, which is to say $La_{2-x}Sr_xCuO_4$. In the three other compounds, the oxygen stoichiometry is given as $4+\delta$ (Hg1201) or $6+\delta$ (YBCO and Ti2201). The deviation of δ from an integer value (either 0 or 1) accounts for the presence of oxygen vacancies[8].

In an electron diffraction experiment, an incident beam of electrons with wavevector $\boldsymbol{q}$ is scattered from a crystal, and the scattering intensity $I(\boldsymbol{k})$ as a function of the *wavevector transfer* $\boldsymbol{k} = \boldsymbol{q}' - \boldsymbol{q}$ is measured. If the scattering is elastic, $|\boldsymbol{q}'| = |\boldsymbol{q}|$, which means k is related to the scattering angle $\vartheta = \cos^{-1}(\hat{\boldsymbol{q}} \cdot \hat{\boldsymbol{q}}')$ by $k = 2q\sin(\frac{1}{2}\vartheta)$. Let us model the $T = 0$ density[9] of the crystal $\rho(\boldsymbol{r})$ as

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{R}} \sum_{j} c_j \, \delta(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\delta}_j) \quad , \tag{2.14}$$

where c_j is the weight for ionic species j. The total scattering intensity $I(\boldsymbol{k})$ is proportional to $|\hat{\rho}(\boldsymbol{k})|^2/N$, where $\hat{\rho}(\boldsymbol{k})$ is the Fourier transform of $\rho(\boldsymbol{r})$ and N is the total number of unit cells in

[8] It is a good exercise to determine the stoichiometry of these compounds based on the figures.
[9] What matters for electron diffraction is the electron density.

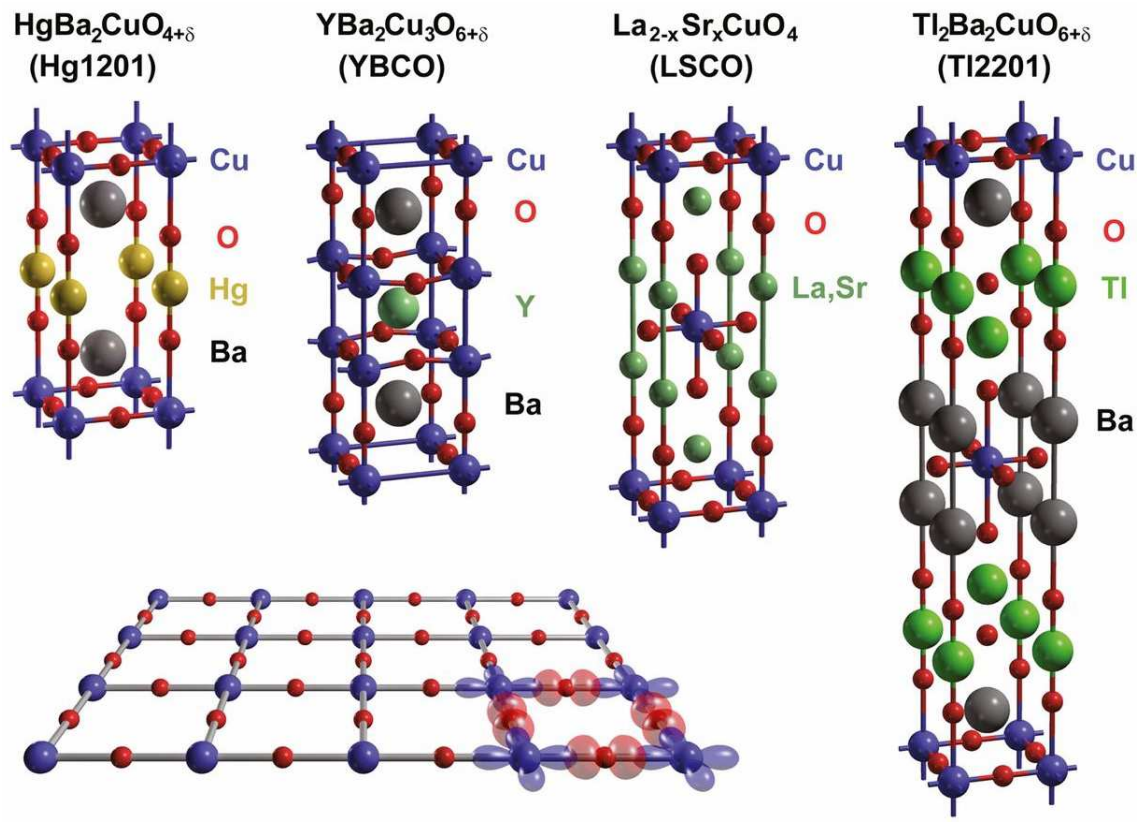


Figure 2.10: Unit cells of four high temperature cuprate superconductors. Lower left shows a sketch of the active electronic orbitals in the Cu-O planes. Image credit: N. Barišić *et al.*, *Proc. Nat. Acad. Sci.* **110**, 12235 (2013).

the crystal. Choosing units where the prefactor is unity, we have

$$\begin{aligned} I(\boldsymbol{k}) &= \frac{1}{N}\left|\hat{\rho}(\boldsymbol{k})\right|^2 = \frac{1}{N}\sum_{\boldsymbol{R},\boldsymbol{R}'} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \sum_{j,j'} c_j\, c_{j'}\, e^{-i\boldsymbol{k}\cdot(\boldsymbol{\delta}_j-\boldsymbol{\delta}_{j'})} \\ &= F(\boldsymbol{k}) \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} = \hat{\Omega}\sum_{\boldsymbol{K}} F(\boldsymbol{K})\,\delta(\boldsymbol{k}-\boldsymbol{K}) \quad , \end{aligned} \tag{2.15}$$

where we have invoked the Poisson summation formula of Eqn. 2.7, and where we have defined the *form factor*

$$F(\boldsymbol{K}) = \left|\sum_{j=1}^{r} c_j\, e^{-i\boldsymbol{K}\cdot\boldsymbol{\delta}_j}\right|^2 \quad . \tag{2.16}$$

Thus we expect δ-function *Bragg peaks* in the scattering intensity at values of the wavevector transfer equal to any reciprocal lattice vector $\boldsymbol{K}$. The form factor $F(\boldsymbol{K})$ modifies the intensity and can even lead to systematic *extinctions* of certain reciprocal lattice vectors. Consider, for example, a one-dimensional lattice with lattice spacing a and basis elements $\delta_1 = 0$ and $\delta_2 = \frac{1}{2}a$.

If $c_1 = c_2 \equiv c$, the form factor is

$$F(K) = |c|^2 \left|1 + e^{iKa/2}\right|^2 \quad . \tag{2.17}$$

This vanishes for $K = 2\pi j/a$ whenever j is odd. So the lesson here is that the $T = 0$ scattering intensity from a crystal is given by a sum of δ-functions and is singular whenever the wavevector transfer is equal to a reciprocal lattice vector. The presence of a basis modifies each Bragg peak by the form factor $F(\boldsymbol{K})$, which in some cases can even extinguish the peak completely[10].

2.1.6 Trigonal crystal system

While the trigonal point group D_{3d} is a normal subgroup of the hexagonal point group D_{6h}, the trigonal Bravais lattice does not result from an infinitesimal distortion of the simple hexagonal lattice. Contrast this situation with that for, *e.g.*, tetragonal *vis-a-vis* cubic, where a tetragonal lattice is obtained by an infinitesimal stretching along one of the principal axes of the cubic lattice. Any trigonal lattice, however, can be expressed as a hexagonal lattice with a three element basis. To see this, define the vectors

$$\boldsymbol{s}_1 = \tfrac{1}{\sqrt{3}}\, a\big(\tfrac{\sqrt{3}}{2}\,\hat{\boldsymbol{x}} - \tfrac{1}{2}\hat{\boldsymbol{y}}\big) \quad , \quad \boldsymbol{s}_2 = \tfrac{1}{\sqrt{3}}\, a\hat{\boldsymbol{y}} \quad , \quad \boldsymbol{s}_3 = \tfrac{1}{\sqrt{3}}\, a\big(-\tfrac{\sqrt{3}}{2}\,\hat{\boldsymbol{x}} - \tfrac{1}{2}\hat{\boldsymbol{y}}\big) \quad . \tag{2.18}$$

Then $\boldsymbol{a}_1 \equiv \boldsymbol{s}_1 - \boldsymbol{s}_3 = a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 \equiv \boldsymbol{s}_2 - \boldsymbol{s}_3 = a\big(\frac{1}{2}\hat{\boldsymbol{x}} + \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ are primitive DLVs for a two-dimensional hexagonal lattice. The vectors $\boldsymbol{d}_j \equiv \boldsymbol{s}_j + \frac{1}{3}\,c\hat{\boldsymbol{z}}$ for $j = 1, 2, 3$ then constitute three primitive DLVs for the trigonal lattice, each of length $d = \frac{1}{3}\sqrt{3a^2 + c^2}$. They also correspond to a three element basis within the first Wigner-Seitz cell of the simple hexagonal lattice. Conventionally, and equivalently, the three element basis may be taken to be $\boldsymbol{\delta}_1 = 0$, $\frac{1}{3}\boldsymbol{a}_1 + \frac{1}{3}\boldsymbol{a}_2 + \frac{1}{3}c\hat{\boldsymbol{z}}$, and $\frac{2}{3}\boldsymbol{a}_1 + \frac{2}{3}\boldsymbol{a}_3 + \frac{2}{3}c\hat{\boldsymbol{z}}$, all of which are associated with the hexagonal unit cell spanned by vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $c\hat{\boldsymbol{z}}$. Note that this is not a Wigner-Seitz cell, and its projection onto the (x, y) plane is a rhombus rather than a hexagon. Although describing the trigonal Bravais lattice as a hexagonal Bravais lattice with a three element basis might seem an unnecessary complication, in fact it proves to be quite convenient because two pairs of axes in the hexagonal system are orthogonal. Similarly, it is convenient to describe the bcc and fcc cubic lattices as simple cubic with a two and four element basis, respectively, to take advantage of the mutually orthogonal primitive direct lattice vectors of the simple cubic structure.

[10] It is a good exercise to compute $I(\boldsymbol{k})$ for the bcc and fcc structures when they are described in terms of a simple cubic lattice with a two or four element basis. The resulting extinctions limit the Bragg peaks to those wavevectors which are in the bcc or fcc reciprocal lattice.

CRYSTALLOGRAPHY	$d = 2$	$d = 3$
systems	4	7
lattices	5	14
point groups	10	32
space groups	17	230
symmorphic	13	73
non-symmorphic	4	157

Table 2.1: True Facts about two and three-dimensional crystallography.

2.1.7 Point groups, space groups and site groups

A group $\mathcal{P} \subset \mathsf{O}(3)$ of symmetry operations of a structure which leaves one point fixed is known as a *point group*[11]. The point group $\mathcal{P}_{\mathcal{L}}$ of a Bravais lattice is the group of rotational symmetries which fix any of the the Bravais lattice sites. This group is shared by all lattices in the same lattice system, and is known as the *holohedry* of the lattice.

In crystals, not every lattice site is equivalent. This may be due to the fact that different ions occupy different sites, but it is also the case for certain monatomic crystals, such as diamond, which consists of two interpenetrating fcc lattices that are not related by Bravais lattice translation. That is, the diamond structure is an fcc Bravais lattice with a two element basis. The full symmetry group of a crystal consists of both rotations and translations and is called the *space group* $\mathcal{S}$. A space group is a subgroup of the Euclidean group: $\mathcal{S} \subset \mathsf{E}(3)$, and a general space group operation $\left\{ g \,\middle|\, \boldsymbol{t} \right\}$ acts as

$$\left\{ g \,\middle|\, \boldsymbol{t} \right\} \boldsymbol{r} = g\,\boldsymbol{r} + \boldsymbol{t} \quad , \tag{2.19}$$

where $g \in \mathsf{O}(3)$. The identity element in $\mathcal{S}$ is $\left\{ \boldsymbol{E} \,\middle|\, \boldsymbol{0} \right\}$, where E is the identity in $\mathsf{O}(3)$, and the inverse is given by

$$\left\{ g \,\middle|\, \boldsymbol{t} \right\}^{-1} = \left\{ g^{-1} \,\middle|\, -g^{-1}\boldsymbol{t} \right\} \quad . \tag{2.20}$$

In order that $\mathcal{S}$ be a group, we must have that

$$\begin{aligned} \left\{ g_2 \,\middle|\, \boldsymbol{t}_2 \right\} \left\{ g_1 \,\middle|\, \boldsymbol{t}_1 \right\} \boldsymbol{r} &= \left\{ g_2 \,\middle|\, \boldsymbol{t}_2 \right\} \left(g_1\,\boldsymbol{r} + \boldsymbol{t}_1 \right) \\ &= g_2\,g_1\,\boldsymbol{r} + g_2\,\boldsymbol{t}_1 + \boldsymbol{t}_2 = \left\{ g_2\,g_1 \,\middle|\, g_2\,\boldsymbol{t}_1 + \boldsymbol{t}_2 \right\} \boldsymbol{r} \quad , \end{aligned} \tag{2.21}$$

is also in $\mathcal{S}$. This requires that the matrices g themselves form a group, called the *crystallographic point group* $\mathcal{P}$. For a Bravais lattice, $\mathcal{P} = \mathcal{P}_{\mathcal{L}}$, but in general a crystal is of lower symmetry than its underlying Bravais lattice, and the crystallographic point group is a subgroup of the holohedry: $\mathcal{P} \subset \mathcal{P}_{\mathcal{L}}$. Note that $\mathcal{S} \not\equiv \mathcal{P} \times \mathcal{T}$, *i.e.* the space group is not simply a direct product

[11] Mathy McMathstein says that a point group is a group of *linear isometries* which have a common fixed point. An isometry is a distance-preserving transformation on a metric space.

of the point group and the translation group, because multiplication of $(g, \boldsymbol{t}) \in \mathcal{P} \times \mathcal{T}$ satisfies $(g_2, \boldsymbol{t}_2)\,(g_1, \boldsymbol{t}_1) = (g_2\, g_1, \boldsymbol{t}_2 + \boldsymbol{t}_1)$. The abelian group $\mathcal{T} \cong \mathbb{Z}^d$ of Bravais lattice translations $\{\, E \,|\, \boldsymbol{R} \,\}$ forms an invariant subgroup of $\mathcal{S}$. If all the symmetry operations of a particular crystal can be written as $\{\, g \,|\, \boldsymbol{R} \,\}$, the crystal's space group is then said to be *symmorphic* and we write $\mathcal{S} = \mathcal{P} \rtimes \mathcal{T}$, where the symbol $\rtimes$ indicates a *semi-direct product* of two groups. *In a symmorphic crystal, one may choose an origin about which all point group symmetries are realized.*

However, it turns out that many crystals have space group elements $\{\, g \,|\, \boldsymbol{t} \,\}$ where $g \in \mathcal{P}$ but $\boldsymbol{t} \notin \mathcal{T}$. Rather, for these symmetry operations, $\boldsymbol{t}$ is a fraction of a Bravais lattice translation. In some cases, with a different choice of origin, these operations can be expressed as a rotation followed by Bravais lattice translation[12]. For crystals with *nonsymmorphic* space groups, however, there is no possible choice of origin about which all elements of $\mathcal{S}$ can be decomposed into a point group operation followed by a Bravais lattice translation. Two examples are shown in Fig. 2.11: the three-dimensional hexagonal close packed (hcp) structure, and the two-dimensional Shastry-Sutherland lattice. An hcp crystal is a simple hexagonal lattice with a two element basis. It occurs commonly in nature and describes, for example, the low temperature high pressure phase of ^{4}He just above its melting curve (about 25 atmospheres at $T = 0\,\mathrm{K}$). The primitive direct lattice vectors of the hcp structure are

$$\boldsymbol{a}_1 = \big(\tfrac{1}{2}\hat{\boldsymbol{x}} - \tfrac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)a \quad , \quad \boldsymbol{a}_2 = \big(\tfrac{1}{2}\hat{\boldsymbol{x}} + \tfrac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)a \quad , \quad \boldsymbol{a}_3 = c\hat{\boldsymbol{z}} \quad , \tag{2.22}$$

with $c = \sqrt{\frac{8}{3}}\,a$. The basis vectors are $\boldsymbol{\delta}_1 = 0$ and $\boldsymbol{\delta}_2 = \frac{1}{3}\boldsymbol{a}_1 + \frac{2}{3}\boldsymbol{a}_2 + \frac{1}{2}\boldsymbol{a}_3$. In the figure, A sublattice sites are depicted in red and B sublattice sites in blue. Note that the B sites lie in the centers of the up-triangles in each A sublattice plane, and displaced by half a unit cell in the $\hat{\boldsymbol{z}}$ diraction. The nonsymmorphic operation in the hcp point group is known as a *screw axis* and it involves a rotation by 60° about the $\hat{\boldsymbol{z}}$ axis through the centers of the A sublattice down triangles, followed by a translation by $\frac{1}{2}\boldsymbol{a}_3$. The crystallographic symbol for a screw operation is n_m, corresponding to a rotation by $2\pi/n$ followed by a translation by m/n of a unit cell along the screw axis. In the hcp structure, the screw operation is thus denoted by 6_3.

The second example is that of the Shastry-Sutherland lattice, which describes the $CuBO_3$ layers in the magnetic compound $SrCu_2(BO_3)_2$. Here we have four sublattices, and the nonsymmorphic operation is known as a *glide mirror*, which involves translation along a plane (or a line in two dimensions) by a half unit cell, followed by a reflection in the plane. See if you can spot the nonsymmorphic symmetry.

A third example is that of diamond, which consists of two interpenetrating fcc lattices, and has a zincblende structure shown in Fig. 2.16. Diamond possesses both a fourfold (4_1) screw axis as well as a glide mirror. While the point group is O_h, there is no point in the diamond lattice about which all operations in O_h are realized. The maximum symmetry at any site is T_d.

In a symmorphic crystal, it is always possible to find some origin within the structural unit

[12] In such cases, the putative nonsymmorphic operation is called *removable*. Otherwise, the nonsymmorphic operation is *essential*.

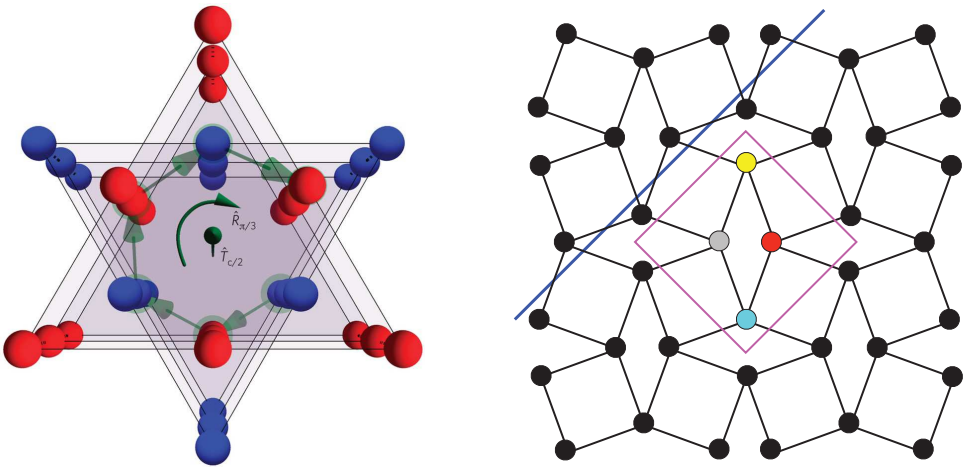


Figure 2.11: Two nonsymmorphic crystal lattices. Left: The hexagonal close packed lattice (space group $P6_3/mmc$) has a two site unit cell (red and blue) and a screw axis symmetry, given by a rotation by 60° followed by a translation of $\frac{1}{2}\boldsymbol{a}_3$ along the c-axis. The underlying Bravais lattice is simple hexagonal. Right: The Shastry-Sutherland lattice (space group $p4g$) has a four site unit cell (shown in center) and a glide mirror (blue line). Translation by half a unit cell along the mirror line followed by a mirror reflection is a lattice symmetry. The underlying Bravais lattice is square.

cell about which all point group symmetries are realized. In a Bravais lattice, this is true with respect to every lattice point, but obviously it is possible to choose an origin about which the group of rotational symmetries is reduced. For example, the point group of the square lattice is C_{4v}, but by choosing an origin in the center of one of the links the symmetry is reduced to C_{2v}. It is sometimes convenient to speak of the group of rotational symmetries with respect to a specific point $\boldsymbol{r}$ in the crystal structure. We call this the *site group* $\mathcal{P}(\boldsymbol{r})$. When $\boldsymbol{r} = \boldsymbol{R} + \boldsymbol{\delta}_j$ is a site in the crystal, *i.e.* a location of one of the ions, we may denote the site group as $\mathcal{P}(\boldsymbol{R}, j)$.

In a nonsymmorphic crystal, in general *no* sites will realize the symmetry of the point group $\mathcal{P}$. Consider, for example, the Shastry-Sutherland lattice in Fig. 2.11. Choosing the origin as the center of the magenta square unit cell, the site group is $\mathcal{P}(\boldsymbol{0}) = C_{2v}$. But the crystallographic point group for this structure is C_{4v}. Since $\mathcal{P}$ is the group of all rotational symmetries about all possible origins, necessarily $\mathcal{P}(\boldsymbol{r}) \subset \mathcal{P}$ for all sites $\boldsymbol{r}$.

Our crystallographer forbears have precisely tabulated for us all the possible lattices, point groups, and space groups in two and three dimensions (see Tab. 2.1). Proving these results is quite tedious, so we shall be content to take them as received wisdom. Note that a bit more than two thirds (157 out of 230) of the three-dimensional space groups are nonsymmorphic. Of those, all but two involve either a screw axis or a glide plane[13].

[13]Space groups no. 24 (also known as $I2_12_12_1$) and no. 199 ($I2_13$) have removable screw axes, but nevertheless there is no single origin about which every symmetry operation can be expressed as $\{\, g \,|\, \boldsymbol{t} \,\}$ with $g \in \mathcal{P}$ and

2.2 More on Point Groups

2.2.1 Standard notation for point group operations

A list of point group operations is provided in Tab. 2.2. We'll also start to use C_n to denote a group element, *i.e.* a rotation by $2\pi/n$ about a primary axis. If we need to distinguish this element from the cyclic group, which we've thus far also called C_n, we'll instead refer to the group as $\mathcal{C}_n$. Note that inversion can be written as $I = S_2$, and that I commutes with all elements of the point group $\mathcal{P}$, *i.e.* $I \in Z(\mathcal{P})$ is in the center of $\mathcal{P}$.

Any improper operation $g \in \mathsf{O}(3)$ has $\det(g) = -1$. This entails that g must have an eigenvalue $\lambda = -1$, and the corresponding eigenvector $\hat{\boldsymbol{m}}$, for which $g\hat{\boldsymbol{m}} = -\hat{\boldsymbol{m}}$, is known as a *reversal axis*. It also follows in all odd dimensions that if g is proper, *i.e.* if $\det(g) = +1$, then g has an eigenvalue $\lambda = +1$, and the corresponding eigenvector $\hat{\boldsymbol{n}}$ which satisfies $g\hat{\boldsymbol{n}} = \hat{\boldsymbol{n}}$ is an *invariant axis*. Improper elements of $\mathsf{O}(n)$ can be written as $Ig(\xi, \hat{\boldsymbol{n}})$, where I is the inversion operator. In even dimensions, the inversion I is equivalent to C_2, but one can form improper rotations via a reflection σ.

The rotoreflection operation is $S_n = \sigma_h^{-1} C_n = C_n \sigma_h^{-1}$. The reason we write σ_h^{-1} rather than σ_h has to do with what happens when we account for electron spin, in which case $\sigma_h^{-1} = \overline{E}\,\sigma_h$, where $\overline{E}$ is spinor reversal, *i.e.* rotation of the spinor component through 2π. Without spin, we have $\sigma_h^{-1} = \sigma_h$, and for n odd, one then has $(S_n)^n = \sigma_h$ and $(S_n)^{n+1} = C_n$, which says that if $S_n \in \mathcal{P}$ then so are both σ_h and C_n. If, on the other hand, n is even, this may not be the case.

2.2.2 Proper point groups

A proper point group $\mathcal{P}$ is a subgroup of $\mathsf{SO}(3)$[14]. The following are the proper point groups:

(i) *Cyclic groups* : The cyclic group C_n (order n) describes n-fold rotations about a fixed axis. The restriction theorem limits crystallographic cyclic groups to the cases $n = 1, 2, 3, 4$, and 6. Again, molecules, which have no translational symmetries, are not limited by the restriction theorem.

(ii) *Dihedral groups* : The group D_n (order $2n$) has a primary n-fold axis and n twofold axes perpendicular to the primary axis. Note that if one started with only one such perpendicular twofold axis, the C_n operations would generate all the others. For n even, the alternating twofold axes break up into two conjugacy classes, whereas for n odd there is only one such class.

$\boldsymbol{t} \in \mathcal{T}$.

[14] Two-dimensional point groups are much simpler to classify as they always involve at most a single rotation axis and/or a planar reflection. They form a subset of the three-dimensional point groups.

SYMBOL	OPERATION
E	identity
C_n	rotation through $2\pi/n$ about primary axis $\hat{\boldsymbol{n}}$; operator equivalent: $e^{-2\pi i \hat{\boldsymbol{n}}\cdot\boldsymbol{J}/n\hbar}$ where $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$
I	inversion ($\boldsymbol{r} \to -\boldsymbol{r}$) ; leaves spinor coordinates invariant and commutes with all other point group operations
σ	C_2 rotation followed by reflection in plane perpendicular to the axis of rotation ; equivalent to IC_2 or $C_2\, I$
σ_h	reflection in a 'horizontal' plane perpendicular to a primary axis
σ_v	reflection in a 'vertical' plane which contains a primary axis
σ_d	reflection in a 'diagonal' plane containing the primary axis of symmetry and which bisects the angle between neighboring twofold axes perpendicular to the primary axis
S_n	rotoreflection: $S_n = \sigma_h^{-1}\, C_n$, *i.e.* rotation by $2\pi/n$ followed by reflection in the plane perpendicular to that axis (note $I = S_2$)
$\bar{E}$	spinor rotation through 2π ; $\bar{E} = e^{-2\pi i \hat{\boldsymbol{n}}\cdot\boldsymbol{S}}$ ($S = \frac{1}{2}$) ; leaves spatial coordinates (x, y, z) invariant
$\bar{g}$	any point group operation g followed by $\bar{E}$

Table 2.2: Standard notation for point group operations.

(iii) *Tetrahedral, octahedral, and icosahedral groups* : When there is more than one n-fold axis with $n > 2$, the rotations about either axis will generate new axes. Geometrically, this process run to its conclusion traces out a regular spherical polygon when one traces the intersections of the successively-generated axes on the unit sphere. There are only five possible such regular polyhedra: tetrahedron, cube, octahedron, dodecahedron, and icosahedron. The second two have the same symmetry operations, as do the last two, so there are only three such groups: T, O, and I.

(iiia) *Tetrahedral group* : T is the symmetry group of proper rotations of the tetrahedron. Embedding the tetrahedron in a cube, as in Fig. 2.12, there are three two-fold axes through the cube faces, plus four threefold axes through the cube diagonals, for a total of 12 operations including the identity. Note $T \cong A_4$, the alternating group on four symbols.

(iiib) *Octahedral group* : O consists of all the symmetry operations from T plus 12 more, arising from six new twofold axes running through the centers of each edge, not parallel to any face, and six more operations arising from extending the twofold

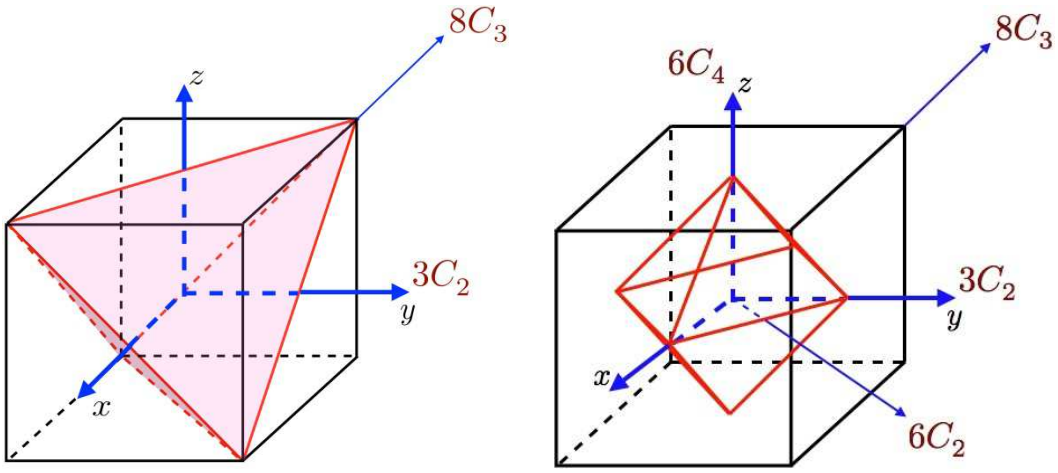


Figure 2.12: Left: Proper rotational symmetries of the tetrahedron, forming the group T. Right: Proper rotational symmetries of the octahedron (or cube), forming the group O.

axes through the faces to fourfold axes (see Fig. 2.9). So, 24 elements in all, shown in Fig. 2.12. Note $O \cong S_4$, the symmetric group on four symbols.

(iiic) *Icosahedral group* : I is the symmetry group of the dodecahedron or icosahedron. There are six fivefold axes, ten threefold axes, and 15 twofold axes, so including the identity there are $1 + 6 \cdot (5-1) + 10 \cdot (3-1) + 15 \cdot (2-1) = 60$ elements. We also have $I \cong A_5$, the alternating group on five symbols.

2.2.3 Commuting operations

The following operations commute:[15]

- Rotations about the same axis.
- Reflections in mutually perpendicular planes. In general the product of reflections in two planes which intersect at an angle α is $\sigma_v \, \sigma_{v'} = C(2\alpha)$, where the rotation is about the axis defined by their intersection line in the direction from the v' plane to the v plane. Thus $\sigma_{v'} = \sigma_v \, C(2\alpha)$.
- Rotations about perpendicular twofold axes: $C_2 \, C_2' = C_2' \, C_2 = C_2''$, where the resulting rotation is about the third perpendicular axis.
- A rotation C_n and a reflection σ_h in a plane perpendicular to the n-fold axis.

[15]See M. Lax, *Symmetry Principles in Solid State and Molecular Physics*, p. 54.

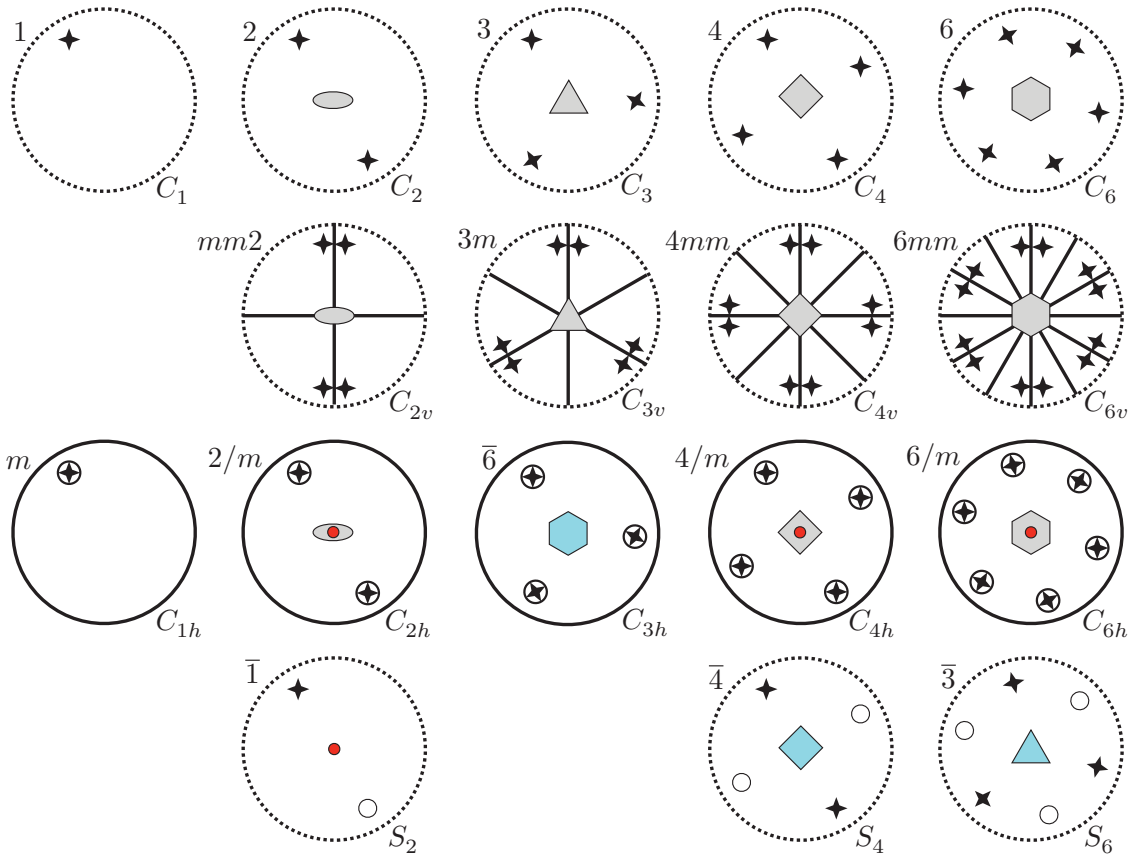


Figure 2.13: Stereographic projections of simple point groups C_n, C_{nv}, C_{nh}, and S_n. Dark lines correspond to reflection planes. C_{6v} looks like what I found the last time I sliced open a kiwi. Note $S_1 \cong C_{1h}$ and $S_3 \cong C_{3h}$. Adapted from Table 4.2 of M. Tinkham, *Group Theory and Quantum Mechanics*.

- Inversion I and any point group operation g (with g a rotation relative to the inversion point)
- A twofold rotation C_2 and a reflection σ_v in a plane containing the rotation axis.

2.2.4 Improper point groups

First, some notation. Since we will start to use C_n to denote the generator of rotations about the primary axis, we'll write $\mathcal{C}_n$ to denote the cyclic group with n elements. Similarly we'll use $\mathcal{S}_{2n}$ to denote the rotoreflection group. In addition to the proper point group $\mathcal{C}_2 = \{E, C_2\} \cong \mathbb{Z}_2$, we will also define two improper $\mathbb{Z}_2$ clones: $\mathcal{C}_i = \{E, I\}$, containing the identity and the inversion operation, and $\mathcal{C}_s \equiv \{E, \sigma_h\}$ containing the identity and the horizontal reflection σ_h. All will play a role in our ensuing discussion.

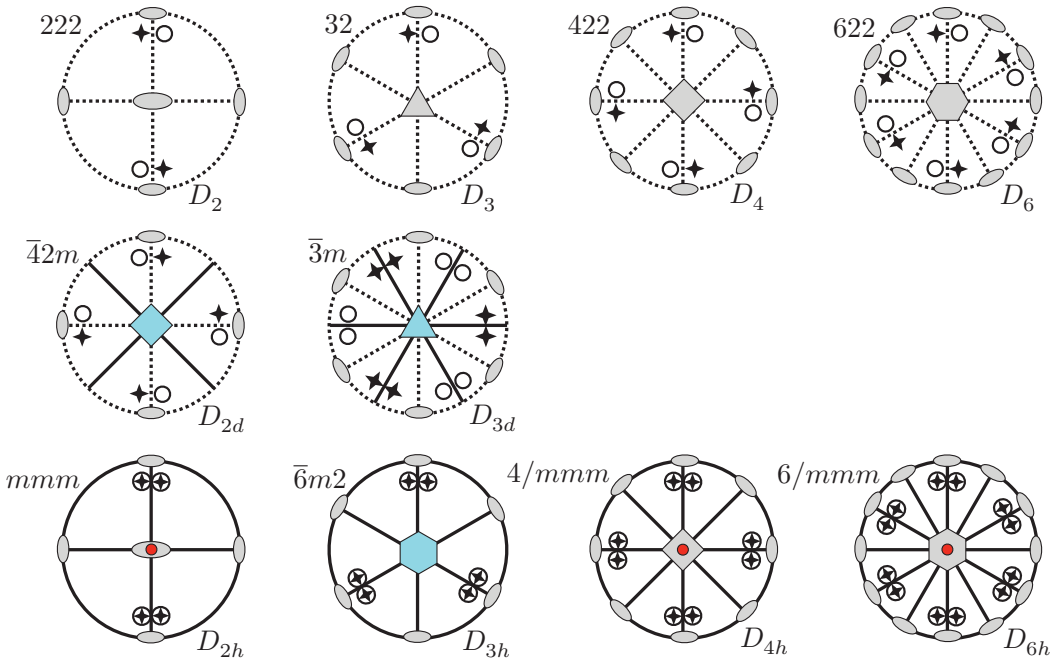


Figure 2.14: Stereograms of simple point groups D_n, D_{nd}, and D_{nh}. Dark lines correspond to reflection planes. Dashed lines correspond to 2-fold rotation axes. Adapted from Table 4.2 of M. Tinkham, *Group Theory and Quantum Mechanics*.

In §2.10 of Lax, the relations between proper and improper point groups are crisply discussed. Suppose a group G contains both proper and improper elements. We write $G = H \cup M$ where $H \triangleleft G$ is a normal subgroup containing all the proper elements, and M, which is not a group (no identity!) contains all the improper elements. Let $m \in M$ be any of the improper operations. Then $mH = M$ since multiplying any proper element by an improper one yields an improper element, and we conclude that H and M contain the same number of elements. Thus $G \cong H \cup mH$ and only one improper generator is needed. Since the inversion operator commutes with all elements of $\mathsf{O}(3)$, we can always form an improper group which contains I by constructing $G = H \cup IH = H \otimes \mathcal{C}_i$. If $G = H \cup mH$ does not contain the inversion operator I, we can always form a *proper* group $\widetilde{G} = H \cup mIH$ which is isomorphic to G. Consider the case of the improper point group $G = \mathcal{C}_{3v}$, where $H = \mathcal{C}_3 = \{E\,,\, C_3\,,\, C_3^{-1}\}$ and $m = \sigma_v$ is a vertical reflection plane containing the threefold axis[16]. Then $\tilde{G} = D_3$, which is proper, and which is isomorphic to C_{3v}. Finally, if G is proper, and if it contains an index two subgroup, *i.e.* a subgroup $H \subset G$ such that $N_G = 2N_H$, then we can construct $\tilde{G} = H \cup I\,(G \backslash H)$, where $G \backslash H$ is G with the elements from H removed. Then $\tilde{G}$ is an improper group with no inversion operation.

OK let's finally meet the improper point groups:

$\mathcal{S}_{2n}$: The rotoreflection group $\mathcal{S}_{2n}$ is a cyclic group of order $2n$ generated by $S_{2n} \equiv \sigma_h^{-1}\, C_{2n}$. In

[16] Acting with C_3 generates the additional vertical reflections: $C_3\,\sigma_v = \sigma_{v'}$ and $C_3\,\sigma_{v'} = \sigma_{v''}$.

the absence of spin, $\sigma^{-1} = \sigma$ for any reflection. Then for n odd, $(S_{2n})^n = \sigma_h\, C_2 = I$, hence $\mathcal{S}_n \cong \mathcal{C}_n \otimes \mathcal{C}_i$.

$\mathcal{C}_{nh}$: The $2n$ element group $\mathcal{C}_{nh} \cong \mathcal{C}_n \otimes \mathcal{C}_s$ has two commuting generators, C_n and σ_h. For n odd, $\mathcal{C}_{nh}$ is cyclic and is generated by the single element $\sigma_h\, C_n$.

$\mathcal{C}_{nv}$: The $2n$ element group $\mathcal{C}_{nv}$ has two noncommuting generators, C_n and σ_v, where σ_v is a reflection in a plane containing the n-fold axis. Repeated application of C_n creates $(n-1)$ additional vertical reflection planes. One has $\mathcal{C}_{nv} \cong D_n$.

D_{nh} : Adding a horizontal reflection plane to D_n, one obtains $D_{nh} \cong D_n \otimes \mathcal{C}_s$. For n even, one also has $D_{nh} \cong D_n \otimes \mathcal{C}_i$. The group has $4n$ elements.

D_{nd} : If instead of adding a horizontal reflection σ_h we add a 'diagonal' reflection σ_d in a plane which bisects the angle between neighboring twofold axes, we arrive at D_{nd}, which also has $4n$ elements.

T_d : Adding a reflection plane passing through one of the edges of the tetrahedron, we double the size of the tetrahedral group from 12 to 24. In Fig. 2.12, such a reflection might permute the vertices 3 and 4. Thus while $T \cong A_4$, we have $T_d \cong S_4$.

T_h : Adding inversion to the proper rotational symmetries of the tetrahedron, we obtain $T_h \cong T \otimes \mathcal{C}_i$, which has 24 elements.

O_h : Adding inversion to the proper rotational symmetries of the cube, we obtain $O_h \cong O \otimes \mathcal{C}_i$, which has 48 elements.

I_h : Adding inversion to the proper rotational symmetries of the icosahedron, we obtain $I_h \cong I \otimes \mathcal{C}_i$, which has 60 elements.

Stereographic projections of the simple point groups are depicted in Figs. 2.13 and 2.14. The subgroup structure of the point groups, which tells us the hierarchy of symmetries, is shown in Fig. 2.17.

Why don't we consider the rotoreflection groups $\mathcal{S}_n$ for n odd? Because for n odd, $\mathcal{S}_n \cong C_{nh}$. For n odd, both S_n and S_{2n} generate cyclic groups of order $2n$. It is perhaps instructive to consider the simplest nontrivial case, $n = 3$:

$$\begin{aligned} \mathcal{S}_3 &= \left\{E\,,\ \sigma_h\, C_3\,,\ C_3^{-1}\,,\ \sigma_h\,,\ C_3\,,\ \sigma_h\, C_3^{-1}\right\} \\ \mathcal{S}_6 &= \left\{E\,,\ \sigma_h\, C_6\,,\ C_3\,,\ \sigma_h\, C_2\,,\ C_3^{-1}\,,\ \sigma_h\, C_6^{-1}\right\} \quad . \end{aligned} \tag{2.23}$$

We see that $\mathcal{C}_{3h}$, which is generated by the pair $(C_3\,,\ \sigma_h)$, contains the same elements as $\mathcal{S}_3$. This result holds for all odd n, because in those cases $\sigma_h \in \mathcal{S}_n$.

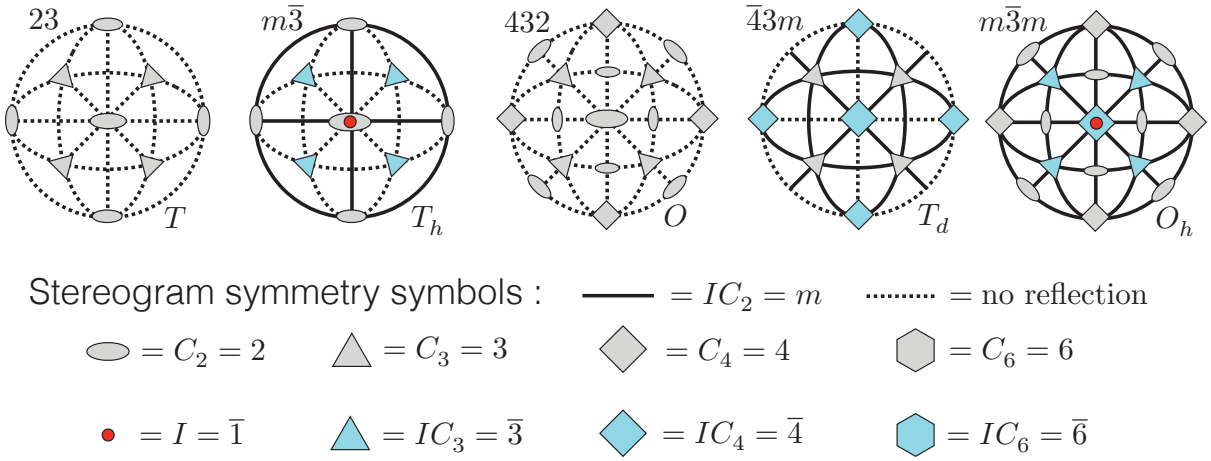


Figure 2.15: Stereograms of tetrahedral and cubic point groups and legend for symbols.

2.2.5 The ten two-dimensional point groups

There are ten two-dimensional point groups, listed in Tab. 2.3. As the only allowed elements are 2, 3, 4, and 6-fold rotations about the z-axis, plus vertical (line) mirrors, the only possible groups are C_1 , C_2 , C_3 , C_4 , C_6 and their mirrored extensions C_{1v} , C_{2v} , C_{3v} , C_{4v} , and C_{6v}. Note that the group C_{1v} is equivalent to C_{1h} , since in $d = 3$ both have a single reflection plane.

2.2.6 The achiral tetrahedral group, T_d

Many materials such as GaAs occur in an AB zincblende structure, which consists of two interpenetrating fcc lattices A and B, separated by $(\frac{a}{4}, \frac{a}{4}, \frac{a}{4})$, where a is the side length of the cube; see Fig. 2.16. As the figure shows, the B sublattice sites within the cube form a tetrahedron. The

LATTICE SYSTEM	POINT GROUPS			
oblique	C_1	C_2		
rectangular	C_{1h}	C_{2v}		
centered rectangular	C_{1h}	C_{2v}		
square	C_4	C_{4v}		
hexagonal	C_3	C_{3v}	C_6	C_{6v}

Table 2.3: The ten two-dimensional point groups. Note $C_{1h} \cong C_{1v}$.

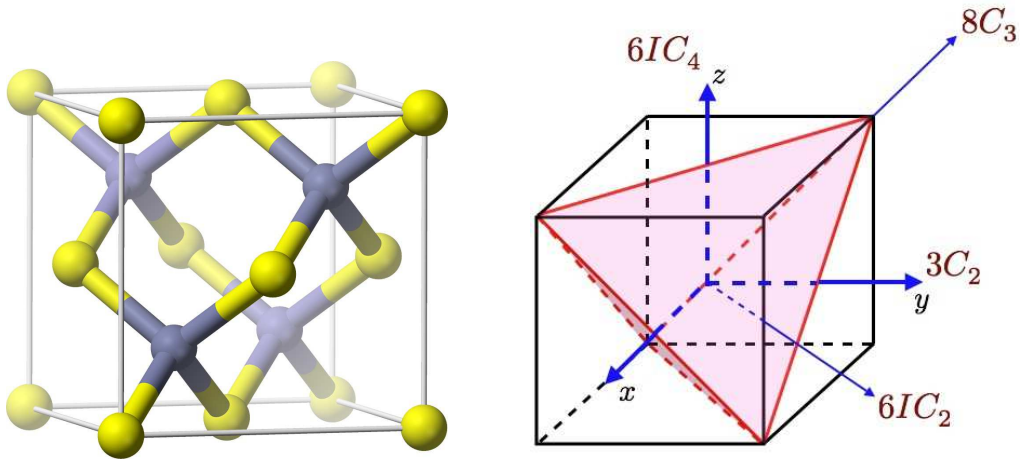


Figure 2.16: Left: The zincblende structure. Right: Proper and improper elements of the group T_d.

crystallographic point group for this structure is T_d, the achiral tetrahedral group. A noteworthy feature is that the zincblende structure has no center of inversion symmetry, hence $I \notin T_d$.

If all the atoms are identical, *i.e.* A = B, then we get the diamond structure, which is the structure of silicon and of course carbon diamond. The diamond lattice is inversion symmetric, with the point of inversion halfway between the A and B sublattice sites. The point group for diamond is the cubic group O_h. This might be surprising upon staring at the structure for a time, because it doesn't possess a cubic symmetry. However, the space group for diamond is non-symmorphic – it has a glide plane and a screw axis.

The group T_d has 24 elements; these are listed in Tab. 2.5. Its character table is provided in Tab. 2.4. These are arranged in five group *classes*. One class is the identity, E. Another class consists of three 180° rotations about the $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$ axes, respectively. A third class, with eight

T_d	E	$8C_3$	$3C_2$	$6\sigma_d$	$6S_4$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
E	2	-1	2	0	0
T_1	3	0	-1	-1	1
T_2	3	0	-1	1	-1

Table 2.4: Character table for the group T_d.

class	x	y	z	$g \in \mathsf{O}(3)$	class	x	y	z	$g \in \mathsf{O}(3)$
E	x	y	z	1	$6IC_2$	$-y$	$-x$	z	$IR_{[110]}(\pi) = \sigma_{(110)}$
$3C_2$	x	$-y$	$-z$	$R_{[100]}(\pi)$	$(6\sigma_d)$	y	x	z	$IR_{[1\bar{1}0]}(\pi) = \sigma_{(1\bar{1}0)}$
	$-x$	y	$-z$	$R_{[010]}(\pi)$		$-z$	y	$-x$	$IR_{[101]}(\pi) = \sigma_{(101)}$
	$-x$	$-y$	z	$R_{[001]}(\pi)$		z	y	x	$IR_{[10\bar{1}]}(\pi) = \sigma_{(10\bar{1})}$
$8C_3$	z	x	y	$R_{[111]}(+\frac{2\pi}{3})$		x	$-z$	$-y$	$IR_{[011]}(\pi) = \sigma_{(011)}$
	y	z	x	$R_{[111]}(-\frac{2\pi}{3})$		x	z	y	$IR_{[01\bar{1}]}(\pi) = \sigma_{(01\bar{1})}$
	z	$-x$	$-y$	$R_{[1\bar{1}1]}(+\frac{2\pi}{3})$	$6IC_4$	$-x$	z	$-y$	$IR_{[100]}(+\frac{\pi}{2}) = \sigma_{(100)}R_{[100]}(-\frac{\pi}{2})$
	$-y$	$-z$	x	$R_{[1\bar{1}1]}(-\frac{2\pi}{3})$	$(6S_4)$	$-x$	$-z$	y	$IR_{[100]}(-\frac{\pi}{2}) = \sigma_{(100)}R_{[100]}(+\frac{\pi}{2})$
	$-z$	x	$-y$	$R_{[11\bar{1}]}(+\frac{2\pi}{3})$		$-z$	$-y$	x	$IR_{[010]}(+\frac{\pi}{2}) = \sigma_{(010)}R_{[010]}(-\frac{\pi}{2})$
	y	$-z$	$-x$	$R_{[11\bar{1}]}(-\frac{2\pi}{3})$		z	$-y$	$-x$	$IR_{[010]}(-\frac{\pi}{2}) = \sigma_{(010)}R_{[010]}(+\frac{\pi}{2})$
	$-z$	$-x$	y	$R_{[1\bar{1}\bar{1}]}(+\frac{2\pi}{3})$		y	$-x$	$-z$	$IR_{[001]}(+\frac{\pi}{2}) = \sigma_{(001)}R_{[001]}(-\frac{\pi}{2})$
	$-y$	z	$-x$	$R_{[1\bar{1}\bar{1}]}(-\frac{2\pi}{3})$		$-y$	x	$-z$	$IR_{[001]}(-\frac{\pi}{2}) = \sigma_{(001)}R_{[001]}(+\frac{\pi}{2})$

Table 2.5: Table of elements and classes for T_d. Here $I\colon (x,y,z) \to (-x,-y,-z)$ is inversion and $\sigma_{(h,k,l)}$ is a reflection in the plane perpendicular to $h\hat{\boldsymbol{x}} + k\hat{\boldsymbol{y}} + l\hat{\boldsymbol{z}}$. For example $\sigma_{(110)}\colon (x,y,z) \to (-y,-x,z)$. Note that each fourfold rotoinversion can be expressed as a rotoreflection, *i.e.* $6IC_4 \cong 6S_4$, comprising a $\pm\frac{\pi}{2}$ rotation about one of the C_2 axes followed by a reflection in the plane perpendicular to that axis. Similarly, each twofold rotoinversion can be expressed as a reflection $6IC_2 \cong 6\sigma_d$ in one of the six diagonal mirror planes.

elements, consists of rotations by $\pm 120^\circ$ about each of the four body diagonals. This amounts to 12 group operations, all of which are proper rotations. The remaining 12 elements involve the inversion operator, I, which takes (x,y,z) to $(-x,-y,-z)$, and are therefore improper rotations, with determinant -1.[17] These elements fall into two classes, one of which consists of 180° rotations about diagonals parallel to one of the sides of the cube (*e.g.* the line $y = x$, $z = 0$), followed by inversion. The last class consists of rotations by $\pm\frac{\pi}{2}$ about $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$, also followed by an inversion, or, respectively, rotations by $\mp\frac{\pi}{2}$ about $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{x}}$ followed by a reflection in the plane perpendicular to the rotation axis.

Rotoreflections *versus* **rotoinversions**

In general every rotoinversion $IR(\hat{\boldsymbol{n}},\alpha)$ may be written as a rotoreflection $\sigma_{\hat{\boldsymbol{n}}}\,R(\hat{\boldsymbol{n}},-\alpha)$ where $\sigma_{\hat{\boldsymbol{n}}}$ is a reflection in the plane perpendicular to $\hat{\boldsymbol{n}}$. Both are improper rotations, *i.e.* elements

[17] Note that I itself is not an element of T_d.

Γ	d_Γ	basis functions ψ^Γ_μ for T_d	basis functions ψ^Γ_μ for O
A_1	1	$1 \quad \text{or} \quad xyz$	1
A_2	1	$x^4(y^2-z^2)+y^4(z^2-x^2)+z^4(x^2-y^2)$	xyz
E	2	$\left\{\sqrt{3}\,(x^2-y^2)\,,\, 2z^2-x^2-y^2\right\}$	$\left\{\sqrt{3}\,(x^2-y^2)\,,\, 2z^2-x^2-y^2\right\}$
T_1	3	$\left\{x\,(y^2-z^2)\,,\, y\,(z^2-x^2)\,,\, z\,(x^2-y^2)\right\}$	$\left\{x\,,\ y\,,\ z\right\}$
T_2	3	$\left\{x\,,\, y\,,\, z\right\}$	$\left\{yz\,,\, zx\,,\, xy\right\}$

Table 2.6: Irreducible representations and basis functions for T_d and O.

$g \in \mathsf{O}(3)$ with $\det g = -1$. Distinguishing these operations is useful when there is a single preferred rotation axis[18].

2.2.7 Tetrahedral *vs.* octahedral symmetry

In the case of the octahedral group, O, the inversion operation is not included in the last two classes, and they are written as $6C_2$ and $6C_4$, respectively. The symmetry operations of O are depicted in fig. 2.12. The groups O and T_d are isomorphic. Both are *enantiomorphic* (*i.e.* chiral), and completing either of them by adding in the inversion operator I results in the full cubic group, O_h, which has 48 elements.

While the groups T_d and O are isomorphic, the symmetry of their basis functions in general differs. Consider, for example, the function $\psi = xyz$. It is easy to see from table 2.5 that every element of T_d leaves ψ invariant. Within O, however, the classes $6\sigma_d$ and $6S_4$ are replaced by $6C_2$ and $6C_4$ when the inversion operation is removed. Each element of these classes then takes ψ to $-\psi$. Thus, within T_d, the function $\psi = xyz$ is indistinguishable from unity, and it transforms according to the trivial A_1 representation. Within O, however, ψ is distinguishable from 1 because ψ reverses sign under the operation of all group elements in classes $6C_2$ and $6C_4$.

In O, the triplets of basis functions $\{x, y, z\}$ and $\{yz, zx, xy\}$ belong to different representations (T_1 and T_2, respectively). In T_d, however, they must belong to the same representation, since one set of functions is obtained from the other by dividing into xyz: $x = (xyz)/(yz)$, *et. cyc.* But xyz transforms as the identity, so 'polar' and 'axial' vectors belong to the same representation of T_d.

[18] I am grateful to Filipp Rybakov for encouraging me to clarify my thinking regarding rotoreflections *versus* rotoinversions, and for correcting some errors in Tab. 2.5.

Finally, let's think about how O differs from O_h. Consider the function

$$\psi = xyz \cdot \left\{ x^4 \left(y^2 - z^2\right) + y^4 \left(z^2 - x^2\right) + z^4 \left(x^2 - y^2\right) \right\} . \tag{2.24}$$

One can check that this function is left invariant by every element of O. It therefore transforms according to the A_1 representation of O. But it reverses sign under parity, so within the full cubic group O_h, it transforms according to separate one-dimensional representation. Note that ψ transforms according to the A_2 representation of T_d.

2.2.8 The 32 crystallographic point groups

Tab. 2.7 lists all possible point group symmetries for three-dimensional crystals. The largest possible symmetry group within a given lattice system is the rightmost point group, corresponding to the symmetry of the underlying Bravais lattice. The point groups may be classified as being *centrosymmetric* (*i.e.* including the inversion operation I), *non-centrosymmetric*, or *enantiomorphic*. A centrosymmetric crystal has an inversion center. Enantiomorphic structures are non-centrosymmetric; they have only rotation axes and include no improper operations. They are intrinsically chiral and not superposable on their mirror image. In addition, a point group may be *polar*, meaning every symmetry operation leaves more than one point fixed (*i.e.* those points along the high symmetry *polar axis*). Thus, a group with more than one axis of rotation or with a mirror plane which does not contain the primary axis cannot be polar. A polar axis is only possible in non-centrosymmetric structures. Ferroelectricity and piezoelectricity can only occur in polar crystals.

LATTICE SYSTEM	POINT GROUPS						
cubic	T	T_d	T_h	O	O_h		
hexagonal	C_6^*	C_{3h}	C_{6h}	D_6	C_{6v}^*	D_{3h}	D_{6h}
trigonal	C_3^*	S_6	D_3	C_{3v}^*	D_{3d}		
tetragonal	C_4^*	S_4	C_{4h}	D_4	C_{4v}^*	D_{2d}	D_{4h}
orthorhombic	D_2	C_{2v}^*	D_{2h}				
monoclinic	C_2^*	C_s^*	C_{2h}				
triclinic	C_1^*	C_i					

Table 2.7: The 32 three-dimensional crystallographic point groups. Color scheme: centrosymmetric, non-centrosymmetric, enantiomorphic (*i.e.* chiral). Polar point groups are marked with an asterisk *.

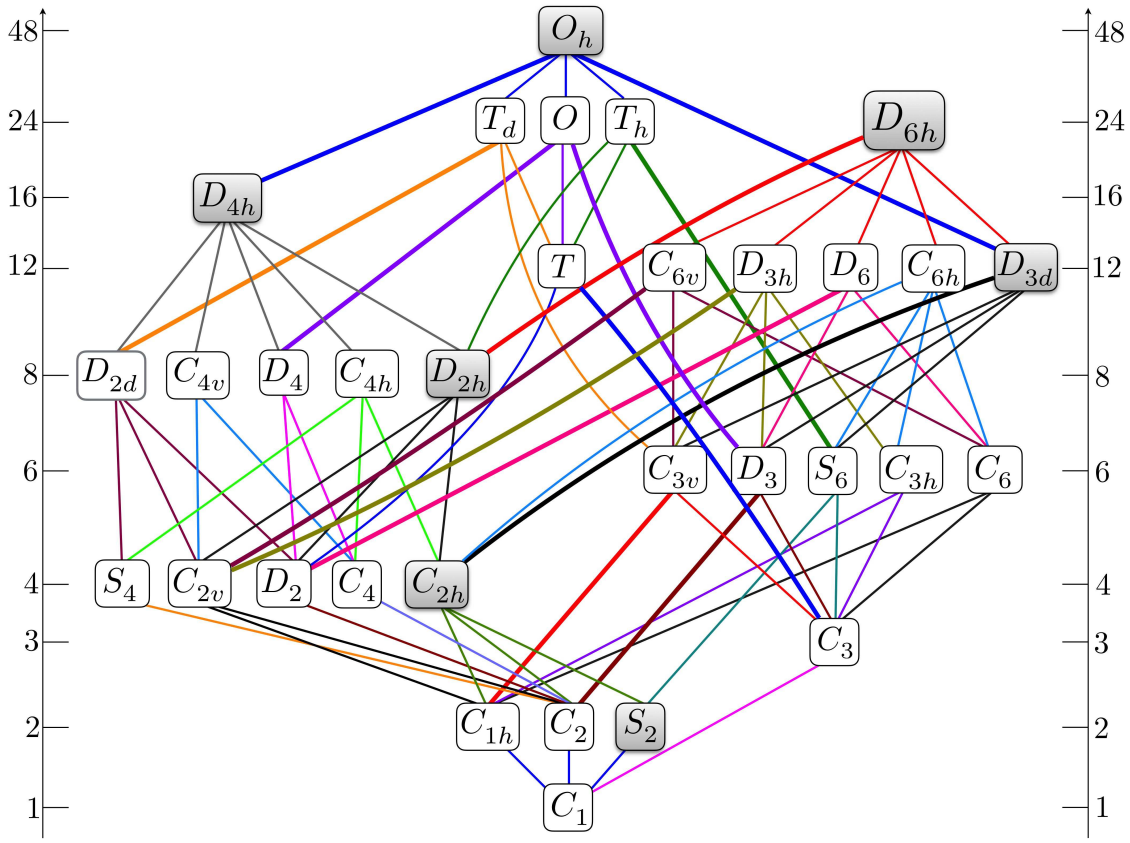


Figure 2.17: The 32 crystallographic point groups, their orders, and their subgroup structure. If the subgroup is not invariant (normal), the line is heavy. Gray boxes indicate holohedral groups, *i.e.* point groups of maximal symmetry within a given lattice system, corresponding to the symmetry of the underlying Bravais lattice itself. (See Tab. 6.1.6 of Lax, or Tab. 5 of Koster *et al.*)

2.2.9 Hermann-Mauguin (international) notation

The notation with which we have thus far identified point groups and their operations (C_{nv}, T_d, σ_h, *etc.*) is named for the German mathematician A. M. Schoenflies (1853-1928). A more informative system, originally due to German crystallographer C. Hermann and subsequently improved by the French minerologist C.-V. Mauguin, goes by the name *Hermann-Mauguin* (HM) or *international* notation. Since most physics publications today use the international notation, we pause to review it and to explain the method to its madness.

HM notation is defined for both point groups as well as their elements. For the individual symmetry operations, the HM symbols are as follows:

(i) n : rotation by $2\pi/n$ about a primary axis (Schoenflies C_n)

$$\begin{array}{ccccc} 2 = C_2 & 3 = C_3 & 4 = C_4 & 5 = C_5 & 6 = C_6 \\ & 3_2 = C_3^{-1} & 4_3 = C_4^{-1} & 5_4 = C_5^{-1} & 6_5 = C_6^{-1} \end{array} \tag{2.25}$$

(ii) m : reflection in a plane (σ)

- m_h : reflection in a plane perpendicular to the primary axis $\hat{\boldsymbol{n}}$ (σ_h)
- m_v : reflection in a plane containing the primary axis (σ_v)
- m_d : reflection in a plane containing the primary axis and bisecting the angle between two perpendicular 2-fold axes (σ_d)

(iii) $\overline{n}$: rotoinversion IC_n (note $\overline{1}$ is inversion, $\overline{2} = m_h$ is horizontal reflection)

$$\begin{array}{cccc} \overline{3} = S_6^{-1} & \overline{4} = S_4^{-1} & \overline{5} = S_{10}^{-1} & \overline{6} = S_3^{-1} \\ \overline{3}_2 = S_6 & \overline{4}_3 = S_4 & \overline{5}_4 = S_{10} & \overline{6}_5 = S_3 \end{array} \tag{2.26}$$

(iv) $\widetilde{n}$: rotoreflection $\sigma_h^{-1} C_n = S_n$

The number assignments associated with rotoinversion look strange at first. Pray tell, why do we have $\overline{3} = S_6^{-1}$ but $\overline{4} = S_4^{-1}$ and $\overline{6} = S_3^{-1}$? Well, since you asked so nicely, I will explain, but it will help if you consult Fig. 2.13. The issue here is that the Schoenflies groups $\mathcal{S}_n$ are generated by the *rotoreflection* operation $S_n \equiv \sigma_h^{-1} C_n$ while the HM symbol $\overline{n}$ denotes *rotoinversion* IC_n. The relation between the two is as follows. Let $C(\alpha)$ denote counterclockwise rotation through an angle α. Then $S(\alpha) = I\, C(\alpha - \pi)$. In other words, $S_n = I\, C_2^{-1} C_n$. According to this definition,

$$S_2 = I \quad , \quad S_3 = I\, C_6^{-1} \quad , \quad S_4 = I\, C_4^{-1} \quad , \quad S_6 = I\, C_3^{-1} \quad . \tag{2.27}$$

Note that $S_5 = I\, C_5^{-3}$, which produces a ten-fold pattern. In general, for n odd, S_n generates a $2n$-fold pattern.

Now let's talk about the HM symbols for the point groups themselves. The basic idea is to identify symmetry-inequivalent axes and reflection planes. For a single n-fold axis, the Schoenflies group is $\mathcal{C}_n$ and the HM symbol is n. If we add a vertical mirror σ_v to $\mathcal{C}_n$, forming $\mathcal{C}_{nv}$, the HM

$\mathbb{Z}_2$ clones	$\{E, C_2\}$	$\{E, I\}$	$\{E, \sigma_h\}$
Schoenflies	$\mathcal{C}_2$	$\mathcal{C}_i$	$\mathcal{C}_s$
HM	2	$\overline{1}$	m

Table 2.8: Two element point group notation.

Schoenflies	HM	2	3	4	5	6	$\mathcal{G}$ (HM)	order
$\mathcal{C}_n$	n	2	3	4	5	6	n	n
$\mathcal{S}_n$ (n odd)	$\overline{(2n)}$		$\overline{6}$		$\overline{10}$		$\overline{(2n)}$	$2n$
$\mathcal{S}_n$ ($n = 4k$)	$\overline{n}$			$\overline{4}$			$\overline{n}$	n
$\mathcal{S}_n$ ($n = 4k+2$)	$\overline{(n/2)}$	$\overline{1}$				$\overline{3}$	$\overline{n}$	n
$\mathcal{C}_{nv}$ (n even)	nmm	$2mm$		$4mm$		$6mm$	n, m_v	$2n$
$\mathcal{C}_{nv}$ (n odd)	nm		$3m$		$5m$		n, m_v	$2n$
$\mathcal{C}_{nh}$ (n even)	$\frac{n}{m}$	$\frac{2}{m}$		$\frac{4}{m}$		$\frac{6}{m}$	n, m_h	$2n$
$\mathcal{C}_{nh}$ (n odd)	$\overline{(2n)}$		$\overline{6}$		$\overline{10}$		n, m_h	$2n$
D_n (n even)	$n22$	222		422		622	$n, 2$	$2n$
D_n (n odd)	$n2$		32		52		$n, 2$	$2n$
D_{nd} (n even)	$\overline{(2n)}\,2m$	$\overline{4}2m$		$\overline{8}2m$		$\overline{12}\,2m$	$n, 2, m_d$	$4n$
D_{nd} (n odd)	$\overline{n}\,\frac{2}{m}$		$\overline{3}\,\frac{2}{m}$		$\overline{5}\,\frac{2}{m}$		$n, 2, m_d$	$4n$
D_{nh} (n even)	$\frac{n}{m}\frac{2}{m}\frac{2}{m}$	$\frac{2}{m}\frac{2}{m}\frac{2}{m}$		$\frac{4}{m}\frac{2}{m}\frac{2}{m}$		$\frac{6}{m}\frac{2}{m}\frac{2}{m}$	$n, 2, m_h$	$4n$
D_{nh} (n odd)	$\overline{(2n)}\,m2$		$\overline{6}m2$		$\overline{10}\,m2$		$n, 2, m_h$	$4n$

Table 2.9: Schoenflies and Hermann-Mauguin (international) notation for simple crystallographic point groups. The last columns list the generators $\mathcal{G}$ and the number of elements. Note $\mathcal{S}_n = \mathcal{C}_{nh}$ for n odd, and that $\overline{(2n)}\,2m = \overline{(2n)}\,m2$.

Schoenflies	T	T_h	T_d	O	O_h	I	I_h
HM	23	$\frac{2}{m}\,\overline{3}$	$\overline{4}3m$	432	$\frac{4}{m}\,\overline{3}\,\frac{2}{m}$	532	$\frac{2}{m}\,\overline{3}\,\overline{5}$
generators	$3, 2$	$3, 2, m_h$	$3, 2, m_d$	$4, 3, 2$	$4, 3, 2, m_h$	$5, 3, 2$	$5, 3, 2, m_h$
order	12	24	24	24	48	60	120

Table 2.10: Schoenflies and Hermann-Mauguin notation for multi-axis point groups. Indices for generators refer to distinct (though not necessarily orthogonal) axes.

symbol is nm if n is odd and nmm is n is even. The reason for the difference is that for n even, the alternating vertical reflections break into two classes, whereas for n odd there is only one class (check the character tables!). If we instead we had added a horizontal mirror σ_h to form $\mathcal{C}_{nh}$, the HM symbol would be $\frac{n}{m}$. However, when n is odd, $\mathcal{C}_{nh}$ is generated by the single rotoinversion $\overline{(2n)}$, and the convention is to use that symbol rather than the equivalent $\frac{n}{m}$ because the operation $\overline{(2n)}$ generates a pattern with more points than either n or m_h (though combined of course they generate the same group). For the dihedral groups D_n, the HM symbol is $n22$ if n is even and $n2$ if n is odd, for reasons similar to those for $\mathcal{C}_{nv}$. In general, for groups with a single primary axis, HM symbols can have up to three positions, which are assigned as follows:

- The first position indicates the rotational symmetry n of the primary axis, or $\overline{n}$ if the symmetry is rotoinversion. It can also be $\frac{n}{m}$ in the case of an n-fold axis plus a horizontal reflection plane.
- The second position indicates symmetry of a secondary axis or plane, and can be 2, m, or $\frac{2}{m}$.
- The third position indicates symmetry of a tertiary axis or plane, and can be 2, m, or $\frac{2}{m}$.

Thus, the HM symbol for D_{nd} is $\overline{n}\,\frac{2}{m}$ if n is odd but is $\overline{(2n)}m$ if n is even, while the HM symbol for D_{nh} is $\overline{n}m2$ if n is odd and $\frac{n}{m}\frac{2}{m}\frac{2}{m}$ if n is even. Notation for two element point groups is given in Tab. 2.8

Finally we come to the tetrahedral, octahedral, and icosahedral groups, all of which have more than one high order ($n > 2$) axis. For the tetrahedral group T, the HM symbol is 23 because the 2-fold axes are oriented parallel to the axes of the cube containing the tetrahedron, as shown in Figs. 2.9 and 2.12. The octahedral group O is written 432 in HM notation, because the fourfold axes are parallel to the cube axes, there are secondary threefold axes along the cube diagonals, and tertiary twofold axes running through the centers of the cube edges. The HM symbol for the icosahedral group I is 532. There are primary fivefold axes, through the vertices, secondary threefold axes through the face centers, and tertiary twofold axes through the edge centers (once again, see Fig. 2.9). Now add an improper element: inversion or a mirror plane. For the pyritohedral group T_h, we start with T and then add mirror planes perpendicular to the twofold axes, turning the threefold axes into inversion axes[19]. Consequently the HM symbol is $\frac{2}{m}\,\overline{3}$. For the achiral tetrahedral group T_d, we add mirrors perpendicular to the diagonal threefold axes, resulting in fourfold inversion axes and the symbol $\overline{4}3m$. When it comes to the cubic group O, we may add either a mirror or inversion. Since they are equivalent, consider the mirror, which bisects the fourfold axes, turning the threefold axes into inversion axes, and generating new mirrors perpendicular to the teriary twofold axes. The HM symbol is then $\frac{4}{m}\,\overline{3}\,\frac{2}{m}$. Finally, adding a mirror to the icosahedron turns I into I_h, with HM symbol $\overline{5}\,\overline{3}\,\frac{2}{m}$.

[19] The seams of a volleyball have pyritohedral symmetry.

No.	HM short	HM full	Schoenflies	No.	HM short	HM full	Schoenflies
1	1	1	C_1	17	$\overline{3}$	$\overline{3}$	$C_{3i}\ (S_6)$
2	$\overline{1}$	$\overline{1}$	$C_i\ (S_2)$	18	32	32	D_3
3	2	2	C_2	19	$3m$	$3m$	C_{3v}
4	m	m	$C_s\ (C_{1h})$	20	$\overline{3}m$	$\overline{3}\frac{2}{m}$	D_{3d}
5	$2/m$	$\frac{2}{m}$	C_{2h}	21	6	6	C_6
6	222	222	$D_2\ (V)$	22	$\overline{6}$	$\overline{6}$	C_{3h}
7	$mm2$	$mm2$	C_{2v}	23	$6/m$	$\frac{6}{m}$	C_{6h}
8	mmm	$\frac{2}{m}\frac{2}{m}\frac{2}{m}$	$D_{2h}\ (V_h)$	24	622	622	D_6
9	4	4	C_4	25	$6mm$	$6mm$	C_{6v}
10	$\overline{4}$	$\overline{4}$	S_4	26	$\overline{6}m2$	$\overline{6}m2$	D_{3h}
11	$4/m$	$\frac{4}{m}$	C_{4h}	27	$6/mmm$	$\frac{6}{m}\frac{2}{m}\frac{2}{m}$	D_{6h}
12	422	422	D_4	28	23	23	T
13	$4mm$	$4mm$	C_{4v}	29	$m\overline{3}$	$\frac{2}{m}\overline{3}$	T_h
14	$\overline{4}2m$	$\overline{4}2m$	$D_{2d}\ (V_d)$	30	432	432	O
15	$4/mmm$	$\frac{4}{m}\frac{2}{m}\frac{2}{m}$	D_{4h}	31	$\overline{4}3m$	$\overline{4}3m$	T_d
16	3	3	C_3	32	$m\overline{3}m$	$\frac{4}{m}\overline{3}\frac{2}{m}$	O_h

Table 2.11: HM and Schoenflies notation for the 32 crystallographic point groups.

2.2.10 Double groups

The group operations act on electron wavefunctions, which are spinor functions of the spatial coordinates $\boldsymbol{r} = (x, y, z)$:

$$\vec{\psi}(\boldsymbol{r}) = \begin{pmatrix} \psi_\uparrow(\boldsymbol{r}) \\ \psi_\downarrow(\boldsymbol{r}) \end{pmatrix} . \tag{2.28}$$

Rotations by an angle θ about an axis $\hat{\boldsymbol{n}}$ are represented by the unitary operator $U(\theta;\hat{\boldsymbol{n}}) = e^{-i\theta\hat{\boldsymbol{n}}\cdot\boldsymbol{J}/\hbar}$, where $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$ is the sum of orbital ($\boldsymbol{L}$) and intrinsic spin ($\boldsymbol{S}$) angular momenta. For crystallographic point groups, $\theta = 2\pi/n$ where $n = 1$, 2, 3, 4, or 6.

When spin is neglected, we have the point groups we have studied. With spin, we must deal with the fact that $\mathsf{SU}(2)$ gives us a *projective representation* of $\mathsf{SO}(3)$. Recall that $\hat{D}(G)$ is a projec-

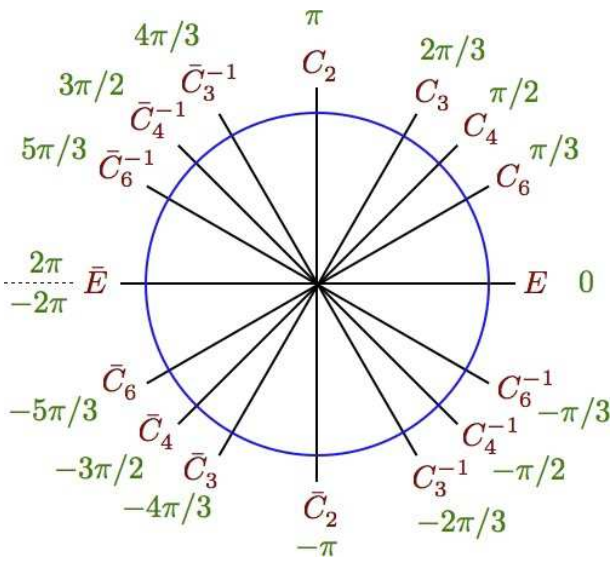


Figure 2.18: Schematic diagram of (common axis) double group rotation generators. Based on Fig. 1 of Koster *et al.* (1963). Note that $C_1 = \overline{E}$, *i.e.* a rotation by 2π.

tive representation of G if

$$\hat{D}(g)\,\hat{D}(h) = \omega(g,h)\,\hat{D}(gh) \tag{2.29}$$

where associativity imposes the following condition on the cocycle $\omega(g,h)$:

$$\frac{\omega(g,h)}{\omega(h,k)} = \frac{\omega(g,hk)}{\omega(gh,k)} \quad . \tag{2.30}$$

In our case, $G = \mathsf{SO}(3)$ and $\hat{D}\big(R(\xi,\hat{\boldsymbol{n}})\big) = \exp(-i\xi\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{J}})$ where $\hat{\boldsymbol{J}} = \hat{\boldsymbol{L}} + \hat{\boldsymbol{S}}$ and $S = \frac{1}{2}$. For example, any C_2 operation has $\xi = \pi$, hence $(C_2)^2 = C_1 = \exp(-2\pi i\hat{\boldsymbol{n}}\cdot\hat{\boldsymbol{S}}) = -1$, which is to say spinor inversion, *i.e.* $\begin{pmatrix} u \\ v \end{pmatrix} \to -\begin{pmatrix} u \\ v \end{pmatrix}$. *For any point group $\mathcal{P}$, the multiplication table for the projective representation $\hat{D}(\mathcal{P})$ looks exactly like that for $\mathcal{P}$, except some entries get multiplied by -1. I.e.* all the cocycles $\omega(g,h)$ are ± 1. We can *lift* this projective representation to an enlarged point group, called the *double group*, $\mathcal{P}'$, by introducing a generator $\overline{E}$, representing spinor inversion, with $\overline{E}^2 = E$. To each element $g \in \mathcal{P}$, there corresponds a counterpart $\bar{g} \equiv g\overline{E}$. Thus, $N_{\mathcal{P}'} = 2N_{\mathcal{P}}$. Note that $\overline{E}$ leaves $\boldsymbol{r}$ unchanged, and that the bar of g^{-1} is the inverse of $\bar{g}$. A schematic illustration of proper rotations within a double group is shown in Fig. 2.18. *Do not confuse the barring operation in double groups with the HM symbol for rotoinversion!*

Remarks about double group multiplication

Some noteworthy aspects regarding multiplication of double group elements:

- ◇ The element $\overline{E}$ is given by $\overline{E} = C_1 = C(\pm 2\pi)$. Note $C(4\pi) = E$.
- ◇ For any group element g, whether barred or unbarred, $g g^{-1} = E$.
- ◇ For the inversion operator I, $I^2 = \bar{I}^2 = E$ and $I\bar{I} = \bar{I}I = \overline{E}$.
- ◇ Any reflection σ obeys $\sigma^2 = \overline{E}$. This is because we can always write $\sigma = IC_2$ where C_2 is a twofold rotation about an axis normal to the reflection plane, whence $\sigma^2 = I^2\, C_2^2 = C_1 = \overline{E}$.
- ◇ For $n > 2$, we define $C_n \equiv \hat{R}(2\pi/n)$ to be a counterclockwise rotation by $2\pi/n$ and $C_n^{-1} \equiv \hat{R}(-2\pi/n)$ to be the inverse operation, *i.e.* clockwise rotation by $2\pi/n$. Then $C_2^2 = C_1 = \overline{E}$, hence $C_2^{-1} = \overline{C}_2$. More fully, according to Fig. 2.18, we have

$$\begin{aligned}
&C_2\, C_3^{-1} = C_6 \qquad && C_2\, C_4^{-1} = C_4 \qquad && C_2\, C_6^{-1} = C_3 \\
&C_2\, C_6 = \overline{C}_3^{-1} \qquad && C_2\, C_4 = \overline{C}_4^{-1} \qquad && C_2\, C_3 = \overline{C}_6^{-1} \\
&C_2\, \overline{C}_3^{-1} = \overline{C}_6 \qquad && C_2\, \overline{C}_4^{-1} = \overline{C}_4 \qquad && C_2\, \overline{C}_6^{-1} = \overline{C}_3 \\
&C_2\, \overline{C}_6 = C_3^{-1} \qquad && C_2\, \overline{C}_4 = C_4^{-1} \qquad && C_2\, \overline{C}_3 = C_6^{-1} \quad ,
\end{aligned} \tag{2.31}$$

  where all rotations are about the same axis.
- ◇ To compute the product of σ_h with a rotation, recall the definition of the rotoreflection operation $S_n \equiv \sigma_h^{-1}\, C_n = I\, C_2^{-1}\, C_n$, which entails $S_n^{-1} = \sigma_h\, C_n^{-1} = I\, C_2\, C_n^{-1}$. One then has

$$\sigma_h\, C_n = \overline{S}_n \quad , \quad \sigma_h\, \overline{C}_n^{-1} = \overline{S}_n^{-1} \quad , \quad \sigma_h\, \overline{C}_n = S_n \quad , \quad \sigma_h\, C_n^{-1} = S_n^{-1} \quad . \tag{2.32}$$

- ◇ We may then apply σ_h to Eqns. 2.31 and 2.32 to obtain results such as

$$C_2\, S_3^{-1} = \overline{S}_6 \quad , \quad C_2\, S_6 = S_3^{-1} \quad , \quad \sigma_h\, S_n = C_n \quad , \quad \sigma_h\, S_n^{-1} = \overline{C}_n^{-1} \quad . \tag{2.33}$$

- ◇ What about σ_v ? If $\{\sigma_v\, ,\, \sigma_{v'}\, ,\, \sigma_{v''}\}$ denote vertical reflection planes oriented at angles 0, $2\pi/3$, and $4\pi/3$, respectively, then we should have either $C_3\, \sigma_v = \sigma_{v'}$ or $C_3\, \sigma_v = \overline{\sigma}_{v'}$. Which is it? If we apply C_3 twice, for either initial case we obtain $C_3^2\, \sigma_v = \sigma_{v''}$. Applying C_3 yet again yields $C_3\, \sigma_{v''} = \overline{\sigma}_v$. Thus we have

$$C_3\, \sigma_v = \overline{\sigma}_{v'} \quad , \quad C_3\, \sigma_{v'} = \overline{\sigma}_{v''} \quad , \quad C_3\, \sigma_{v''} = \overline{\sigma}_v \quad . \tag{2.34}$$

  Note then that $\sigma_{v'}\, \sigma_v = C_3$ and $\sigma_v\, \sigma_{v'} = C_3^{-1}$, *et. cyc.*

To summarize, let $C(\alpha)$ denote counterclockwise rotation through an angle α, and let $C_n = C(\alpha_n)$ *etc.* with $\alpha_n = 2\pi/n$. Then

$$\overline{C}(\alpha) = C(\alpha - 2\pi) \quad , \quad S(\alpha) = I\, C(\alpha - \pi) \quad , \quad \overline{S}(\alpha) = I\, C(\alpha + \pi) \tag{2.35}$$

and

$$\sigma = I\, C(\pi) \quad , \quad \overline{\sigma} = I\, C(-\pi) \quad . \tag{2.36}$$

T_d'	E	$\overline{E}$	$8C_3$	$8\overline{C}_3$	$\begin{matrix}3C_2\\3\overline{C}_2\end{matrix}$	$\begin{matrix}6\sigma_d\\6\overline{\sigma}_d\end{matrix}$	$6S_4$	$6\overline{S}_4$
$\Gamma_1\ (A_1)$	1	1	1	1	1	1	1	1
$\Gamma_2\ (A_2)$	1	1	1	1	1	−1	−1	−1
$\Gamma_3\ (E)$	2	2	−1	−1	2	0	0	0
$\Gamma_4\ (T_1)$	3	3	0	0	−1	−1	1	1
$\Gamma_5\ (T_2)$	3	3	0	0	−1	1	−1	−1
Γ_6	2	−2	1	−1	0	0	$\sqrt{2}$	$-\sqrt{2}$
Γ_7	2	−2	1	−1	0	0	$-\sqrt{2}$	$\sqrt{2}$
Γ_8	4	−4	−1	1	0	0	0	0

Table 2.12: Character table for the double group of T_d.

Character tables for double groups

One might at first suspect that any conjugacy class $\mathcal{C}$ of the point group $\mathcal{P}$ spawns two classes within the double group $\mathcal{P}'$, *i.e.* $\mathcal{C}$ and $\overline{\mathcal{C}} \equiv \overline{E}\,\mathcal{C}$. This is always true provided the elements of C don't square to the identity. But for twofold axes C_2 and reflections σ, a theorem due to Opechowski (1940) guarantees:

- For proper twofold operations, C_2 and $\bar{C}_2$ adjoin to the same class if either
  - there exists a second twofold axis perpendicular to the initial axis, or
  - there exists a reflection plane containing the initial axis.
- For improper twofold operations, σ and $\bar{\sigma}$ adjoin to the same class if either
  - there exists a second reflection plane perpendicular to the initial one, or
  - there exists a twofold axis lying within the initial plane.

In these cases, the resulting total number of classes in $\mathcal{P}'$ is less than twice that for $\mathcal{P}$. As an example, consider the tetrahedral group T_d. There are three twofold axis: $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$. All are bilateral because a rotation by π about $\hat{\boldsymbol{x}}$ reverses the direction of both $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{z}}$, *etc.* Accordingly, in the character table Tab. 2.12 for the double group of T_d, the classes C_2 and $\bar{C}_2$ are adjoined, as are σ_d and $\bar{\sigma}_d$.

With the exception of those twofold operations satisfying the conditions in Opechowski's theorem, the classes $\mathcal{C}$ and $\overline{\mathcal{C}}$ are distinct in the double group. Any IRREP of $\mathcal{P}$ will be an IRREP of $\mathcal{P}'$ with $\chi(\overline{\mathcal{C}}) = \chi(\mathcal{C})$. But since the number of elements is doubled in $\mathcal{P}'$, there must be new

IRREPs specific to the double group. For these additional IRREPs, one has $\chi(\overline{\mathcal{C}}) = -\chi(\mathcal{C})$, hence if $\mathcal{C}$ and $\overline{\mathcal{C}}$ adjoin to $\mathcal{C} \cup \overline{\mathcal{C}}$ by Opechowski, one must have $\chi(\mathcal{C} \cup \overline{\mathcal{C}}) = 0$. Checking Tab. 2.12, we see that in the extra IRREPs $\Gamma_{6,7,8}$, $\chi(\overline{\mathcal{C}}) = -\chi(\mathcal{C})$ except in the case of the adjoined classes, for which $\chi(\mathcal{C} \cup \overline{\mathcal{C}}) = 0$.

We can understand that twofold rotations and reflections are special in this regard from the result we obtained for $\mathsf{SU}(2)$ characters,

$$\chi^{(j)}(\xi) = \frac{\sin\,(j+\frac{1}{2})\xi}{\sin\,\frac{1}{2}\xi} \tag{2.37}$$

for rotation by an angle ξ about any axis. Thus $\chi^{(j)}(\alpha + 2\pi) = (-1)^{2j}\chi^{(j)}(\alpha)$. For $j = \frac{1}{2}$, or indeed for any half odd integer j, we have $\chi(\pi) = \chi(3\pi) = 0$. Thus C_2 and $\overline{C}_2$ have the same character. A similar result holds for reflections, because $\sigma = IC_2$ and $\overline{\sigma} = I\overline{C}_2$. Therefore the classes C_2 and $\overline{C}_2$ are not distinguished by character, nor are σ and $\overline{\sigma}$. This is true in any IRREP in which $\chi(\overline{E}) = -\chi(E)$.

2.2.11 The three amigos : D_4, C_{4v}, D_{2d}

Let's try to apply some of what we've just learned to the groups D_4, C_{4v}, and D_{2d}. All these eight-element groups are isomorphic to each other. The character table for all three is given in Tab. 2.13. Although they are all isomorphic, they include different sets of symmetry operations, and therefore they will have different basis representations.

Let's now discuss all the classes of these three groups. Recall that

$$R(\xi, \hat{\boldsymbol{n}})_{ab} = n^a n^b + \left(\delta^{ab} - n^a n^b\right)\cos\xi - \epsilon_{abc}\, n^c \sin\xi \quad . \tag{2.38}$$

- C_2 : This class is present in all three groups. It consists of a single element which is rotation by π about the $\hat{\boldsymbol{z}}$ axis, and represented by the 3×3 matrix

$$R_z^{\pi} \equiv R(\pi, \hat{\boldsymbol{z}}) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.39}$$

- $2C_4$: Present in D_4 and C_{4v}. Contains the elements

$$R_z^{\pi/2} \equiv R(\tfrac{\pi}{2}, \hat{\boldsymbol{z}}) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad R_z^{-\pi/2} \equiv R(-\tfrac{\pi}{2}, \hat{\boldsymbol{z}}) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.40}$$

  These elements are inverses of each other.

D_4 C_{4v} D_{2d}	E E E	$2C_4$ $2C_4$ $2S_4$	C_2 C_2 C_2	$2C_2'$ $2\sigma_v$ $2C_2'$	$2C_2''$ $2\sigma_d$ $2\sigma_d$	D_4 basis	C_{4v} basis	D_{2d} basis
A_1	1	1	1	1	1	x^2+y^2	x^2+y^2 or z	x^2+y^2
A_2	1	1	1	-1	-1	L_z or z	L_z	L_z
B_1	1	-1	1	1	-1	x^2-y^2	x^2-y^2	x^2-y^2
B_2	1	-1	1	-1	1	xy	xy	xy or z
E	2	0	-2	0	0	$\{x\,,\,y\}$	$\{x\,,\,y\}$	$\{x\,,\,y\}$

Table 2.13: Character table for the point groups D_4, C_{4v}, and D_{2d}.

- $2S_4$: Present only in D_{2d} . These are rotoreflections, *i.e.* $2C_4$ followed by $z \to -z$:

$$S_z^{\pi/2} \equiv S(\tfrac{\pi}{2}, \hat{\boldsymbol{z}}) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad , \quad S_z^{-\pi/2} \equiv S(-\tfrac{\pi}{2}, \hat{\boldsymbol{z}}) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad . \tag{2.41}$$

  These two are also inverses within $\mathsf{O}(3)$. In general we have $S(\alpha) = IC(\alpha - \pi)$, in which case $S_z^{\pi/2} = IR_z^{-\pi}R_z^{\pi/2}$ and $S_z^{-\pi/2} = IR_z^{-\pi}R_z^{-\pi/2}$. Why do we distinguish R_z^{π} and $R_z^{-\pi}$ when they are represented by the same matrix? This will be important when we construct the corresponding matrix representation for the double groups[20].

- $2C_2'$: Present in D_4 and D_{2d} , this class consists of twofold rotations about $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$:

$$R_x^{\pi} \equiv R(\pi, \hat{\boldsymbol{x}}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad , \quad R_y^{\pi} \equiv R(\pi, \hat{\boldsymbol{y}}) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad . \tag{2.42}$$

- $2\sigma_v$: This occurs only in C_{4v} and corresponds to reflections $x \to -x$ and $y \to -y$:

$$\Sigma_x \equiv IR_x^{\pi} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad \Sigma_y \equiv IR_y^{\pi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.43}$$

[20] See the explanation of Eqn. 2.32.

- $2C_2''$: Occurring only in D_4, these operations are twofold rotations about the diagonal axes $y = x$ and $y = -x$:

$$R^{\pi}_{xy} \equiv R\big(\pi, \tfrac{\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}}}{\sqrt{2}}\big) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad , \quad R^{\pi}_{x\overline{y}} \equiv R\big(\pi, \tfrac{\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}}}{\sqrt{2}}\big) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad . \tag{2.44}$$

- $2\sigma_d$: Occurring in C_{4v} and D_{2d}, this class of reflections is equivalent to IC_2'', hence

$$\Sigma_{xy} \equiv IR^{\pi}_{xy} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad , \quad \Sigma_{x\overline{y}} \equiv IR^{\pi}_{x\overline{y}} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{2.45}$$

I apologize for the loose notation where we are using the same symbols to refer to group elements as well as their 3×3 matrix representations. Notice that all the matrices representing elements of C_{4v} have a block-diagonal structure with an upper left 2×2 block and a lower right 1×1 block, where the latter is always 1. This is because we never need to speak of the z-direction when we talk about C_{4v} as all its operations involve x and y alone.

Now let's talk about the basis functions. The projectors onto the various representations are given by

$$\Pi^{\Gamma} = \frac{d_{\Gamma}}{N_G} \sum_{g \in G} \chi^{\Gamma^*}(g)\, D(g) \quad , \tag{2.46}$$

where $N_G = 8$ for the three amigos. It should be clear how the basis functions in Tab. 2.13 are eigenfunctions of these projectors, but let's note the following to obviate any confusion. First of all, what do we mean by L_z as a basis function of the A_2 IRREP in the case of C_{4v} and D_{2d}? We mean the angular momentum operator, $L_z = xp_y - yp_x$. We know that $L_\alpha = \varepsilon_{\alpha\beta\gamma} r^\beta p^\gamma$ transforms as a vector under proper rotations, however it is known as an *axial vector* because it transforms differently under improper rotations. That is, under the operation σ_h (which, *nota bene* is present in none of our three groups), z is odd but L_z is even. Similarly, under either of the σ_v operations, z is even but L_z is odd. For D_4, the basis function $f(z) = z$ corresponds to the A_2 IRREP because it is even under E, $2C_4$, and C_2 and odd under $2C_2'$ and $2C_2''$. But in C_{4v}, whose operations all leave z invariant, $f(z) = z$ transforms as the A_1 IRREP. And for D_{2d}, where $2S_4$ and $2C_2'$ reverse z but $2\sigma_d$ does not, $f(z) = z$ transforms as the B_2 IRREP! Note that other valid choices of basis functions are possible. For example, rather than the pair $\{x\,,\,y\}$, we could have chosen $\{L_x\,,\,L_y\}$ as basis functions for the E IRREP.

Double group matrices and projectors

Now let's tackle the corresponding double groups. We will need the 2×2 matrices representing the various point group operations. Recall for a rotation by ξ about $\hat{\boldsymbol{n}}$,

$$\exp(-i\xi\hat{\boldsymbol{n}}\cdot\boldsymbol{\sigma}/2) = \cos(\tfrac{1}{2}\xi) - i\sin(\tfrac{1}{2}\xi)\,\hat{\boldsymbol{n}}\cdot\boldsymbol{\sigma} \quad . \tag{2.47}$$

We'll write the elements of $D^{(1/2)}(G)$ as $U(g)$. We then have

$$U(R_z^\pi) = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \quad , \quad U(R_z^{\pi/2}) = \begin{pmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \quad , \quad U(R_z^{-\pi/2}) = \begin{pmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix} \quad . \tag{2.48}$$

For the rotoreflections,

$$U(S_z^{\pi/2}) = \begin{pmatrix} e^{i\pi/4} & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix} \quad , \quad U(S_z^{\pi/2}) = \begin{pmatrix} -e^{-i\pi/4} & 0 \\ 0 & -e^{i\pi/4} \end{pmatrix} \quad . \tag{2.49}$$

Note that $U(S_z^{\pm\pi/2}) = I\,U(R_z^{-\pi})\,U(R_z^{\pm\pi/2})$, where $R_z^{-\pi} = -R_z^\pi$ and that I acts as the identity matrix on spinors. Note that $U(S_z^{\pi/2}) = U(R_z^{-\pi/2})$. Next, we need

$$U(\Sigma_x) = U(R_x^\pi) = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} \quad , \quad U(\Sigma_y) = U(R_y^\pi) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad . \tag{2.50}$$

Since the only difference between the twofold rotations and the corresponding reflections in the planes perpendicular to their axes is the inversion I, their representations in $D^{1/2}(G)$ are identical. The remaining matrices are

$$U(\Sigma_{xy}) = U(R_{xy}^\pi) = \begin{pmatrix} 0 & -e^{i\pi/4} \\ e^{-i\pi/4} & 0 \end{pmatrix} \quad , \quad U(\Sigma_{x\overline{y}}) = U(R_{x\overline{y}}^\pi) = \begin{pmatrix} 0 & e^{-i\pi/4} \\ -e^{i\pi/4} & 0 \end{pmatrix} \quad . \tag{2.51}$$

Note that their product is $U(\Sigma_{x\overline{y}})\,U(\Sigma_{xy}) = U(R_z^\pi)$. Note also that $\det U(g) = 1$ since each $U(g) \in \mathsf{SU}(2)$.

Appealing to the character table in Tab. 2.14, we can now construct the double group projectors. We write the projectors as

$$\Pi^\Gamma = \frac{d_\Gamma}{N_G} \sum_{g \in G} \chi^{\Gamma^*}(g)\, D(g) \otimes U(g) \quad . \tag{2.52}$$

where G is any of D_4', C_{4v}', and D_{2d}', and $N_G = 16$, since each of the double groups of the three amigos has 16 elements. For the IRREPs $\{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5\}$ we may use the basis functions $\psi_\mu^\Gamma(\boldsymbol{r})$ from the proper point groups. *I.e.* we can simply ignore all the U-matrices and pretend there is no spin component. More correctly, we can consider the spin component of each basis function to be a *singlet*,

$$|\,\mathrm{S}\,\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle\Big) \quad . \tag{2.53}$$

One can check that $U(g)\,|\,\mathrm{S}\,\rangle = |\,\mathrm{S}\,\rangle$ for all g, which follows from $\det U(g) = 1$. For Γ_6 and Γ_7, though, the projectors annihilate any basis function of the form $f(\boldsymbol{r})\,|\,\mathrm{S}\,\rangle$. However, a basis

D_4'	E	$\overline{E}$	$2C_4$	$2\overline{C}_4$	$C_2 \cup \overline{C}_2$	$2C_2' \cup 2\overline{C}_2'$	$2C_2'' \cup 2\overline{C}_2''$	
C_{4v}'	E	$\overline{E}$	$2C_4$	$2\overline{C}_4$	$C_2 \cup \overline{C}_2$	$2\sigma_v \cup 2\overline{\sigma}_v$	$2\sigma_d \cup 2\overline{\sigma}_d$	
D_{2d}'	E	$\overline{E}$	$2S_4$	$2\overline{S}_4$	$C_2 \cup \overline{C}_2$	$2C_2' \cup 2\overline{C}_2'$	$2\sigma_d \cup 2\overline{\sigma}_d$	basis
Γ_1	1	1	1	1	1	1	1	x^2+y^2
Γ_2	1	1	1	1	1	-1	-1	L_z
Γ_3	1	1	-1	-1	1	1	-1	x^2-y^2
Γ_4	1	1	-1	-1	1	-1	1	xy
Γ_5	2	2	0	0	-2	0	0	$\{x\,,\,y\}$ or $\{L_x\,,\,L_y\}$
Γ_6	2	-2	$\sqrt{2}$	$-\sqrt{2}$	0	0	0	$\{\|\uparrow\rangle\,,\,\|\downarrow\rangle\}$
Γ_7	2	-2	-$\sqrt{2}$	$\sqrt{2}$	0	0	0	$\Gamma_3 \times \Gamma_6$ or $\Gamma_4 \times \Gamma_6$

Table 2.14: Character table for the double groups of D_4, C_{4v}, and D_{2d}.

function of the form $|\uparrow\rangle$ or $|\downarrow\rangle$ (*i.e.* with no spatial dependence) does the trick. In spinor notation, we have

$$\begin{aligned} \frac{2}{16}\left[\chi^{\Gamma_6}(E)\,U(E)+\chi^{\Gamma_6}(\overline{E})\,U(\overline{E})\right]\begin{pmatrix}u\\v\end{pmatrix} &= \frac{1}{2}\begin{pmatrix}u\\v\end{pmatrix} \\ \frac{2}{16}\left[\chi^{\Gamma_6}(2C_4)+\chi^{\Gamma_6}(2\overline{C}_4)\,\overline{E}\right]\left[U(R_z^{\pi/2})+U(R_z^{-\pi/2})\right]\begin{pmatrix}u\\v\end{pmatrix} &= \frac{1}{2}\begin{pmatrix}u\\v\end{pmatrix} \quad . \end{aligned} \tag{2.54}$$

Thus, $\begin{pmatrix}u\\v\end{pmatrix}$ is an eigenfunction of the projector Π^{Γ_6}. In order to keep this spinor from being annihilated by Π^{Γ_7}, we need to multiply it by a scalar function $\psi(\boldsymbol{r})$ which reverses the sign from the characters of the classes $2C_4$ and $2\overline{C}_4$. According to Tab. 2.13, the basis function from either the B_1 or the B_2 IRREPs will work. This explains the basis functions in Tab. 2.14[21]. Other valid choices of basis functions are of course possible.

Do we always need the double group?

Although electrons carry spin $S=\frac{1}{2}$, we usually don't need to invoke the double group formalism if the spin-orbit coupling is sufficiently weak. That is, we may use $\boldsymbol{L}$ rather than $\boldsymbol{J}$ as the generator of rotations, since $\left[\hat{H}, L^\alpha\right]=0$. Each electronic energy level is of course doubly

[21] The spin component of the basis functions for the Γ_1 through Γ_5 IRREPs should be considered to be the singlet $|\,\mathrm{S}\,\rangle$.

degenerate due to the spin, which just "comes along for the ride". In the presence of significant spin-orbit coupling, $\left[\hat{H}, L^\alpha\right] \neq 0$ but $\left[\hat{H}, J^\alpha\right] = 0$. Thus we must use the total angular momentum $\boldsymbol{J}$ as the generator of rotations, which entails the double point group symmetries.

2.3 Space Groups

The full group of symmetry operations of an n-dimensional crystal is called its *space group*, $\mathcal{S}$. Any crystallographic space group is a subgroup of the Euclidean group: $\mathcal{S} \subset \mathsf{E}(n)$. Space groups are infinite discrete groups. Two-dimensional space groups are called *wallpaper groups*. An accounting of the total number of lattices, point groups, and space groups for two and three dimensional crystals is provided in Tab. 2.1.

2.3.1 Space group elements and their properties

Each element $\{\, g \,|\, \boldsymbol{t} \,\} \in \mathcal{S}$ represents a compounded operation of rotation by a rotation g (either proper or improper) and a translation $\boldsymbol{t}$. When $g = E$, the space group operations are pure translations, and are all of the form $\{\, E \,|\, \boldsymbol{R} \,\}$, where $\boldsymbol{R} \in \mathcal{L}$ is a vector in the underlying Bravais lattice. As discussed in §2.1.7, the operations $\{\, g \,|\, \boldsymbol{t} \,\}$ form a group, with

$$\begin{aligned} \{\, g \,|\, \boldsymbol{t} \,\}\{\, g' \,|\, \boldsymbol{t}' \,\} &= \{\, gg' \,|\, g\boldsymbol{t}' + \boldsymbol{t} \,\} \\ \{\, g \,|\, \boldsymbol{t} \,\}^{-1} &= \{\, g^{-1} \,|\, -g^{-1}\boldsymbol{t} \,\} \quad . \end{aligned} \tag{2.55}$$

We see that the rotations g must themselves form a group, which is the *point group* $\mathcal{P}$ of the crystal. Pure translations $\{\, E \,|\, \boldsymbol{R} \,\}$ by a direct lattice vector are part of the space group, and indeed form a normal subgroup thereof: $\{\, g \,|\, \boldsymbol{t} \,\}^{-1}\{\, E \,|\, \boldsymbol{R} \,\}\{\, g \,|\, \boldsymbol{t} \,\} = \{\, E \,|\, g^{-1}\boldsymbol{R} \,\}$. Thus, $g^{-1}\boldsymbol{R} \in \mathcal{L}$ for any $g \in \mathcal{P}$, which means, as noted above in §2.1.7, that the point group $\mathcal{P}$ of any crystal is a subgroup of the point group $\mathcal{P}_{\mathcal{L}}$ of its underlying Bravais lattice (*i.e.* the holohedry).

Fom Eqn. 2.55, we have the group conjugation property

$$\{\, h \,|\, \boldsymbol{s} \,\}^{-1}\{\, g \,|\, \boldsymbol{t} \,\}\{\, h \,|\, \boldsymbol{s} \,\} = \{\, h^{-1}gh \,|\, h^{-1}g\boldsymbol{s} - h^{-1}\boldsymbol{s} + h^{-1}\boldsymbol{t} \,\} \equiv \{\, g' \,|\, \boldsymbol{t}' \,\} \quad , \tag{2.56}$$

for which the rotation is $g' = h^{-1}gh$ and the translation is $\boldsymbol{t}' = h^{-1}g\boldsymbol{s} - h^{-1}\boldsymbol{s} + h^{-1}\boldsymbol{t}$. When $h = E$, we have $g' = g$ and

$$\boldsymbol{t} - \boldsymbol{t}' = (E - g)\,\boldsymbol{s} \quad . \tag{2.57}$$

Suppose we further demand $\boldsymbol{t}' = 0$, *i.e.* that the conjugated operation is equivalent to a pure rotation, with no translation, about a different choice of origin. We see that this is possible if we choose $\boldsymbol{s}$ such that $\boldsymbol{t} = (E - g)\,\boldsymbol{s}$.

Now it was noted in §2.2.1 that when the dimension n of space is odd, $g \in \mathsf{O}(n)$ always preserves some axis, meaning it has an eigenvalue $\lambda = 1$. The other two eigenvalues may be

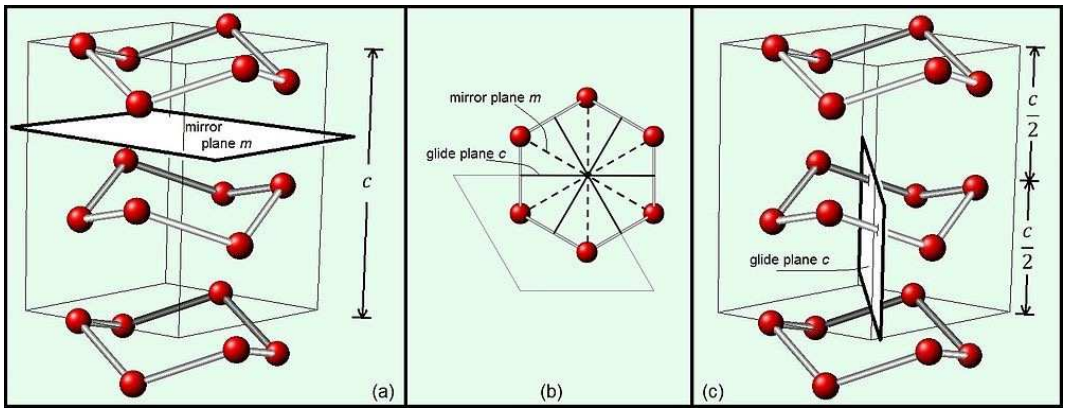


Figure 2.19: Structure of hexagonal H_2O ice, with red spheres showing location of oxygen atoms. The space group is $P6_3/mmc$. The 6_3 symbol indicates a sixfold screw axis. The first two m symbols indicate mirror planes perpendicular and parallel to the c-axis. The c symbol indicates a glide plane where the translation is along the c-axis. (Image credit: *Wikipedia*).

written as $e^{\pm i\alpha}$ where $\alpha = 2\pi/n$ with $n = 2$, 3, 4, or 6. (The case $n = 1$ corresponds to the identity E.) A mirror reflection, which is an improper operation, has an inversion axis corresponding to an eigenvalue $\lambda = -1$, with all remaining eigenvalues $\lambda = +1$. Proper rotations therefore have an *invariant axis*, while mirror reflections have an *invariant plane*. Thus we can write

$$\begin{aligned} \text{proper rotation}: \quad & r = |\,\hat{\boldsymbol{e}}_1\,\rangle\langle\,\hat{\boldsymbol{e}}_1\,| + e^{i\alpha}\,|\,\hat{\boldsymbol{e}}_2\,\rangle\langle\,\hat{\boldsymbol{e}}_2\,| + e^{-i\alpha}\,|\,\hat{\boldsymbol{e}}_3\,\rangle\langle\,\hat{\boldsymbol{e}}_3\,| \\ \text{mirror reflection}: \quad & m = -|\,\hat{\boldsymbol{e}}_1\,\rangle\langle\,\hat{\boldsymbol{e}}_1\,| + |\,\hat{\boldsymbol{e}}_2\,\rangle\langle\,\hat{\boldsymbol{e}}_2\,| + |\,\hat{\boldsymbol{e}}_3\,\rangle\langle\,\hat{\boldsymbol{e}}_3\,| \quad . \end{aligned} \tag{2.58}$$

We now see that if $g = r$ is a proper rotation, $\boldsymbol{t} = (E - r)\,\boldsymbol{s}$ cannot be solved for $\boldsymbol{s}$ if $\boldsymbol{t}$ has any component along the invariant axis $\hat{\boldsymbol{e}}_1$. Similarly, if $g = m$ is a mirror, $\boldsymbol{t} = (E - m)\,\boldsymbol{s}$ cannot be solved for $\boldsymbol{s}$ if $\boldsymbol{t}$ has any component in the invariant plane spanned by $\{\hat{\boldsymbol{e}}_2, \hat{\boldsymbol{e}}_3\}$. Space group operations $\{\, r \,|\, \boldsymbol{t} \,\}$ for which $\boldsymbol{t}$ is parallel to the invariant axis of r are called *screws*, while those for which $\boldsymbol{t}$ is parallel to an invariant plane of m are called *glides*. As we shall see, the possible values of $\boldsymbol{t}$ are strongly constrained in either case. Screws and glides may be considered *intrinsic translations* because they cannot be removed simply by a new choice of origin.

Next we note that if $\{\, g \,|\, \boldsymbol{t} \,\} \in \mathcal{S}$, we can always choose the translation component $\boldsymbol{t}$ to either be in the direct lattice or to lie within the first Wigner-Seitz (WS) cell[22]. If $\boldsymbol{t} \equiv \boldsymbol{\tau} \notin \mathcal{T}$, then it must be unique for a given g, because if both $\{\, g \,|\, \boldsymbol{\tau} \,\}$ and $\{\, g \,|\, \boldsymbol{\tau}' \,\}$ are in $\mathcal{S}$, then so is $\{\, g \,|\, \boldsymbol{\tau}' \,\}^{-1}\{\, g \,|\, \boldsymbol{\tau} \,\} = \{\, E \,|\, g^{-1}(\boldsymbol{\tau} - \boldsymbol{\tau}') \,\}$, which means that $g^{-1}(\boldsymbol{\tau} - \boldsymbol{\tau}') \in \mathcal{T}$ and therefore $\boldsymbol{\tau} - \boldsymbol{\tau}' \in \mathcal{T}$. Since by assumption both $\boldsymbol{\tau}$ and $\boldsymbol{\tau}'$ lie within the first WS cell, we must have $\boldsymbol{\tau}' = \boldsymbol{\tau}$. Thus, all space group elements are of the form $\{\, g \,|\, \boldsymbol{R} + \boldsymbol{\tau}_g \,\}$, where $\boldsymbol{\tau}_g$ may either be zero or a unique nonzero vector within the first WS cell. Now the point group $\mathcal{P}$ is of finite order, so each

[22] A translation $\boldsymbol{t}$ which is not a direct lattice vector can always be brought into the first WS cell by a direct lattice translation.

Nos.	lattice	$\mathcal{P}^{(\mathrm{Sch})}$	$\mathcal{P}^{(\mathrm{HM})}$	order	$\mathcal{S}$ (sym)	$\mathcal{S}$ (n-sym)
1	oblique	C_1	1	1	$p1$	
2	oblique	C_2	2	2	$p2$	
3 - 4	rectangular	C_{1v}	m	2	pm	pg
5 - 6	rectangular	C_{2v}	$2mm$	4	pmm	pmg
7	centered rectangular	C_{1v}	m	2	cm	
8 - 9	centered rectangular	C_{2v}	$2mm$	2	cmm	pgg
10	square	C_4	4	4	$p4$	
11 - 12	square	C_{4v}	$4mm$	8	$p4m$	$p4g$
13	hexagonal	C_3	3	3	$p3$	
14 - 15	hexagonal	C_{3v}	$3m$	6	$p3m1$, $p31m$	
16	hexagonal	C_6	6	6	$p6$	
17	hexagonal	C_{6v}	$6mm$	12	$p6m$	

Table 2.15: The 17 wallpaper groups and their short notation.

element $g \in \mathcal{P}$ satisfies $g^n = E$ where n is finite and taken to be the smallest positive integer which satisfies this relation. Therefore

$$\left\{ g \,\middle|\, \boldsymbol{\tau}_g \right\}^n = \left\{ g^n \,\middle|\, \boldsymbol{\tau}_g + g\boldsymbol{\tau}_g + \ldots + g^{n-1}\boldsymbol{\tau}_g \right\} \quad , \tag{2.59}$$

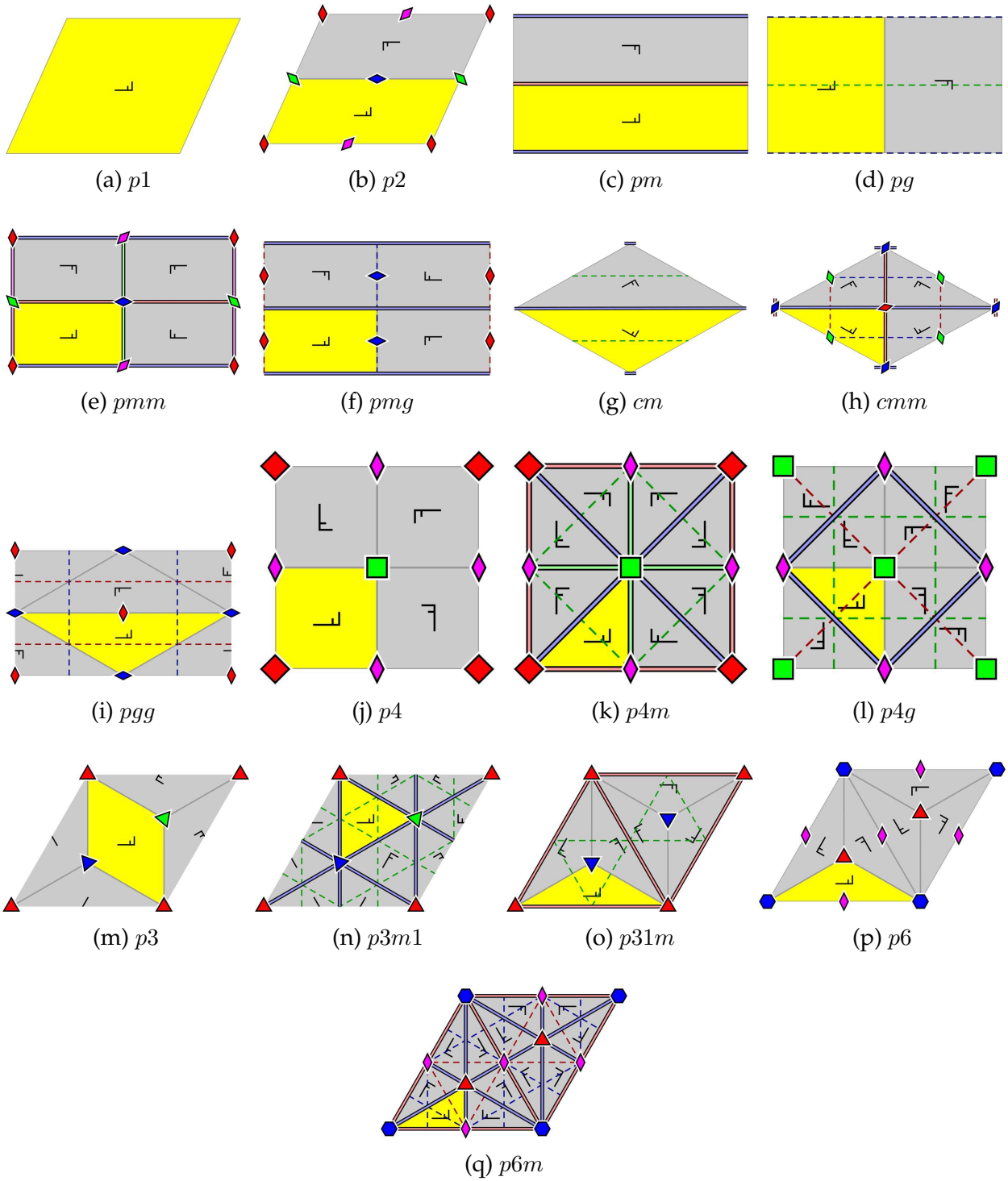
and since $g^n = E$, we must have that $\boldsymbol{\tau}_g + g\boldsymbol{\tau}_g + \ldots + g^{n-1}\boldsymbol{\tau}_g = \boldsymbol{R}$ is a direct lattice vector. Note that for $g = r$ we can have $n = 2$, 3, 4, or 6, while for $g = m$ we necessarily have $n = 2$.

According to Eqn. 2.58, we have

$$E + g + g^2 + \ldots + g^{n-1} = n\, P_{\scriptscriptstyle\parallel}(g) \quad , \tag{2.60}$$

where $P_{\scriptscriptstyle\parallel}(r) \equiv |\,\hat{\boldsymbol{e}}_1\,\rangle\langle\,\hat{\boldsymbol{e}}_1\,|$ is the projector onto the invariant axis of r, and $P_{\scriptscriptstyle\parallel}(m) \equiv |\,\hat{\boldsymbol{e}}_2\,\rangle\langle\,\hat{\boldsymbol{e}}_2\,| + |\,\hat{\boldsymbol{e}}_3\,\rangle\langle\,\hat{\boldsymbol{e}}_3\,|$ the projector onto the invariant plane of m. Thus we conclude $n\, P_{\scriptscriptstyle\parallel}(g)\,\boldsymbol{\tau}_g = \boldsymbol{R}$, which is to say that the nonremovable part of the translation $\boldsymbol{\tau}_g$, *i.e.* its projection onto the rotation axis or mirror plane, is equal to $\boldsymbol{R}/n$. Note also that in $d = 2$, there is no preserved rotation axis, since it would be orthogonal to the (x, y) plane. Therefore two dimensional point groups can at most have glides and no screws.

We may now identify all possible screws with the symbols 2_1, 3_1, 4_1, 4_2, 6_1, 6_2, and 6_3, as well as their enantiomorphous counterparts 3_2, 4_3, 6_4, and 6_5. The symbol n_m indicates a rotation



(a) $p1$ (b) $p2$ (c) pm (d) pg

(e) pmm (f) pmg (g) cm (h) cmm

(i) pgg (j) $p4$ (k) $p4m$ (l) $p4g$

(m) $p3$ (n) $p3m1$ (o) $p31m$ (p) $p6$

(q) $p6m$

Figure 2.20: Unit cells for the 17 two-dimensional space groups (wallpaper groups). (Image credit: *Wikipedia*.)

by $2\pi/n$ followed by a translation by a fraction m/n of a unit cell along the screw axis. Glide planes are denoted by the symbols a, b, c, n, and d, depending on the direction of the translation component. Let the symmetry axes of the crystal be $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. Then

- For a glides, $\boldsymbol{\tau} = \frac{1}{2}\boldsymbol{a}$.
- For b glides, $\boldsymbol{\tau} = \frac{1}{2}\boldsymbol{b}$.
- For c glides, $\boldsymbol{\tau} = \frac{1}{2}\boldsymbol{c}$.
- For n glides, $\boldsymbol{\tau} = \frac{1}{2}(\boldsymbol{a}+\boldsymbol{b})$, $\frac{1}{2}(\boldsymbol{b}+\boldsymbol{c})$, $\frac{1}{2}(\boldsymbol{a}+\boldsymbol{c})$, or $\frac{1}{2}(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c})$.
- For d glides, $\boldsymbol{\tau} = \frac{1}{4}(\boldsymbol{a}+\boldsymbol{b})$, $\frac{1}{4}(\boldsymbol{b}+\boldsymbol{c})$, $\frac{1}{4}(\boldsymbol{a}+\boldsymbol{c})$, or $\frac{1}{4}(\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c})$.

The d-glide is called the *diamond glide* and is present in the diamond lattice.

Be forewarned that it is possible for a *symmorphic* space group to include screw and glide operations provided they are *removable* by choosing a different origin. Such nonsymmorphic operations are called *inessential*. In other words, if $\mathcal{S}$ contains nonsymmorphic operations (screws or glides), but there exists some $\rho \equiv \{\, h \,|\, \boldsymbol{s} \,\}$ such that all elements of $\rho^{-1}\mathcal{S}\rho$ are of the form $\{\, g \,|\, \boldsymbol{R} \,\}$, then $\mathcal{S}$ is symmorphic. A nonsymmorphic space group contains *essential* (*i.e.* unremovable) screws or glides[23].

2.3.2 Factor groups

In the dim and distant past – specifically, in §1.3.1 – we discussed the concept of a *factor group*. Recall that if $H \subset G$ is a subgroup, there is a unique *left coset decomposition* of G as $G = \bigcup_i r_i\, H$ where $i \in \{1, \ldots, N_G/N_H\}$. If $H \triangleleft G$ is a normal subgroup, meaning $gHg^{-1} \in H$ for all $g \in G$, the cosets $r_i\, H$ form a group under multiplication, called the *factor group* G/H.

Since the abelian group $\mathcal{T}$ of Bravais lattice translations is a normal subgroup of the space group, we can decompose $\mathcal{S}$ as

$$\mathcal{S} = \bigcup_g \{\, g \,|\, \boldsymbol{\tau}_g \,\}\, \mathcal{T} = \mathcal{T} + \{\, g_2 \,|\, \boldsymbol{\tau}_{g_2} \,\}\, \mathcal{T} + \ldots + \{\, g_{N_{\mathcal{P}}} \,|\, \boldsymbol{\tau}_{g_{N_{\mathcal{P}}}} \,\}\, \mathcal{T} \quad . \tag{2.61}$$

This says that the space group $\mathcal{S}$ is generated by all Bravais lattice translations $\{\, E \,|\, \boldsymbol{R} \,\}$ and all operations $\{\, g \,|\, \boldsymbol{\tau}_g \,\}$. If, as in §2.3.5 below, we impose periodic boundary conditions, so that space is compactified into a three-dimensional torus of $N_1 \times N_2 \times N_3$ unit cells, then the translation group $\mathcal{T}$ is of finite order $|\mathcal{T}| = N_1 N_2 N_3$, and the order of the space group is $|\mathcal{S}| = |\mathcal{P}| \cdot |\mathcal{T}|$.

[23] As noted above, there are two nonsymmorphic space groups which contain neither screws nor glides, but for which one can nevertheless not write $\mathcal{S} = \mathcal{P} \rtimes \mathcal{T}$.

crystal system	type	symmorphic space groups
triclinic	P	$P1$, $P\overline{1}$
monoclinic	P	$P2$, Pm , $P2/m$
	A/C	$C2$, Cm , $C2/m$
orthorhombic	P	$P222$, $Pmm2$, $Pmmm$
	A/C	$C222$, $Cmm2$, $Cmmm$, $Amm2$
	I	$I222$, $Imm2$, $Immm$
	F	$F222$, $Fmm2$, $Fmmm$
tetragonal	P	$P4$, $P\overline{4}$, $P4/m$, $P422$, $P4mm$
		$P\overline{4}2m$, $P\overline{4}m2$, $P4/mmm$
	I	$I4$, $I\overline{4}$, $I4/m$, $I422$, $I4mm$
		$I\overline{4}2m$, $I\overline{4}m2$, $I4/mmm$
trigonal	P	$P3$, $P\overline{3}$, $P321$, $P3m1$, $P\overline{3}m1$
		$P312$, $P31m$, $P\overline{3}1m$
(rhombohedral)	R	$R3$, $R\overline{3}$, $R32$, $R3m$, $R\overline{3}m$
hexagonal	P	$P6$, $P\overline{6}$, $P6/m$, $P622$, $P6mm$
		$P\overline{6}m2$, $P\overline{6}2m$, $P6/mmm$
cubic	P	$P23$, $Pm3$, $P432$, $P\overline{4}3m$, $Pm\overline{3}m$
	I	$I23$, $Im3$, $I432$, $I\overline{4}3m$, $Im\overline{3}m$
	F	$F23$, $Fm3$, $F432$, $F\overline{4}3m$, $Fm\overline{3}m$

Table 2.16: The 73 symmorphic three-dimensional space groups and their short notation. Bravais lattice types are primitive (P), base-centered (A/C), body-centered (I), and face-centered (F). Space groups printed in red indicate cases where there are two inequivalent $\mathcal{P}$-invariant space lattice orientations.

The set of operations $\{\, g \,|\, \boldsymbol{\tau}_g \,\}$ is thus the factor group $\mathcal{F} \equiv \mathcal{S}/\mathcal{T}$. While there exists a bijective map $\{\, g \,|\, \boldsymbol{\tau}_g \,\} \longleftrightarrow \{\, g \,|\, \mathbf{0} \,\}$ between the factor group $\mathcal{F}$ and the point group $\mathcal{P}$, multiplication within the factor group is always modulo $\mathcal{T}$. Group multiplication of the factor group elements results in a projective representation of the point group,

$$\{\, g \,|\, \boldsymbol{\tau}_g \,\}\, \{\, h \,|\, \boldsymbol{\tau}_h \,\} = \{\, E \,|\, \boldsymbol{R}_{g,h} \,\}\, \{\, gh \,|\, \boldsymbol{\tau}_{gh} \,\} \quad , \tag{2.62}$$

and one can *lift* the projective representation of $\mathcal{P}$ to its *central extension*, which is to say $\mathcal{S}$. Here

$$\boldsymbol{R}_{g,h} = \boldsymbol{\tau}_g + g\boldsymbol{\tau}_h - \boldsymbol{\tau}_{gh} \tag{2.63}$$

must be in the Bravais lattice. Note that the cocycles here are actually translation operators rather than actual phases. Below we shall see how by diagonalizing the translation part of the space group, the cocycles become phases.

The case of diamond

Diamond is a rather typical nonsymmorphic space group. Recall the primitive direct lattice vectors for the fcc Bravais lattice,

$$\boldsymbol{a}_1 = \tfrac{1}{2}a_0\,(0,1,1) \quad , \quad \boldsymbol{a}_2 = \tfrac{1}{2}a_0\,(1,0,1) \quad , \quad \boldsymbol{a}_3 = \tfrac{1}{2}a_0\,(1,1,0) \quad , \tag{2.64}$$

where a_0 is the side length of the simple cubic lattice whose four element basis describes the fcc structure. The space group of diamond is $\mathcal{S} = Fd\overline{3}m$, this the point group is $m\overline{3}m$, which is O_h. Thus there are 48 cosets in the factor group $\mathcal{F}$, which is the order of O_h. These cosets break up into two collections. One consists of operations of the form $\left\{\, h \,\middle|\, \mathbf{0} \,\right\} \mathcal{T}$ where $h \in T_d$. The other consists of operations of the form $\left\{\, I \,\middle|\, \boldsymbol{\tau} \,\right\} \left\{\, h \,\middle|\, \mathbf{0} \,\right\} \mathcal{T}$ where I is the inversion operator and $\boldsymbol{\tau} = \frac{1}{4}\boldsymbol{a}_1 + \frac{1}{4}\boldsymbol{a}_2 + \frac{1}{4}\boldsymbol{a}_3 = \frac{1}{4}a_0\,(1,1,1)$. The elements from the first collection thus constitute a group in their own right, which is the zincblende space group $\widetilde{\mathcal{S}} = F\overline{4}3m$. This is a normal subgroup of $\mathcal{S}$ of index two, *i.e.* $\mathcal{S}/\widetilde{\mathcal{S}} \cong \mathbb{Z}_2$. Explicitly, we then have $\mathcal{S} = \widetilde{\mathcal{S}} \cup \left\{\, I \,\middle|\, \boldsymbol{\tau} \,\right\} \widetilde{\mathcal{S}}$.

2.3.3 How to make a symmorphic space group

The simplest recipe:

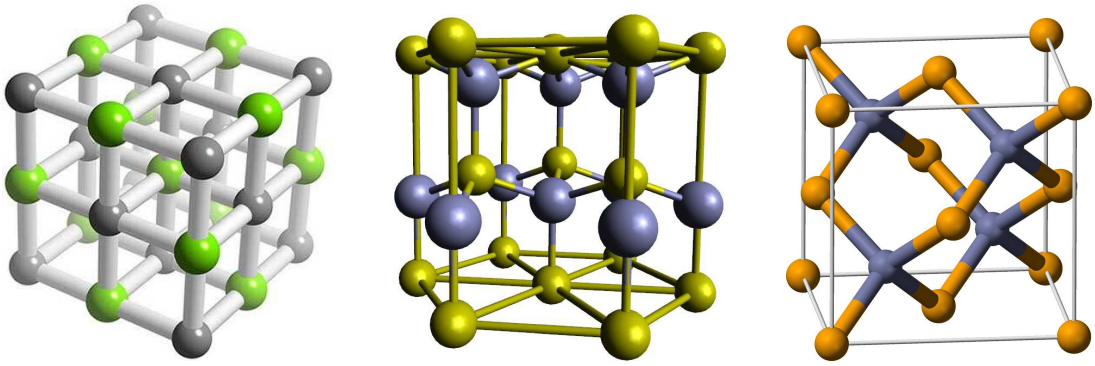
(i) Start with a lattice system.

(ii) Choose a point group consistent with the lattice system.

(iii) Choose an allowed lattice type (*i.e.* centering).

(iv) Congratulations, you've just specified a symmorphic point group.

To see this method in practice, let's try it out in two dimensions, where there 13 of the 17 space (wallpaper) groups are symmorphic. There are four crystal systems (oblique, rectangular, square, hexagonal), and the rectangular system can either have a primitive or a centered unit cell. For oblique lattices the allowed point groups are C_1 and C_2, so two possibilities. For rectangular lattices, the allowed point groups are C_{1v} and C_{2v}. There are two possible centerings, for a total of four possibilities. For square lattices, $\mathcal{P}$ can be either C_4 or C_{4v} – another two. For hexagonal, either C_3, C_{3v}, C_6, or C_{6v}, so four total. We arrive at 12 so we are missing a space group. The reason is there can be two inequivalent orientations of the space lattice which the point group leaves invariant, thereby leading to another space group. This happens in the



(a) Rocksalt, $Fm\overline{3}m$ (b) Wurzite, $P6_3mc$ (c) Zincblende, $F\overline{4}3m$

Figure 2.21: Some common AB crystal structures and their space groups.

case of the hexagonal lattice with C_{3v} ($3m$) point group symmetry. There are two space groups, called $p3m1$ and $p31m$.

A table of the 17 wallpaper groups is provided in Tab. 2.15, and sketches of the unit cells of each of them are depicted in Fig. 2.20. Study the nonsymmorphic cases pg, pmg, pgg, and $p4g$ to see if you can identify the glide mirrors. Note also how the naming convention works: the leading p or c character stands for primitive or centered. Information about the point group is contained in the space group label. Finally, the symbol g is used to indicate the presence of a glide mirror.

The naming convention for three-dimensional space groups is somewhat more complex, but the procedure is as described in the above recipe. There are seven distinct crystal systems, and Bravais lattice types are either primitive (P), base-centered (A/C), body-centered (I), or face-centered (F). Consider an fcc lattice with point group O_h ($m\overline{3}m$ in HM short notation). The corresponding symmorphic space group is $Fm\overline{3}m$, the full symbol for which is $F\frac{4}{m}\overline{3}\frac{2}{m}$. Proceeding in this way, accounting for all the crystal systems, their allowed point groups, and possible centerings, one obtains 66 symmorphic space groups. As in the two-dimensional case, when inequivalent orientations of the space lattice are both preserved by the point group, we get an extra space group. Such cases are indicated in red in Tab. 2.16. For example, for the case $C_{2v} = mm2$, the A and C centering types lead to different space groups, $Amm2$ and $Cmm2$, respectively. They are distinct space groups because in the latter case the centering is along a twofold axis, while in the former it is not.

2.3.4 Nonsymmorphic space groups

Returning to our example space group $F\frac{4}{m}\overline{3}\frac{2}{m}$, a check of the tables[24] reveals that there are a total of four space groups generated from the fcc lattice and point group $O_h = \frac{4}{m}\overline{3}\frac{2}{m}$. The other

[24] See http://www.wikiwand.com/en/List_of_space_groups.

three are $F\frac{4}{m}\overline{3}\frac{2}{c}$, $F\frac{4_1}{d}\overline{3}\frac{2}{m}$, and $F\frac{4_1}{d}\overline{3}\frac{2}{c}$; their short names are $Fm\overline{3}c$, $Fd\overline{3}m$, and $Fd\overline{3}c$, respectively. These three space groups are all nonsymmorphic and involve either screws (4_1), glides (c, d), or both. The second of these three corresponds to carbon diamond. Schoenflies' names for the four point groups generated from fcc and O_h were O_h^5, O_h^6, O_h^7, and O_h^8, respectively, which convey little information other than the order in which he derived them from the point group O_h.[25]

Of the 230 three-dimensional space groups, 157 are nonsymmorphic and contain operations $\{ g \,|\, \boldsymbol{\tau}_g \}$ where $\boldsymbol{\tau}_g \notin \mathcal{T}$ is not in the direct lattice, and no single change of origin can reduce all the $\boldsymbol{\tau}_g$ to zero or to a direct lattice vector.

Some of the nonsymmorphic space groups with screw axes have mirror images, and together are known as *enantiomorphic pairs*. For example, space groups $(P4_1\,,\ P4_3)$ form such a pair, as do $(P4_12_12\,,\ P4_32_12)$, $(P3_112\,,\ P3_212)$, $(P6_222\,,\ P6_422)$, *etc.*

2.3.5 Translations and their representations

The set of translations $\mathcal{T}$ is a subgroup of $\mathcal{S}$, consisting of the elements $\{ E \,|\, \boldsymbol{R} \}$, where $\boldsymbol{R} = \sum_{j=1}^{d} m_j \, \boldsymbol{a}_j$ is a sum over the primitive direct lattice vectors with integer coefficients. It is convenient to work with discrete groups of finite order, so to this end we invoke periodic boundary conditions, which places our system on a d-dimensional torus extending for N_j unit cells in the $\boldsymbol{a}_j$ direction for each $j \in \{1, \ldots, d\}$. This means that $\boldsymbol{R}$ is equivalent to $\boldsymbol{R} + \sum_{j=1}^{d} l_j \, \boldsymbol{L}_j$ with $L_j = N_j \, \boldsymbol{a}_j$ and each $l_j \in \mathbb{Z}$. Our Bravais lattice translation group $\mathcal{T}$ now has $N = \prod_{j=1}^{d} N_j$ elements, which is the total number of unit cells in the real space torus.

Next we ask about irreducible representations of $\mathcal{T}$. Since $\mathcal{T}$ is an abelian group, all its IRREPs are one-dimensional. If $\psi(\boldsymbol{r})$ is a basis function for a unitary one-dimensional IRREP of $\mathcal{T}$, then

$$\{ E \,|\, \boldsymbol{R} \}\, \psi(\boldsymbol{r}) = \psi\big(\{ E \,|\, \boldsymbol{R} \}^{-1} \boldsymbol{r}\big) = \psi(\boldsymbol{r} - \boldsymbol{R}) = e^{-i\omega(\boldsymbol{R})}\, \psi(\boldsymbol{r}) \quad . \tag{2.65}$$

In order that the group multiplication law be satisfied, we must have $e^{-i\omega(\boldsymbol{R})}\, e^{-i\omega(\boldsymbol{R}')} = e^{-i\omega(\boldsymbol{R}+\boldsymbol{R}')}$, which tells us that $\omega(\boldsymbol{R})$ is linear in $\boldsymbol{R}$, *i.e.*

$$\omega(m_1 \, \boldsymbol{a}_1 + \ldots + m_d \, \boldsymbol{a}_d) = m_1 \, \omega(\boldsymbol{a}_1) + \ldots + m_d \, \omega(\boldsymbol{a}_d) \tag{2.66}$$

to within an additive multiple of 2π. We may define $\omega(\boldsymbol{a}_j) \equiv \theta_j$, in which case the IRREP is labeled by the set of angles $\boldsymbol{\theta}$. Furthermore, we must have $\omega(\boldsymbol{R}) = \omega(\boldsymbol{R} + \boldsymbol{L}_j)$ for all $j \in \{1, \ldots, d\}$, which says that $N_j \, \theta_j$ is congruent to zero modulo 2π, *i.e.* $\theta_j = 2\pi l_j / N_j$, where $l_j \in \{1, \ldots, N_j\}$. So the θ_j values are quantized and there are $N = \prod_j N_j$ distinct values of the vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$.

[25] Schoenflies' O_h^1 through O_h^4 correspond to primitive cubic lattices, and O_h^9 and O_h^{10} to bcc lattices.

system	$\mathcal{P}^{(\mathrm{Sch})}$	$\mathcal{P}^{\mathrm{HM}}$	$N_{\mathcal{P}}$	nonsymmorphic space groups
triclinic	C_1	1	1	none
	C_i	$\overline{1}$	2	none
monoclinic	C_2	2	2	$P2_1$
	C_s	m	2	Pc , Cc
	C_{2h}	$2/m$	4	$P2_1m$, $P2/c$, $P2_1/c$, $C2/c$
orthorhombic	D_2	222	4	$P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, $I2_12_12_1$
	C_{2v}	$mm2$	4	$Pmc2_1$, $Pcc2$, $Pma2$, $Pca2_1$, $Pnc2$, $Pmn2_1$, $Pba2$,
				$Pna2_1$, $Pnn2$
				$Cmc2_1$, $Ccc2$, $Abm2$, $Ama2$, $Aba2$, $Fdd2$, $Iba2$, $Ima2$
	D_{2h}	mmm	8	$Pnnn$, $Pccm$, $Pban$, $Pmma$, $Pnna$, $Pmna$, $Pcca$, $Pbam$,
				$Pccm$, $Pbcm$, $Pnnm$, $Pmmn$, $Pbcn$, $Pbca$, $Pnma$
				$Cmcm$, $Cmca$, $Cccm$, $Cmma$, $Ccca$, $Fddd$, $Ibam$, $Ibcm$, $Imma$
tetragonal	C_4	4	4	$P4_1$, $P4_2$, $P4_3$, $I4_1$
	S_4	$\overline{4}$	4	none
	C_{4h}	$4/m$	8	$P4_2/m$, $P4/n$, $P4_2/n$, $I4_1/a$
	D_4	422	8	$P42_12_1$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, $I4_122$
	C_{4v}	$4mm$	8	$P4bm$, $P4_2cm$, $P4_2nm$, $P4cc$, $P4nc$, $P4_2mc$, $P4_2bc$
				$I4cm$, $I4_1md$, $I\overline{4}2d$
	D_{2d}	$\overline{4}2m$	8	$P\overline{4}2c$, $P\overline{4}2_1m$, $P\overline{4}2_1c$, $P\overline{4}c2$, $P\overline{4}c2$, $P\overline{4}n2$, $I\overline{4}c2$, $I\overline{4}2d$
	D_{4h}	$4/mmm$	16	$P4/mcc$, $P4/nbm$, $P4/nnc$, $P4/mbm$, $P4/mnc$, $P4/nmm$,
				$P4/ncc$, $P4_2/mmc$, $P4_2/mcm$, $P4_2/nbc$, $P4_2/nnm$, $P4_2/mbc$,
				$P4_2/mnm$, $P4_2/nmc$, $P4_2/ncm$
				$I4/mcm$, $I4_1/amd$, $I4_1/acd$
trigonal	C_3	3	3	$P3_1$, $P3_2$
	S_6	$\overline{3}$	3	none
	D_3	32	6	$P3_112$, $P3_121$, $P3_212$, $P3_221$
	C_{3v}	$3m$	6	$P31c$, $P3c1$, $R3c$
	D_{3d}	$\overline{3}m$	12	$P\overline{3}1c$, $P\overline{3}c1$, $R\overline{3}c$
hexagonal	C_6	6	6	$P6_1$, $P6_2$, $P6_3$, $P6_4$, $P6_5$
	C_{3h}	$\overline{6}$	6	none
	C_{6h}	$6/m$	12	$P6_3/m$
	D_6	622	12	$P6_122$, $P6_222$, $P6_322$, $P6_422$, $P6_522$
	C_{6v}	$6mm$	12	$P6cc$, $P6_3cm$, $P6_3mc$
	D_{3h}	$\overline{6}m2$	12	$P\overline{6}c2$, $P\overline{6}2c$
	D_{6h}	$6/mmm$	24	$P6/mcc$, $P6_3/mcm$, $P6_3/mmc$
cubic	T	23	12	$P2_13$, $I2_13$
	T_h	$m\overline{3}$	24	$Pn\overline{3}$, $Pa\overline{3}$, $Fd\overline{3}$, $Ia\overline{3}$
	O	432	24	$P4_132$, $P4_232$, $P4_332$, $I4_132$, $F4_132$
	T_d	$\overline{4}3m$	24	$P\overline{4}3n$, $F\overline{4}3c$, $I\overline{4}3d$
	O_h	$m\overline{3}m$	48	$Pn\overline{3}n$, $Pm\overline{3}n$, $Pn\overline{3}n$, $Fm\overline{3}c$, $Fd\overline{3}m$, $Fd\overline{3}c$, $Ia\overline{3}d$

Table 2.17: The 157 nonsymmorphic three-dimensional space groups.

Recall the definition of the reciprocal lattice vectors $\boldsymbol{b}_j$ which satisfy $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\,\delta_{ij}$. Then if we define the wavevector $\boldsymbol{k} \equiv \sum_{j=1}^{d} \theta_j\,\boldsymbol{b}_j/2\pi$, we then have $\omega(\boldsymbol{R}) = \boldsymbol{k}\cdot\boldsymbol{R}$, and our basis functions may be written as $\psi_{\boldsymbol{k}}(\boldsymbol{r}) = u(\boldsymbol{r})\,e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$ where $u(\boldsymbol{r}-\boldsymbol{R}) = u(\boldsymbol{r})$ for all $\boldsymbol{R}\in\mathcal{L}$ is a periodic *cell function*. Any cell function may be expanded as a discrete Fourier series, *viz.*

$$u(\boldsymbol{r}) = \sum_{\boldsymbol{K}} C_{\boldsymbol{K}}\,e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \quad , \tag{2.67}$$

where $\boldsymbol{K} = \sum_{j=1}^{d} n_j\,\boldsymbol{b}_j$ is a reciprocal lattice vector, which satisfies $\exp(i\boldsymbol{K}\cdot\boldsymbol{R}) = 1$ for all direct lattice vectors $\boldsymbol{R}$, and the $\{C_{\boldsymbol{K}}\}$ are a set of coefficients. What we have just shown is known as *Bloch's theorem*, which says that the eigenfunctions of any Hamiltonian $\hat{H}$ which commutes with all Bravais lattice translations may be written in the form $\psi_{\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,u(\boldsymbol{r})$, where $u(\boldsymbol{r})$ is a cell function and $\boldsymbol{k}$ lies within the first Brillouin zone of the reciprocal lattice. The reason that $\boldsymbol{k}$ is confined to this region is that $\boldsymbol{k}\to\boldsymbol{k}+\boldsymbol{K}$ amounts to a change of the cell function $u(\boldsymbol{r})\to u(\boldsymbol{r})\,e^{i\boldsymbol{K}\cdot\boldsymbol{r}}$. Note that quantization of $\boldsymbol{\theta}$ entails quantization of $\boldsymbol{k}$ to one of N possible values.

The character of the space group element $\{\,E\,|\,\boldsymbol{R}\,\}$ in the $\boldsymbol{k}$ IRREP is thus $\chi^{(\boldsymbol{k})}(\boldsymbol{R}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$, in suitably abbreviated notation. The great orthogonality and completeness theorems then tell us

$$\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} = N\,\delta_{\boldsymbol{k},\boldsymbol{k}'} \qquad , \qquad \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} = N\,\delta_{\boldsymbol{R},\boldsymbol{R}'} \quad . \tag{2.68}$$

In the limit $N\to\infty$, these equations become

$$\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} = \hat{\Omega}\sum_{\boldsymbol{K}}\delta(\boldsymbol{k}'-\boldsymbol{k}-\boldsymbol{K}) \qquad , \qquad \Omega\int\limits_{\hat{\Omega}}\frac{d^dk}{(2\pi)^d}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} = \delta_{\boldsymbol{R},\boldsymbol{R}'} \quad . \tag{2.69}$$

The first of these is the generalized Poisson summation formula from Eqn. 2.7. In the second, the integral is over the first Brillouin zone, $\hat{\Omega}$. Recall $\mathsf{vol}(\Omega) = \Omega$ and $\mathsf{vol}(\hat{\Omega}) = \hat{\Omega} = (2\pi)^d/\Omega$.

2.3.6 Space group representations

We follow Lax §8.6 and §8.7. When solving for electronic or vibrational states of a crystal, the first order of business is to classify eigenstates by wavevector, *i.e.* to diagonalize the operations $\{\,E\,|\,\boldsymbol{R}\,\}$ in the space group $\mathcal{S}$. For states of crystal momentum $\boldsymbol{k}$, we have $\{\,E\,|\,\boldsymbol{R}\,\}\,|\,\boldsymbol{k},\lambda\,\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,|\,\boldsymbol{k},\lambda\,\rangle$, where λ denotes other quantum numbers not related to crystal momentum.

Acting on Bloch states, a general space group operation has the following action:

$$\begin{aligned}\{\,g\,|\,\boldsymbol{t}\,\}\,\overbrace{e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,u(\boldsymbol{r})}^{\psi_{\boldsymbol{k}}(\boldsymbol{r})} &= \exp\Big[i\boldsymbol{k}\cdot\{\,g\,|\,\boldsymbol{t}\,\}^{-1}\boldsymbol{r}\Big]\,u\big(\{\,g\,|\,\boldsymbol{t}\,\}^{-1}\boldsymbol{r}\big) \\ &= e^{ig\boldsymbol{k}\boldsymbol{r}}\,u\big(g^{-1}(\boldsymbol{r}-\boldsymbol{t})\big) \equiv e^{ig\boldsymbol{k}\cdot\boldsymbol{r}}\,e^{-ig\boldsymbol{k}\cdot\boldsymbol{t}}\,\widetilde{u}(\boldsymbol{r}) = \widetilde{\psi}_{g\boldsymbol{k}}(\boldsymbol{r}) \quad ,\end{aligned} \tag{2.70}$$

where if $u(\boldsymbol{r}) = \sum_{\boldsymbol{K}} C_{\boldsymbol{K}}\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}}$ is the original cell function, then

$$\widetilde{u}(\boldsymbol{r}) = \sum_{\boldsymbol{K}} C_{g^{-1}\boldsymbol{K}}\, e^{-i\boldsymbol{K}\cdot\boldsymbol{t}}\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \equiv \sum_{\boldsymbol{K}} \widetilde{C}_{\boldsymbol{K}}\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \tag{2.71}$$

is a new cell function, *i.e.* it satisfies $\widetilde{u}(\boldsymbol{r}+\boldsymbol{R}) = \widetilde{u}(\boldsymbol{r})$ for all direct lattice vectors $\boldsymbol{R}$. Thus, application of $\{\, g \,|\, \boldsymbol{t} \,\} \in \mathcal{S}$ to a Bloch function $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ generates a new Bloch function $\widetilde{\psi}_{g\boldsymbol{k}}(\boldsymbol{r})$ at wavevector $g\boldsymbol{k}$.[26]

Group and star of the wavevector $\boldsymbol{k}$

If $g\boldsymbol{k} = \boldsymbol{k} + \boldsymbol{K}$, then $\{\, g \,|\, \boldsymbol{t} \,\}$ does not change the wavevector of the Bloch function. We define the *point group* $\mathcal{P}_{\boldsymbol{k}}$ *of the wavevector* $\boldsymbol{k}$ to be those point group operations $g \in \mathcal{P}$ which leave $\boldsymbol{k}$ unchanged up to a reciprocal lattice vector[27]. The space group of the wavevector $\mathcal{S}_{\boldsymbol{k}}$ is then all $\{\, g \,|\, \boldsymbol{t} \,\} \in \mathcal{S}$ for which $g \in \mathcal{P}_{\boldsymbol{k}}$. The *star of the wavevector* $\boldsymbol{k}$ is defined to be the set of points including $\boldsymbol{k}$ and all its images $g\boldsymbol{k}$, where $g \in \mathcal{P} \setminus \mathcal{P}_{\boldsymbol{k}}$.[28]

Algebra and representation of the space group

Recall the results of Eqns. 2.62 and 2.63. From

$$\{\, g \,|\, \boldsymbol{\tau}_g \,\}\, \{\, h \,|\, \boldsymbol{\tau}_h \,\} = \{\, E \,|\, \boldsymbol{R}_{g,h} \,\}\, \{\, gh \,|\, \boldsymbol{\tau}_{gh} \,\} = \{\, gh \,|\, \boldsymbol{\tau}_{gh} \,\}\, \{\, E \,|\, (gh)^{-1}\boldsymbol{R}_{g,h} \,\} \quad , \tag{2.72}$$

we see that, acting on a Bloch state,

$$\{\, g \,|\, \boldsymbol{\tau}_g \,\}\, \{\, h \,|\, \boldsymbol{\tau}_h \,\}\, \psi_{\boldsymbol{k}}(\boldsymbol{r}) = e^{-igh\boldsymbol{k}\cdot\boldsymbol{R}_{g,h}}\, \{\, gh \,|\, \boldsymbol{\tau}_{gh} \,\}\, \psi_{\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{2.73}$$

and so if g and h are both elements of $\mathcal{P}_{\boldsymbol{k}}$, then

$$\{\, g \,|\, \boldsymbol{\tau}_g \,\}\, \{\, h \,|\, \boldsymbol{\tau}_h \,\} = e^{-i\boldsymbol{k}\cdot\boldsymbol{R}_{g,h}}\, \{\, gh \,|\, \boldsymbol{\tau}_{gh} \,\} \tag{2.74}$$

when acting on Bloch states of crystal momentum $\boldsymbol{k}$, where $\boldsymbol{R}_{g,h} = \boldsymbol{\tau}_g + g\boldsymbol{\tau}_h - \boldsymbol{\tau}_{gh}$ is a direct lattice vector. The above equation establishes a projective representation for $\mathcal{S}_{\boldsymbol{k}}$. Alternatively, one may define the operators

$$\Lambda_{\boldsymbol{k}}(g) \equiv \{\, g \,|\, \boldsymbol{\tau}_g \,\}\, e^{i\boldsymbol{k}\cdot\boldsymbol{\tau}_g} = \{\, g \,|\, \boldsymbol{R} + \boldsymbol{\tau}_g \,\}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\tau}_g)} \quad , \tag{2.75}$$

which act on states of crystal momentum $\boldsymbol{k}$, and which satisfy the projective algebra

$$\begin{aligned} \Lambda_{\boldsymbol{k}}(g)\,\Lambda_{\boldsymbol{k}}(h) &= \omega_{\boldsymbol{k}}(g,h)\,\Lambda(gh) \\ \omega_{\boldsymbol{k}}(g,h) &= e^{i\boldsymbol{k}\cdot(\boldsymbol{\tau}_h - g\boldsymbol{\tau}_h)} = e^{i\boldsymbol{K}_g\cdot\boldsymbol{\tau}_h} \end{aligned} \tag{2.76}$$

because $\boldsymbol{k}\cdot g\boldsymbol{\tau}_h = g^{-1}\boldsymbol{k}\cdot\boldsymbol{\tau}_h \equiv (\boldsymbol{k} - \boldsymbol{K}_g)\cdot\boldsymbol{\tau}_h$, with $\boldsymbol{K}_g = \boldsymbol{k} - g^{-1}\boldsymbol{k} = \boldsymbol{k} - \boldsymbol{k}g$.

[26] The phase $e^{-ig\boldsymbol{k}\cdot\boldsymbol{t}}$ amounts to a gauge transformation.

[27] $\mathcal{P}_{\boldsymbol{k}}$ is also known as the *little group* of $\boldsymbol{k}$.

[28] We use the notation $A \setminus B$ to denote set subtraction, with $B \subseteq A$. *I.e.* $A \setminus B = A - B$, which is to say the set of elements in A that are not in B.

Representations of symmorphic space groups

When $\mathcal{S}$ is symmorphic, $\boldsymbol{\tau}_g = 0$ for all $g \in \mathcal{P}$, hence $\omega_{\boldsymbol{k}}(g,h) = 1$ for all $\boldsymbol{k}$. We don't have to worry about projective representations of the little groups, and therefore

$$\begin{aligned} D^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \mathbf{0} \,\}\big) &= D^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \\ \chi^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \mathbf{0} \,\}\big) &= \chi^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \quad , \end{aligned} \tag{2.77}$$

i.e. we can use the ordinary point group representation matrices.

Representations of nonsymmorphic space groups

If $\boldsymbol{k} \notin \partial\hat{\Omega}$ lies in the interior of the Brillouin zone and not on its boundary, then both $\boldsymbol{k}$ and $g^{-1}\boldsymbol{k}$ lie inside $\hat{\Omega}$, which means $\boldsymbol{k}_g = 0$ and the cocycle is unity: $\omega_{\boldsymbol{k}}(g,h) = 1$. Thus we have

$$\begin{aligned} D^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \boldsymbol{\tau}_g \,\}\big) &= e^{-i\boldsymbol{k}\cdot\boldsymbol{\tau}_g}\, D^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \\ \chi^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \boldsymbol{\tau}_g \,\}\big) &= e^{-i\boldsymbol{k}\cdot\boldsymbol{\tau}_g}\, \chi^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \quad , \end{aligned} \tag{2.78}$$

where Γ can only be the trivial representation if $\boldsymbol{k} \neq \mathbf{0}$. Again, we only need the ordinary point group representation matrices.

If $\boldsymbol{k} \in \partial\hat{\Omega}$, then $\mathcal{P}_{\boldsymbol{k}}$ may be nontrivial. In this case there are two possibilities:

(i) If there is a one-dimensional IRREP of $\mathcal{S}_{\boldsymbol{k}}$, $d_{\boldsymbol{k}}(g)$, with $d_{\boldsymbol{k}}(g)\, d_{\boldsymbol{k}}(h) = \omega_{\boldsymbol{k}}(g,h)\, d_{\boldsymbol{k}}(gh)$, define the ratio $\widetilde{\Lambda}_{\boldsymbol{k}}(g) \equiv \Lambda_{\boldsymbol{k}}(g)/d_{\boldsymbol{k}}(g)$. The operators $\widetilde{\Lambda}_{\boldsymbol{k}}(g)$ then satisfy $\widetilde{\Lambda}_{\boldsymbol{k}}(g)\, \widetilde{\Lambda}_{\boldsymbol{k}}(h) = \widetilde{\Lambda}_{\boldsymbol{k}}(gh)$, *i.e.* the point group multiplication table. Thus,

$$\begin{aligned} D^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \boldsymbol{\tau}_g \,\}\big) &= e^{-i\boldsymbol{k}\cdot\boldsymbol{\tau}_g}\, d_{\boldsymbol{k}}(g)\, D^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \\ \chi^{\Gamma\,;\mathcal{S}_{\boldsymbol{k}}}\big(\{\, g \,|\, \boldsymbol{\tau}_g \,\}\big) &= e^{-i\boldsymbol{k}\cdot\boldsymbol{\tau}_g}\, d_{\boldsymbol{k}}(g)\, \chi^{\Gamma\,;\mathcal{P}_{\boldsymbol{k}}}(g) \quad . \end{aligned} \tag{2.79}$$

and again we can use the ordinary point group representations.

(ii) If there is no one-dimensional IRREP of $\mathcal{S}_{\boldsymbol{k}}$, if one wishes to avoid needless work, one can consult tables, *e.g.* in appendix F of Lax, or appendix C of Dresselhaus, Dresselhaus, and Jorio.

2.4 Fourier Space Crystallography

Thus far our understanding of crystallography has been based on real space structures and their transformation properties under point and space group operations. An equivalent approach, originally due to Bienenstock and Ewald (1962), and formalized and further developed

by Mermin and collaborators in the 1990s, focuses on the Fourier modes $\hat{\rho}(\boldsymbol{K})$ of the density $\rho(\boldsymbol{r})$. This is known in the literature as *Fourier space crystallography*[29]. Writing $\rho(\boldsymbol{r})$ as a Fourier sum,

$$\rho(\boldsymbol{r}) = \sum_{\boldsymbol{K}} \hat{\rho}(\boldsymbol{K})\, e^{i\boldsymbol{K}\cdot\boldsymbol{r}} \quad , \tag{2.80}$$

where each $\boldsymbol{K} \in \hat{\mathcal{L}}$. Since $\rho(\boldsymbol{r}) \in \mathbb{R}$ is real, we have $\hat{\rho}(-\boldsymbol{K}) = \hat{\rho}^*(\boldsymbol{K})$ for all $\boldsymbol{K} \in \hat{\mathcal{L}}$. The inverse of the above relation is

$$\hat{\rho}(\boldsymbol{K}) = \int d^d r\, \rho(\boldsymbol{r})\, e^{-i\boldsymbol{K}\cdot\boldsymbol{r}} \quad . \tag{2.81}$$

Note that if $\rho'(\boldsymbol{r}) = \rho(\boldsymbol{r} + \boldsymbol{d})$ then $\hat{\rho}'(\boldsymbol{K}) = \hat{\rho}(\boldsymbol{K})\, e^{i\chi(\boldsymbol{K})}$ where $\chi(\boldsymbol{K}) = \boldsymbol{K}\cdot\boldsymbol{d}$ is a linear function on $\hat{\mathcal{L}}$.

2.4.1 Space group symmetries

We now ask how the $\hat{\rho}(\boldsymbol{K})$ transform under space group operations of the crystal. The general space group operation may be written as $\big\{\, g \,\big|\, \boldsymbol{R} + \boldsymbol{\tau}_g \,\big\}$. We have already accounted for the symmetries under Bravais lattice translations, which says that $\rho(\boldsymbol{r})$ is given as the above Fourier sum. So now restrict our attention to operations of the form $\big\{\, g \,\big|\, \boldsymbol{\tau}_g \,\big\}$. If $\rho(\boldsymbol{r})$ is invariant under all space group operations, we must have

$$\rho(\boldsymbol{r}) = \big\{\, g \,\big|\, \boldsymbol{\tau}_g \,\big\}\, \rho(\boldsymbol{r}) = \rho\big(\big\{\, g \,\big|\, \boldsymbol{\tau}_g \,\big\}^{-1}\boldsymbol{r}\big) = \rho\big(g^{-1}(\boldsymbol{r} - \boldsymbol{\tau}_g)\big) \quad . \tag{2.82}$$

Taking the Fourier transform, we have

$$\hat{\rho}(\boldsymbol{K}) = \int d^d r\, \rho\big(g^{-1}(\boldsymbol{r} - \boldsymbol{\tau}_g)\big)\, e^{-i\boldsymbol{K}\cdot\boldsymbol{r}} = \hat{\rho}(\boldsymbol{K}g)\, e^{-i\boldsymbol{K}\cdot\boldsymbol{\tau}_g} \quad , \tag{2.83}$$

which is easily established by changing the integration variables[30] from $\boldsymbol{r}$ to $\boldsymbol{r}' = g^{-1}(\boldsymbol{r} - \boldsymbol{\tau}_g)$. Note that g denotes both an abstract element of the point group $\mathcal{P}$ as well as its 3×3 matrix representation, and that by $\boldsymbol{K}g$ we treat $\boldsymbol{K}$ as a row vector and multiply by the matrix of g on the right. We therefore have

$$\hat{\rho}(\boldsymbol{K}g) = \hat{\rho}(\boldsymbol{K})\, e^{i\phi_g(\boldsymbol{K})} \quad , \tag{2.84}$$

where $\phi_g(\boldsymbol{K}) = \boldsymbol{K}\cdot\boldsymbol{\tau}_g$ acts linearly on $\hat{\mathcal{L}}$, with $\phi_g(\boldsymbol{0}) \cong 0$ for all $g \in \mathcal{P}$ and $\phi_E(\boldsymbol{K}) \cong 0$ for all $\boldsymbol{K} \in \hat{\mathcal{L}}$. Here the symbol $\cong$ denotes equality modulo 2π. We call $\phi_g(\boldsymbol{K})$ a *phase function* on the reciprocal lattice.

We then have

$$\begin{aligned} \hat{\rho}(\boldsymbol{K}gh) &= \hat{\rho}(\boldsymbol{K}g)\, e^{i\phi_h(\boldsymbol{K}g)} = \hat{\rho}(\boldsymbol{K})\, e^{i\phi_g(\boldsymbol{K})}\, e^{i\phi_h(\boldsymbol{K}g)} \\ &= \hat{\rho}(\boldsymbol{K})\, e^{i\phi_{gh}(\boldsymbol{K})} \quad , \end{aligned} \tag{2.85}$$

[29] Here we follow the pedagogical treatment in A. König and N. D. Mermin, *Am. J. Phys.* **68**, 525 (2000), with some minor notational differences.

[30] Since $g \in \mathsf{O}(n)$, we have that the Jacobian of the transformation is $|\det g| = 1$.

and therefore the *group compatibility condition* for the phase functions is

$$\phi_{gh}(\boldsymbol{K}) \cong \phi_h(\boldsymbol{K}g) + \phi_g(\boldsymbol{K}) \quad , \tag{2.86}$$

which is the same condition as that in eqn. 2.63.

Suppose $\rho'(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$ differ by a translation. Then $\hat{\rho}'(\boldsymbol{K}) = \hat{\rho}(\boldsymbol{K})\, e^{i\chi(\boldsymbol{K})}$, hence

$$\begin{aligned} \hat{\rho}'(\boldsymbol{K}g) &= \hat{\rho}'(\boldsymbol{K})\, e^{i\phi'_g(\boldsymbol{K})} = \hat{\rho}(\boldsymbol{K})\, e^{i\chi(\boldsymbol{K})}\, e^{i\phi'_g(\boldsymbol{K})} \\ &= \hat{\rho}(\boldsymbol{K}g)\, e^{i\chi(\boldsymbol{K}g)} = \hat{\rho}(\boldsymbol{K})\, e^{i\phi_g(\boldsymbol{K})}\, e^{i\chi(\boldsymbol{K}g)} \quad , \end{aligned} \tag{2.87}$$

and therefore

$$\phi'_{\boldsymbol{g}}(\boldsymbol{K}) \cong \phi_g(\boldsymbol{K}) + \overbrace{\chi(\boldsymbol{K}g - \boldsymbol{K})}^{\chi(\boldsymbol{K}g) - \chi(\boldsymbol{K})} \quad . \tag{2.88}$$

We say that the the above equation constitutes a *gauge transformation* and thus that the functions $\phi'_g(\boldsymbol{K})$ and $\phi_g(\boldsymbol{K})$ are *gauge equivalent*. We then have the following:

- ◇ *A space group $\mathcal{S}$ is symmorphic iff there exists a gauge in which $\phi_g(\boldsymbol{K}) \cong 0$ for all $g \in \mathcal{P}$ and all $\boldsymbol{K} \in \hat{\mathcal{L}}$.*

2.4.2 Extinctions

In §2.1.5 we noted how in certain crystals, the amplitude of Bravais lattice Bragg peaks observed in a diffraction experiment can be reduced or even extinguished due to the crystal structure. Bragg peak extinction is thus a physical manifestation of the crystallographic point group symmetry, and as such must be encoded in the gauge-invariant content of the phase functions. Suppose that $\boldsymbol{K}g = \boldsymbol{K}$. Then

$$\hat{\rho}(\boldsymbol{K}) = \hat{\rho}(\boldsymbol{K}g) = \hat{\rho}(\boldsymbol{K})\, e^{i\phi_g(\boldsymbol{K})} \quad , \tag{2.89}$$

and thus if $\phi_g(\boldsymbol{K}) \ncong 0$, we necessarily have $\hat{\rho}(\boldsymbol{K}) = 0$, *i.e.* the Bragg peak at $\boldsymbol{K}$ is extinguished. $\boldsymbol{K}g = g^{\mathsf{T}}\boldsymbol{K} = \boldsymbol{K}$ means that $\boldsymbol{K}$ lies within the invariant subspace of g (and that of $g^{\mathsf{T}} = g^{-1}$ as well, of course). Now the only nontrivial ($g \neq E$) point group operations (in three dimensions) with invariant subspaces are (i) proper rotations r, and (ii) mirror reflections m. Every proper rotation has an invariant axis, and every mirror reflection has an invariant plane. We now consider the consequences of each for extinctions.

- *Mirrors* : If m is a mirror, then $m^2 = E$. Consider a reciprocal lattice vector $\boldsymbol{K} = \boldsymbol{K}m$ lying in the invariant plane of m. Then

$$0 \cong \phi_E(\boldsymbol{K}) \cong \phi_{m^2}(\boldsymbol{K}) \cong \phi_m(\boldsymbol{K}m) + \phi_m(\boldsymbol{K}) \cong 2\phi_m(\boldsymbol{K}) \quad . \tag{2.90}$$

Thus, $2\phi_m(\boldsymbol{K}) \cong 0$ which means either $\phi_m(\boldsymbol{K}) \cong 0$ or $\phi_m(\boldsymbol{K}) \cong \pi$. *Unless* $\phi_m(\boldsymbol{K}) = 0$ *for all* $\boldsymbol{K} = \boldsymbol{K}m$ *in the mirror plane, we say that* m *is a glide mirror.* Let $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ be basis vectors for the two-dimensional sublattice of $\hat{\mathcal{L}}$ in the invariant plane of m. Linearity of the phase functions says

$$\phi_m(n_1\,\boldsymbol{\beta}_1 + n_2\,\boldsymbol{\beta}_2) = n_1\,\phi_m(\boldsymbol{\beta}_1) + n_2\,\phi_m(\boldsymbol{\beta}_2) \quad . \tag{2.91}$$

Suppose now that $\phi_m(\boldsymbol{\beta}_1) \cong \phi_m(\boldsymbol{\beta}_2) \cong 0$. In this case, the mirror is ordinary and we have not a glide, *i.e.* there are no extinctions due to m. Next suppose $\phi_m(\boldsymbol{\beta}_1) \cong \pi$ and $\phi_m(\boldsymbol{\beta}_2) \cong 0$. In this case, we have extinctions for all $\boldsymbol{K} = n_1\,\boldsymbol{\beta}_1 + n_2\,\boldsymbol{\beta}_2$ with n_1 odd, for all n_2. A corresponding result holds for the case $\phi_m(\boldsymbol{\beta}_1) \cong 0$ and $\phi_m(\boldsymbol{\beta}_2) \cong \pi$. Finally, suppose $\phi_m(\boldsymbol{\beta}_1) \cong \phi_m(\boldsymbol{\beta}_2) \cong \pi$. Then $\boldsymbol{K}$ is extinguished whenever $n_1 + n_2$ is odd.

- *Proper rotations* : In this case, $r^n = E$ with $n = 2$, 3, 4, or 6. Suppose $\boldsymbol{K} = \boldsymbol{K}r$ lies along the invariant axis of r. Then

$$0 \cong \phi_E(\boldsymbol{K}) \cong \phi_{r^n}(\boldsymbol{K}) \cong n\,\phi_r(\boldsymbol{K}) \quad , \tag{2.92}$$

which says $\phi_r(\boldsymbol{K}) = 2\pi j/n$. If $\phi_r(\boldsymbol{K}) = 0$ for all $\boldsymbol{K} = \boldsymbol{K}r$, the rotation is ordinary. *If* $\phi_r(\boldsymbol{K}) \ncong 0$ *for any* $\boldsymbol{K} = \boldsymbol{K}r$ *along the invariant axis, we say that* r *is a screw.* Let $\boldsymbol{\beta}_1$ be the basis vector for $\boldsymbol{K}$ points along the invariant axis. Then $\phi_r(\boldsymbol{\beta}_1) \cong 2\pi j/n$, with $j \in \{0, \ldots, n-1\}$. The case $j = 0$ corresponds to an ordinary rotation. For $\boldsymbol{K} = l\,\boldsymbol{\beta}_1$, we have $\phi_r(\boldsymbol{K}) \cong 2\pi jl/n$, and Bragg vectors with $jl \neq 0$ modulo n are extinguished.

- *Special circumstances* : Suppose an n-fold proper rotation r lies within the invariant plane of a mirror m. Then $rmr = m$, *i.e.* $mrm = r^{-1}$. This is the case, for example, for the groups C_{nv}, D_{nd}, and D_{nh}. Let $\boldsymbol{K} = \boldsymbol{K}r = \boldsymbol{K}m$. Then

$$\begin{aligned} \phi_m(\boldsymbol{K}) &= \phi_{rmr}(\boldsymbol{K}) \cong \phi_{mr}(\boldsymbol{K}r) + \phi_r(\boldsymbol{K}) \\ &\cong \phi_r(\boldsymbol{K}rm) + \phi_m(\boldsymbol{K}r) + \phi_r(\boldsymbol{K}) \cong 2\phi_r(\boldsymbol{K}) + \phi_m(\boldsymbol{K}) \quad . \end{aligned} \tag{2.93}$$

We then have $2\phi_r(\boldsymbol{K}) \cong 0$, and so the screw symmetry is restricted to two possible cases: either $\phi_r(\boldsymbol{K}) \cong 0$ or $\phi_r(\boldsymbol{K}) \cong \pi$. Such a screw requires n even and $j = \frac{1}{2}n$.

Suppose next that the n-fold rotation axis is perpendicular to a mirror plane, as in the groups C_{nh} and D_{nh}. In this case $mr = rm$, and we have

$$\begin{aligned} \phi_{mr}(\boldsymbol{K}) &= \phi_r(\boldsymbol{K}m) + \phi_m(\boldsymbol{K}) \\ \phi_{rm}(\boldsymbol{K}) &= \phi_m(\boldsymbol{K}r) + \phi_r(\boldsymbol{K}) \quad . \end{aligned} \tag{2.94}$$

There are two interesting possibilities. First, if $\boldsymbol{K} = \boldsymbol{K}r$ is along the invariant axis of r, then $\boldsymbol{K}m = -\boldsymbol{K}$, and we have $\phi_r(\boldsymbol{K}) \cong \phi_r(-\boldsymbol{K}) \cong -\phi_r(\boldsymbol{K})$, hence $2\phi_r(\boldsymbol{K}) \cong 0$, which entails the same restrictions as in the case where $rmr = m$ analyzed above. Second, if $\boldsymbol{K}m = \boldsymbol{K}$, then we obtain $\phi_m(\boldsymbol{K}r) = \phi_m(\boldsymbol{K})$, which says that the diffraction pattern in the invariant plane, including any extinctions, is symmetric under the r operation.

2.4.3 Sticky bands

Consider now the Schrödinger equation $\hat{H}\psi = E\psi$, where[31]

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r}) \quad , \tag{2.95}$$

where $V(\boldsymbol{r})$ is invariant under space group operations. Typically $V(\boldsymbol{r})$ is purely due to (screened) Coulomb interactions between a given electron and the combined electron-ion charge density $\rho(\boldsymbol{r})$, in which case

$$V(\boldsymbol{r}) = \int d^d r'\, v(\boldsymbol{r} - \boldsymbol{r}')\, \rho(\boldsymbol{r}') \quad , \tag{2.96}$$

where $v(\boldsymbol{r}) = v(r)$ is the screened potential at separation $\boldsymbol{r}$. According to Bloch's theorem, eigenfunctions $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ of H are labeled by crystal momentum $\boldsymbol{k} \in \hat{\Omega}$ as well as by a *band index* n, and may be written as

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{K}} C_{n\boldsymbol{k}}(\boldsymbol{K})\, e^{i(\boldsymbol{K}+\boldsymbol{k})\cdot\boldsymbol{r}} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{2.97}$$

where $u_{n\boldsymbol{k}}(\boldsymbol{r}) = u_{n\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R})$ is the cell function for band n, which is periodic in the direct lattice. The Schrödinger equation for band n can then be written as

$$E\, C_{n\boldsymbol{k}}(\boldsymbol{K}) = \sum_{\boldsymbol{K}'} \overbrace{\left[\frac{\hbar^2(\boldsymbol{K}+\boldsymbol{k})^2}{2m}\,\delta_{\boldsymbol{K},\boldsymbol{K}'} + \hat{V}(\boldsymbol{K}-\boldsymbol{K}')\right]}^{\langle\, \boldsymbol{K}\,|\,\hat{H}(\boldsymbol{k})\,|\,\boldsymbol{K}'\,\rangle} C_{n\boldsymbol{k}}(\boldsymbol{K}') \quad , \tag{2.98}$$

where $\hat{V}(\boldsymbol{K}) = \hat{v}(\boldsymbol{K})\,\hat{\rho}(\boldsymbol{K})$, since the Fourier transform of a convolution is the product of the Fourier transforms. Since $v(\boldsymbol{r})$ is isotropic, we have $\hat{v}(\boldsymbol{q}\,g) = \hat{v}(\boldsymbol{q})$ for all $\boldsymbol{q}$, and therefore $\hat{V}(\boldsymbol{K}g) = \hat{V}(\boldsymbol{K})\, e^{i\phi_g(\boldsymbol{K})}$. Let us define $\hat{\omega}(\boldsymbol{q}) \equiv \hbar^2\boldsymbol{q}^2/2m$, which is the isotropic free particle dispersion. Note that

$$\hat{\omega}(\boldsymbol{K}g + \boldsymbol{k}) = \hat{\omega}\big((\boldsymbol{K}+\boldsymbol{k})\,g + (\boldsymbol{k} - \boldsymbol{k}\,g)\big) \quad . \tag{2.99}$$

We now (re-)introduce the notion of the *little group* of a wavevector:

> DEFINITION : Given a wavevector $\boldsymbol{k} \in \hat{\Omega}$, the set of all $g \in \mathcal{P}$ for which $\boldsymbol{K}_g \equiv \boldsymbol{k} - \boldsymbol{k}g$ is in $\hat{\mathcal{L}}$ is called the *little group* of $\boldsymbol{k}$, and notated $\mathcal{P}_{\boldsymbol{k}}$.

Since $\boldsymbol{k}g$ must also lie within $\hat{\Omega}$, we have that $\mathcal{P}_{\boldsymbol{k}} = \{E\}$ if $\boldsymbol{k}$ lies at a generic point in the interior of the first Brillouin zone, *i.e.* not along an axis of rotational symmetry or along a mirror plane.

[31] In this section, we will use hats to denote operators as well as Fourier transformed quantities, so keep on your toes to recognize the meaning of the hat symbol in context.

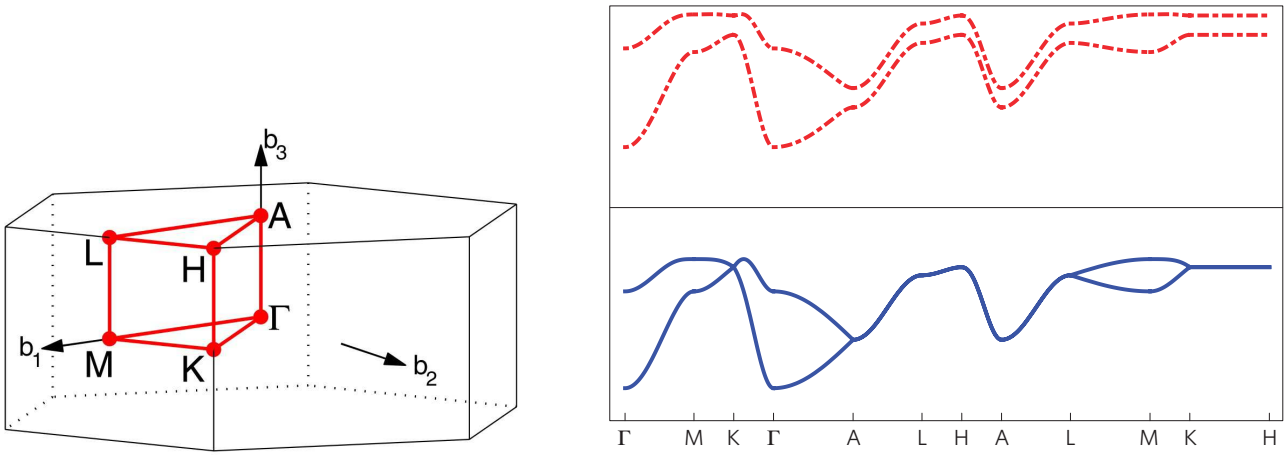


Figure 2.22: Stickiness of tight binding energy bands $\varepsilon_n(\boldsymbol{q})$ in an hcp crystal. Left: First Brillouin zone of the hexagonal Bravais lattice, with high symmetry points identified. Right: Tight binding energy levels for the hcp structure are shown in blue. Note the degeneracies at $\boldsymbol{q} = \mathrm{K}$, $\boldsymbol{q} = \mathrm{H}$, and all along the $\mathrm{A}-\mathrm{H}-\mathrm{L}-\mathrm{A}$ triangle on the top face and along the $\mathrm{K}-\mathrm{H}$ edge. When an alternating site energy on the two sublattices is present (dashed red curves), the screw symmetry is broken, and the space group is reduced from $P6_3/mmc$ to $P\overline{6}m2$.

However, if $\boldsymbol{k}$ lies in an invariant (one-dimensional) subspace of a proper rotation $r \in \mathcal{P}$ or within an invariant (two-dimensional) subspace of a mirror $m \in \mathcal{P}$, then $\boldsymbol{K}_r = \boldsymbol{k}r - \boldsymbol{k}$ and $\boldsymbol{K}_m = \boldsymbol{k}m - \boldsymbol{k}$ vanish, respectively, and the operations r and/or m are thus included in $\mathcal{P}_{\boldsymbol{k}}$.

For wavevectors $\boldsymbol{k} \in \partial\hat{\Omega}$ lying on the boundary of $\hat{\Omega}$, the little group $\mathcal{P}_{\boldsymbol{k}}$ can contain other elements. Consider for example the case of a square lattice, for which $\hat{\Omega}$ is itself a square, and let $\boldsymbol{k} = \frac{1}{2}\boldsymbol{b}_1$, which lies at the center of one of the edges. Let $\mathcal{P} = C_{4v}$, which is generated by r (90° rotation) and σ (x-axis reflection). Then E and σ are in $P_{\boldsymbol{k}}$ because they leave $\boldsymbol{k}$ fixed and hence $\boldsymbol{K}_g = \boldsymbol{0}$, but so are r^2 and σr^2, which send $\boldsymbol{k} \to -\boldsymbol{k}$, in which case $\boldsymbol{K}_g = \boldsymbol{b}_1 \in \hat{\mathcal{L}}$. It should be clear that $\mathcal{P}_{\boldsymbol{k}} \subset \mathcal{P}$ is a subgroup of the crystallographic point group, containing those operations $g \in \mathcal{P}$ which leave $\boldsymbol{k}$ invariant or changed by a reciprocal lattice vector. Note that if $g, h \in \mathcal{P}_{\boldsymbol{k}}$, then

$$\boldsymbol{K}_{gh} = \boldsymbol{k} - \boldsymbol{k}gh = (\boldsymbol{k} - \boldsymbol{k}h) + (\boldsymbol{k}h - \boldsymbol{k}gh) = \boldsymbol{K}_h + \boldsymbol{K}_g\, h \quad . \tag{2.100}$$

For each element g of the little group $\mathcal{P}_{\boldsymbol{k}}$, define the unitary operator $\hat{U}(g)$ such that

$$\hat{U}^\dagger(g) \,\big|\, \boldsymbol{K} \,\big\rangle = e^{i\phi_g(\boldsymbol{K})} \big|\, \boldsymbol{K}g - \boldsymbol{K}_g \,\big\rangle \quad . \tag{2.101}$$

We then have

$$\begin{aligned}\big\langle\, \boldsymbol{K} \,\big|\, \hat{U}(g)\,\hat{H}(\boldsymbol{k})\,\hat{U}^\dagger(g) \,\big|\, \boldsymbol{K}' \,\big\rangle &= \big\langle\, \boldsymbol{K}g - \boldsymbol{K}_g \,\big|\, \hat{H}(\boldsymbol{k}) \,\big|\, \boldsymbol{K}'g - \boldsymbol{K}_g \,\big\rangle\, e^{i\phi_g(\boldsymbol{K}'-\boldsymbol{K})} \\ &= \hat{\omega}\big(\boldsymbol{K}g - \boldsymbol{K}_g + \boldsymbol{K}\big)\,\delta_{\boldsymbol{K},\boldsymbol{K}'} + \hat{V}(\boldsymbol{K}-\boldsymbol{K}') \\ &= \hat{\omega}\big((\boldsymbol{K}+\boldsymbol{k})\,g\big)\,\delta_{\boldsymbol{K},\boldsymbol{K}'} + \hat{V}(\boldsymbol{K}-\boldsymbol{K}') = \big\langle\, \boldsymbol{K} \,\big|\, \hat{H}(\boldsymbol{k}) \,\big|\, \boldsymbol{K}' \,\big\rangle\end{aligned} \tag{2.102}$$

for all $\boldsymbol{k}$, $\boldsymbol{K}$, and $\boldsymbol{K}'$. This tells us that $\big[\hat{H}(\boldsymbol{k}), \hat{U}(g)\big] = 0$ for all $\boldsymbol{k} \in \hat{\Omega}$ and $g \in \mathcal{P}_{\boldsymbol{k}}$. Next, we have

$$\begin{aligned}\hat{U}^\dagger(h)\,\hat{U}^\dagger(g) \,\big|\, \boldsymbol{K} \,\big\rangle &= e^{i\phi_g(\boldsymbol{K})}\, e^{i\phi_h(\boldsymbol{K}g-\boldsymbol{K}_g)} \,\big|\, \boldsymbol{K}gh - \boldsymbol{K}_g\,h - \boldsymbol{K}_h \,\big\rangle \\ \hat{U}^\dagger(gh) \,\big|\, \boldsymbol{K} \,\big\rangle &= e^{i\phi_{gh}(\boldsymbol{K})} \,\big|\, \boldsymbol{K}gh - \boldsymbol{K}_{gh} \,\big\rangle \quad .\end{aligned} \tag{2.103}$$

Invoking Eqn. 2.100, we see that the ket vectors on the RHS of the above two equations are identical. Appealing to the compatibility condition Eqn. 2.86, we conclude $\hat{U}^\dagger(h)\,\hat{U}^\dagger(g) = \hat{U}^\dagger(gh)\,e^{-i\phi_h(\boldsymbol{K}_g)}$, *i.e.*

$$\hat{U}(g)\,\hat{U}(h) = \hat{U}(gh)\,e^{i\phi_h(\boldsymbol{K}_g)} \quad , \tag{2.104}$$

which is to say a projective representation of the little group.

Suppose $\hat{H}(\boldsymbol{k})\,|u_{\boldsymbol{k}}\rangle = E(\boldsymbol{k})\,|u_{\boldsymbol{k}}\rangle$, where $|u_{\boldsymbol{k}}\rangle$ is a Bloch cell function, and where we have dropped the band index n. Since $\big[\hat{H}(\boldsymbol{k}), \hat{U}(g)\big] = 0$, the state $\hat{U}(g)\,|u_{\boldsymbol{k}}\rangle$ is also an eigenstate of $\hat{H}(\boldsymbol{k})$ with eigenvalue $E(\boldsymbol{k})$. If $|u_{\boldsymbol{k}}\rangle$ is nondegenerate, then we must have $\hat{U}(g)\,|u_{\boldsymbol{k}}\rangle = \lambda_g(\boldsymbol{k})\,|u_{\boldsymbol{k}}\rangle$ for all $g \in \mathcal{P}_{\boldsymbol{k}}$. But then $\big[\hat{U}(g), \hat{U}(h)\big]|u_{\boldsymbol{k}}\rangle = 0$, which means

$$e^{i\phi_h(\boldsymbol{K}_g)}\,\hat{U}(gh)\,|u_{\boldsymbol{k}}\rangle = e^{i\phi_g(\boldsymbol{K}_h)}\,\hat{U}(hg)\,|u_{\boldsymbol{k}}\rangle \quad . \tag{2.105}$$

Thus, if $gh = hg$, we must have either (i) $\phi_h(\boldsymbol{K}_g) = \phi_g(\boldsymbol{K}_h)$ or else (ii) $|u_{\boldsymbol{k}}\rangle = 0$, *i.e.* there is no such nondegenerate eigenstate at wavevector $\boldsymbol{k}$. Therefore,

⋆ *If $gh = hg$ and $\phi_h(\boldsymbol{K}_g) \neq \phi_g(\boldsymbol{K}_h)$, all the eigenstates of $\hat{H}(\boldsymbol{k})$ appear in degenerate multiplets.*

That is, two or more bands become "stuck" together at these special $\boldsymbol{k}$ points. Note that the sticking conditions cannot be satisfied in a symmorphic space group, because the phase functions can all be set to zero by a choice of gauge (*i.e.* by a choice of origin for the point group operations). Note also that under a gauge transformation, the change in $\phi_h(\boldsymbol{K}_g) - \phi_g(\boldsymbol{K}_h)$ is

$$\Delta\big(\phi_h(\boldsymbol{K}_g) - \phi_g(\boldsymbol{K}_h)\big) = \chi(\boldsymbol{K}_g\,h - \boldsymbol{K}_g) - \chi(\boldsymbol{K}_h\,g - \boldsymbol{K}_h) = \chi(\boldsymbol{k}\,gh - \boldsymbol{k}\,hg) \quad , \tag{2.106}$$

which vanishes when $gh = hg$.

Since $\phi_g(\boldsymbol{0}) = 0$ for all g, the sticking conditions require that either $\boldsymbol{K}_g$ or $\boldsymbol{K}_h$ be nonzero. This is possible only when $\boldsymbol{k} \in \partial\hat{\Omega}$ lies on the boundary of the first Brillouin zone, for otherwise

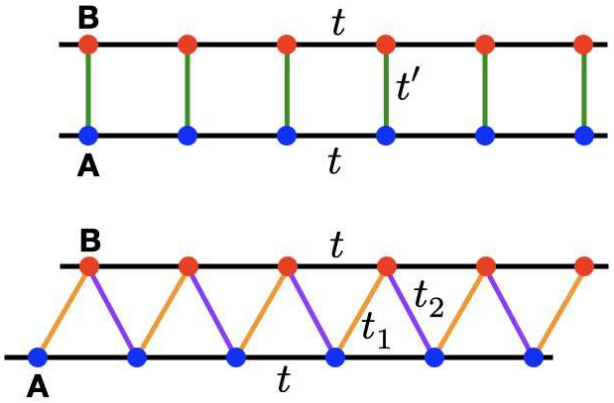


Figure 2.23: Two one-dimensional lattices: the ladder (top) and the trestle (bottom). The trestle is nonsymmorphic when the two classes of diagonal links are equivalent, *i.e.* when $t_1 = t_2$.

the vectors $\boldsymbol{K}_g$ and $\boldsymbol{K}_h$ are too short to be reciprocal lattice vectors[32]. Thus, in nonsymmorphic crystals, band sticking occurs only along the boundary. Consider, for example, the case of diamond, with nonsymmorphic space group $F\frac{4_1}{d}\overline{3}\frac{2}{m}$ ($Fd\overline{3}m$ in the short notation). The diamond structure consists of two interpenetrating fcc Bravais lattices, and exhibits a 4_1 screw axis and a diamond (d) glide[33]. Let $\boldsymbol{k} = \frac{1}{2}\boldsymbol{K}$, where $\boldsymbol{K}$ is the shortest reciprocal lattice vector along the screw axis. Then $\boldsymbol{K}_r = \boldsymbol{k} - \boldsymbol{k}r = 0$ because $\boldsymbol{k}$ is along the invariant axis of the fourfold rotation r, hence r is in the little group. Diamond is centrosymmetric, meaning that its point group contains the inversion operator I, which commutes with all other point group elements. Clearly $\boldsymbol{K}_I = \boldsymbol{k} - I\boldsymbol{k} = 2\boldsymbol{k} = \boldsymbol{K}$, so I is in the little group as well. The sticking conditions then require $\phi_r(\boldsymbol{K}) \neq 0$, which is the condition we found for r to be a screw in the first place. So we have band sticking at $\boldsymbol{k} = \frac{1}{2}\boldsymbol{K}$. This is a special case of the following general rule: in nonsymmorphic centrosymmetric crystals, there is band sticking at every $\boldsymbol{k} = \frac{1}{2}\boldsymbol{K}$ where $\boldsymbol{K}$ is an extinguished reciprocal lattice (Bragg) vector.

Band sticking can also occur along *continuous lines* along the zone boundary. This is possible when the point group contains perpendicular mirrors, such as in the case D_{3h}. Let $\boldsymbol{k}$ lie along the line where the horizontal Brillouin zone surface intersects the vertical mirror plane. The vertical component of $\boldsymbol{k}$ is thus $\frac{1}{2}\boldsymbol{K}$, where $\boldsymbol{K}$ is the shortest vertical reciprocal lattice vector, but otherwise $\boldsymbol{k}$ can lie anywhere along this line. Then $\boldsymbol{K}_{m_h} = \boldsymbol{K}$ and $\boldsymbol{K}_{m_v} = 0$, for all $\boldsymbol{k}$ along the line, where $m_{h,v}$ are the horizontal and vertical mirror operations, respectively. The sticking condition is $\phi_{m_v}(\boldsymbol{K}) \neq 0$, which says that m_v is a glide mirror and $\boldsymbol{K}$ is extinguished. Introducing a perturbation which breaks the nonsymmorphic symmetries unsticks the bands and revives the extinguished Bragg vectors. An example is the hcp structure, shown in Fig. 2.22.

As an example, consider the two one-dimensional lattices in fig. 2.23. We shall call these the

[32] My childhood dreams of becoming a reciprocal lattice vector were dashed for the same reason.
[33] Diamond has a diamond (d) glide. The d is for "duh".

d	Name	Examples	Space group	$\mathcal{S}$
2	Shastry–Sutherland	$SrCu_2(BO_3)_2$	*p*4*g*	2
3	hcp	Be, Mg, Zn	$P6_3/mmc$	2
3	Diamond	C, Si	$Fd\bar{3}m$	2
3	Pyrochlore	$Dy_2Ti_2O_7$ (spin ice)	$Fd\bar{3}m$	2
3	–	α-SiO_2, GeO_2	$P3_121$	3
3	–	$CrSi_2$	$P6_222$	3
3	–	$Pr_2Si_2O_7$, $La_2Si_2O_7$	$P4_1$	4
3	Hex. perovskite	$CsCuCl_3$	$P6_1$	6

Figure 2.24: Examples of space groups and their nonsymmorphic ranks.

ladder (top) and the trestle (bottom). Their respective tight binding Hamiltonians are given by

$$\hat{H}^{\text{ladder}}(k) = -\begin{pmatrix} 2t\cos k & t' \\ t' & 2t\cos k \end{pmatrix} \quad , \quad \hat{H}^{\text{trestle}}(k) = -\begin{pmatrix} 2t\cos k & t_1 + t_2\, e^{-ik} \\ t_1 + t_2\, e^{ik} & 2t\cos k \end{pmatrix} \quad . \tag{2.107}$$

The dispersions are

$$\begin{aligned} E_\pm^{\text{ladder}}(k) &= -2t\cos k \pm t' \\ E_\pm^{\text{trestle}}(k) &= -2t\cos k \pm \left| t_1 + t_2 e^{-ik} \right| \quad . \end{aligned} \tag{2.108}$$

The gaps $\Delta(k) = E_+(k) = E_-(k)$ are given by

$$\begin{aligned} \Delta^{\text{ladder}}(k) &= 2t' \\ \Delta^{\text{trestle}}(k) &= 2\left| t_1 + t_2 e^{-ik} \right| = 2\sqrt{(t_1 - t_2)^2 + 4t_1 t_2 \cos^2(\tfrac{1}{2}k)} \quad . \end{aligned} \tag{2.109}$$

Note that the ladder always has a direct gap, whereas the gap for the trestle vanishes at $k = \pi$ provided that $t_1 = t_2$. This latter condition ensures that the trestle is nonsymmorphic, where the nonsymmorphic space group element is reflection about the midline parallel to the two legs, followed by a lattice translation of half a unit cell.

A more detailed result was derived by Michel and Zak[34]. In nonsymmorphic crystals, energy bands stick together in groups of $\mathcal{S}$, where $\mathcal{S} = 2$, 3, 4, or 6 is the *nonsymmorphic rank* of the space group. In such cases, groups of $\mathcal{S}$ bands are stuck at high symmetry points or along high symmetry lines in the Brillouin zone, and one must fill an integer multiple of $\mathcal{S}$ bands of spinless electrons in order to construct a band insulating state.

[34] L. Michel and J. Zak, *Phys. Rev. B* **59**, 5998 (1999).

Chapter 3

Deformations of Crystals

3.1 Elasticity

3.1.1 Stress and strain tensors

An elastic medium is described by a local deformation field $\boldsymbol{u}(\boldsymbol{r})$, corresponding to the elastic displacement of the solid at $\boldsymbol{r}$. The *strain tensor* is defined by the dimensionless expression

$$\varepsilon_{ij}(\boldsymbol{r}) = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \quad . \tag{3.1}$$

Note that $\varepsilon = \varepsilon^{\mathsf{T}}$ is a symmetric tensor by definition. Similarly, the *stress tensor* $\sigma_{ij}(\boldsymbol{r})$ is defined by

$$dF_i(\boldsymbol{r}) = -\sigma_{ij}(\boldsymbol{r})\, n_j\, d\Sigma \quad , \tag{3.2}$$

where $d\boldsymbol{F}(\boldsymbol{r})$ is the differential force on a surface element $d\Sigma$ whose normal is the vector $\hat{\boldsymbol{n}}$. Angular momentum conservation requires that the stress tensor also be symmetric[1]. The stress and strain tensors are related by the rank four *elastic modulus tensor*, *viz.*

$$\sigma_{ij}(\boldsymbol{r}) = C_{ijkl}\,\varepsilon_{kl}(\boldsymbol{r}) = \frac{\delta f}{\delta \varepsilon_{ij}(\boldsymbol{r})} \quad , \tag{3.3}$$

where the second equality is a statement of thermal equilibrium akin to $p = -\partial F/\partial V$. Here,

$$f(\boldsymbol{r}) = f_0 + \tfrac{1}{2}\, C_{ijkl}\,\varepsilon_{ij}(\boldsymbol{r})\,\varepsilon_{kl}(\boldsymbol{r}) + \mathcal{O}(\varepsilon^3) \tag{3.4}$$

[1] Integrate the differential torque $d\boldsymbol{N} = \boldsymbol{r} \times d\boldsymbol{F}$ over the entire body. Integrating by parts, one obtains a surface term and a volume term. The volume torque density is $-\epsilon_{ijk}\,\sigma_{jk}$, which must vanish, thereby entailing the symmetry $\sigma = \sigma^{\mathsf{T}}$.

is the local free energy density. Since ε is a dimensionless tensor, the elastic moduli have dimensions of energy density, typically expressed in cgs units as $\mathrm{dyn/cm^2}$. For an *isotropic* material, the only $\mathsf{O}(3)$ invariant terms in the free energy to order ε^2 are proportional to either $(\mathsf{Tr}\ \varepsilon)^2$ or to $\mathsf{Tr}(\varepsilon^2)$. Thus,

$$f = f_0 + \tfrac{1}{2}\lambda\,(\mathsf{Tr}\ \varepsilon)^2 + \mu\ \mathsf{Tr}\,(\varepsilon^2) \quad . \tag{3.5}$$

The parameters λ and μ are called the *Lamé coefficients*[2]. For isotropic elastic materials, then,

$$\sigma_{ij} = \frac{\partial f}{\partial \varepsilon_{ij}} = \lambda\ \mathsf{Tr}\ \varepsilon\ \delta_{ij} + 2\mu\,\varepsilon_{ij} \quad . \tag{3.6}$$

In the literature, one often meets up with the quantity $K \equiv \lambda + \frac{2}{3}\mu$, in which case the free energy density becomes

$$f = f_0 + \tfrac{1}{2}K\,(\mathsf{Tr}\ \varepsilon)^2 + \mu\ \mathsf{Tr}\left(\varepsilon - \tfrac{1}{3}\ \mathsf{Tr}\ \varepsilon\right)^2 \tag{3.7}$$

The reason is that the tensor $\widetilde{\varepsilon} \equiv \varepsilon - \frac{1}{3}(\mathsf{Tr}\ \varepsilon)\cdot\mathbf{1}$ is traceless, and therefore the constant K tells us about *bulk deformations* while μ tells us about *shear deformations*. One then requires $K > 0$ and $\mu > 0$ for thermodynamic stability. We then may write, for isotropic materials,

$$\begin{aligned} \sigma &= K\,(\mathsf{Tr}\ \varepsilon)\cdot\mathbf{1} + 2\mu\,\widetilde{\varepsilon} \\ \varepsilon &= \frac{1}{9K}\,(\mathsf{Tr}\ \sigma)\cdot\mathbf{1} + \frac{1}{2\mu}\,\widetilde{\sigma} \quad , \end{aligned} \tag{3.8}$$

with $\widetilde{\sigma} \equiv \sigma - \frac{1}{3}(\mathsf{Tr}\ \sigma)\cdot\mathbf{1}$ the traceless part of the stress tensor[3].

If one solves for the homogeneous deformation[4] of a rod of circular cross section, the only nonzero element of the stress tensor is $\sigma_{zz} = p$, where p is the pressure on either of the circular faces of the rod. One then finds that $\varepsilon_{xx} = \varepsilon_{yy} = \left(\frac{1}{9K} - \frac{1}{6\mu}\right)p$ and $\varepsilon_{zz} = \left(\frac{1}{9K} + \frac{1}{3\mu}\right)p$ are the only nonzero elements of the strain tensor. Thus,

$$Y \equiv \frac{\sigma_{zz}}{\varepsilon_{zz}} = \frac{9K\mu}{3K+\mu} \qquad , \qquad \beta \equiv -\frac{\varepsilon_{xx}}{\varepsilon_{zz}} = \frac{3K-2\mu}{2(3K+\mu)} \quad . \tag{3.9}$$

The quantity Y is called the *Young's modulus*, and must be positive. The quantity β is the *Poisson ratio* β and satisfies $\beta \in \left[-1, \frac{3}{2}\right]$. A material like tungsten carbide has a very large Young's modulus of $Y = 53.4 \times 10^{11}\,\mathrm{dyn/cm^2}$ at STP, which means that you have to pull like hell in order to get it to stretch a little. Normally, when you stretch a material, it narrows in the transverse directions, which corresponds to a positive Poisson ratio. Materials for which $\beta < 0$ are called *auxetics*. When stretched, an auxetic becomes thicker in the directions perpendicular to the applied force. Examples include various porous foams and artificial macrostructures.

[2] If you were wondering why we've suddenly switched to roman indices C_{ijkl} instead of Greek $C_{\alpha\beta\mu\nu}$, it is to obviate any confusion with the Lamé parameter μ.

[3] In d space dimensions, one has $K = \lambda + 2d^{-1}\mu$ and $\widetilde{m} = m - d^{-1}\ \mathsf{Tr}\ m$ is the traceless part of any matrix m.

[4] In a homogeneous deformation, the strain and stress tensors are constant throughout the body.

$(\alpha\beta)$:	(11)	(22)	(33)	(23)	(31)	(12)
a :	1	2	3	4	5	6

Table 3.1: Abbreviation for symmetric compound indices $(\alpha\beta)$.

3.1.2 Elasticity and symmetry

Since

$$C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij} \quad , \tag{3.10}$$

we may use the composite index notation in Tab. 3.1 to write the rank four tensor $C_{ijkl} \equiv C_{ab} = C_{ba}$ as a symmetric 6×6 matrix, with 21 independent elements before accounting for symmetry considerations. The linear stress-strain relation is then given by

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{13} & C_{23} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{24} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{25} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{26} & C_{36} & C_{46} & C_{56} & C_{66} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ 2\varepsilon_4 \\ 2\varepsilon_5 \\ 2\varepsilon_6 \end{pmatrix} \quad . \tag{3.11}$$

Since the elastic tensor is rank four, it is symmetric under inversion.

And now, let the symmetry commence!

- For triclinic crystals with point group C_1 or C_i, there are no symmetries to apply to C_{ab}, hence there are 21 independent elastic moduli. However, one can always rotate axes, and given the freedom to choose three Euler angles, this means we can always choose axes in such a way that three of the 21 moduli vanish, leaving 18. Again, this requires a nongeneric choice of axes.

- For monoclinic crystals, there is symmetry under $z \to -z$, and as in the example of the piezoelectric tensor $d_{\mu\nu\lambda}$, we have that C_{ijkl} vanishes if the index 3(z) appears an odd number of times, which means, in composite index notation,

$$C_{14} = C_{15} = C_{24} = C_{25} = C_{34} = C_{35} = C_{46} = C_{56} = 0 \quad , \tag{3.12}$$

leaving 13 independent elastic moduli for point groups C_2, C_s, and C_{2h}. The 6×6 matrix

C_{ab} thus takes the form

$$C_{ab}^{\text{MONO}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & C_{16} \\ C_{12} & C_{22} & C_{23} & 0 & 0 & C_{26} \\ C_{13} & C_{23} & C_{33} & 0 & 0 & C_{36} \\ 0 & 0 & 0 & C_{44} & C_{45} & 0 \\ 0 & 0 & 0 & C_{45} & C_{55} & 0 \\ C_{16} & C_{26} & C_{36} & 0 & 0 & C_{66} \end{pmatrix} \quad . \tag{3.13}$$

- For orthorhombic crystals, $x \to -x$ and $y \to -y$ are each symmetries. Adding $z \to -z$ in the case of D_{2h} doesn't buy us any new restrictions since C is symmetric under inversion. We then have $C_{ab} = 0$ whenever $a \in \{1, 2, 3\}$ and $b \in \{4, 5, 6\}$. The general form of C_{ab} is then

$$C_{ab}^{\text{ORTHO}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} \quad . \tag{3.14}$$

- For the tetragonal system, we can rotate (x, y, z) to $(-y, x, z)$. For the lower symmetry point groups among this system, namely C_4, S_4, and C_{4h}, the most general form is

$$C_{ab}^{\text{TET}}[C_4, S_4, C_{4h}] = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & C_{16} \\ C_{12} & C_{11} & C_{13} & 0 & 0 & -C_{16} \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ C_{16} & -C_{16} & 0 & 0 & 0 & C_{66} \end{pmatrix} \quad , \tag{3.15}$$

  which has seven independent moduli. For the higher symmetry tetragonal point groups D_4, C_{4v}, D_{2d}, and D_{4h}, we have $C_{16} = 0$ because of the twofold axes which send (x, y, z) into $(x, -y, -z)$ and $(-x, y, -z)$, and there are only six independent moduli.

- For the trigonal point groups, our lives are again complicated by the C_3 rotations. One convenient way to deal with this is to define $\xi \equiv x + iy$ and $\bar{\xi} \equiv x - iy$, with

$$\begin{aligned} \varepsilon_{\xi\xi} &= \xi_i \, \xi_j \, \varepsilon_{ij} = \varepsilon_{xx} - \varepsilon_{yy} + 2i \, \varepsilon_{xy} \\ \varepsilon_{\xi\bar{\xi}} &= \xi_i \, \bar{\xi}_j \, \varepsilon_{ij} = \varepsilon_{xx} + \varepsilon_{xy} \\ \varepsilon_{z\xi} &= \xi_i \, \varepsilon_{zi} = \varepsilon_{zx} + i \, \varepsilon_{zy} \\ \varepsilon_{z\bar{\xi}} &= \bar{\xi}_i \, \varepsilon_{zi} = \varepsilon_{zx} - i \, \varepsilon_{zy} \quad , \end{aligned} \tag{3.16}$$

  where $\xi_i = \partial_i \xi$ where $x_1 = x$ and $x_2 = y$, and $\bar{\xi}_i = \partial_i \bar{\xi}$. A C_3 rotation then takes $\xi \to e^{2\pi i/3} \, \xi$ and $\bar{\xi} \to e^{-2\pi i/3} \, \bar{\xi}$. The only allowed elements of C_{ijkl} are

$$C_{zzzz} \quad , \quad C_{zz\xi\bar{\xi}} \quad , \quad C_{\xi\xi\bar{\xi}\bar{\xi}} \quad , \quad C_{\xi\bar{\xi}\xi\bar{\xi}} \quad , \quad C_{z\xi z\bar{\xi}} \quad , \quad C_{z\xi\xi\xi} \quad , \quad C_{z\bar{\xi}\bar{\xi}\bar{\xi}} \quad , \tag{3.17}$$

and their corresponding elements obtained by permuting $C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}$. The first five of these are real, and the last two are complex conjugates: $C_{z\bar\xi\bar\xi\bar\xi} = C^*_{z\xi\xi\xi}$. So there are seven independent elastic moduli for the point groups C_3 and S_6. Note the general rule that we must have either no complex indices, one ξ and one $\bar\xi$ index, two each of ξ and $\bar\xi$, three ξ, or three $\bar\xi$. All other coefficients vanish by C_3 symmetry. We may now construct the elastic free energy density,

$$f = f_0 + \tfrac{1}{2}C_{zzzz}\,\varepsilon_{zz}^2 + C_{\xi\xi\bar\xi\bar\xi}\,\varepsilon_{\xi\xi}\,\varepsilon_{\bar\xi\bar\xi} + 2\,C_{\xi\bar\xi\xi\bar\xi}\,\varepsilon_{\xi\bar\xi}^2 + 2\,C_{zz\xi\bar\xi}\,\varepsilon_{zz}\,\varepsilon_{\xi\bar\xi} + 4\,C_{z\xi z\bar\xi}\,\varepsilon_{z\xi}\,\varepsilon_{z\bar\xi} + 2\,C_{z\xi\xi\xi}\,\varepsilon_{z\xi}\,\varepsilon_{\xi\xi} + 2\,C_{z\bar\xi\bar\xi\bar\xi}\,\varepsilon_{z\bar\xi}\,\varepsilon_{\bar\xi\bar\xi} \quad . \tag{3.18}$$

Note the coefficient of four in front of the $C_{z\xi z\bar\xi}$ term, which arises from summing over the eight equal contributions,

$$\tfrac{1}{2}\big(C_{z\xi z\bar\xi}+C_{z\xi\bar\xi z}+C_{\xi zz\bar\xi}+C_{\xi z\bar\xi z}+C_{z\bar\xi z\xi}+C_{z\bar\xi\xi z}+C_{\bar\xi zz\xi}+C_{\bar\xi z\xi z}\big)\,\varepsilon_{z\xi}\,\varepsilon_{z\bar\xi} = 4\,C_{z\xi z\bar\xi}\,\varepsilon_{z\xi}\,\varepsilon_{z\bar\xi} \quad . \tag{3.19}$$

From the free energy, one can identify the coefficients of $\varepsilon_a\,\varepsilon_b$, where a and b are composite indices, and thereby determine the general form for C_{ab}, which is

$$C_{ab}^{\text{TRIG}}[C_3, S_6] = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & -C_{25} & 0 \\ C_{12} & C_{11} & C_{13} & -C_{14} & C_{25} & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ C_{14} & -C_{14} & 0 & C_{44} & 0 & C_{25} \\ -C_{25} & C_{25} & 0 & 0 & C_{44} & C_{14} \\ 0 & 0 & 0 & C_{25} & C_{14} & \frac{1}{2}(C_{11}-C_{12}) \end{pmatrix} \quad , \tag{3.20}$$

Adding in reflections or twofold axes, as we have in the higher symmetry groups in this system, *i.e.* D_3, C_{3v}, and D_{3d} allows for $\xi \leftrightarrow \bar\xi$, in which case $C_{z\xi\xi\xi} = C_{z\bar\xi\bar\xi\bar\xi}$, reducing the number of independent moduli to six, with $C_{25} = 0$.

- For all seven hexagonal system point groups, we have $C_{z\xi\xi\xi} = C_{z\bar\xi\bar\xi\bar\xi} = 0$, because C_6 rotations take ξ to $\xi\,e^{i\pi/3}$, hence $C_{z\xi\xi\xi}$ to $-C_{z\xi\xi\xi}$. C_{3h} and D_{3h} don't contain this element, but do contain the mirror reflection $z \to -z$, hence in all cases the elastic tensor resembles that for the trigonal case, but with $C_{14} = C_{25} = 0$. Hence there are five independent moduli, with

$$C_{ab}^{\text{HEX}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(C_{11}-C_{12}) \end{pmatrix} \quad , \tag{3.21}$$

- For the cubic system (five point groups), the only independent elements are C_{xxxx}, C_{xxyy},

C_{xyxy}, and their symmetry-related counterparts such as C_{zzzz}, C_{yzyz}, *etc*. Thus,

$$C_{ab}^{\text{CUB}} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} , \tag{3.22}$$

- For an isotropic material, $C_{11} = C_{22} + 2C_{44}$. The Lamé parameters are $\lambda = C_{12}$ and $\mu = C_{44}$.

3.2 Phonons in Crystals

Crystalline solids support propagating waves called *phonons*, which are quantized vibrations of the lattice. Recall that the quantum mechanical Hamiltonian for a harmonic oscillator, $\hat{H} = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2\, q^2$, may be written as $\hat{H} = \hbar\omega_0\,(a^\dagger a + \frac{1}{2})$, where a and $a^\dagger$ are 'ladder operators' satisfying commutation relations $\left[a\,,a^\dagger\right] = 1$.

3.2.1 One-dimensional chain

Consider the linear chain of masses and springs depicted in fig. 3.1. We assume that our system consists of N mass points on a large ring of circumference L. In equilibrium, the masses are spaced evenly by a distance $b = L/N$. That is, $x_n^0 = nb$ is the equilibrium position of particle n. We define $u_n = x_n - x_n^0$ to be the difference between the position of mass n and The Hamiltonian is then

$$\begin{aligned} \hat{H} &= \sum_n \left[\frac{p_n^2}{2m} + \tfrac{1}{2}\kappa\,(x_{n+1} - x_n - a)^2 \right] \\ &= \sum_n \left[\frac{p_n^2}{2m} + \tfrac{1}{2}\kappa\,(u_{n+1} - u_n)^2 \right] + \tfrac{1}{2}N\kappa(b-a)^2\,, \end{aligned} \tag{3.23}$$

where a is the unstretched length of each spring, m is the mass of each mass point, κ is the force constant of each spring, and N is the total number of mass points. If $b \neq a$ the springs are under tension in equilibrium, but as we see this only leads to an additive constant in the Hamiltonian, and hence does not enter the equations of motion.

The classical equations of motion are

$$\dot{u}_n = \frac{\partial \hat{H}}{\partial p_n} = \frac{p_n}{m} \tag{3.24}$$

$$\dot{p}_n = -\frac{\partial \hat{H}}{\partial u_n} = \kappa \left(u_{n+1} + u_{n-1} - 2u_n \right) . \tag{3.25}$$

Taking the time derivative of the first equation and substituting into the second yields

$$\ddot{u}_n = \frac{\kappa}{m} \left(u_{n+1} + u_{n-1} - 2u_n \right) . \tag{3.26}$$

We now write

$$u_n = \frac{1}{\sqrt{N}} \sum_k \tilde{u}_k \, e^{ikna} , \tag{3.27}$$

where periodicity $u_{N+n} = u_n$ requires that the k values are quantized so that $e^{ikNa} = 1$, *i.e.* $k = 2\pi j/Na$ where $j \in \{0, 1, \ldots, N-1\}$. The inverse of this discrete Fourier transform is

$$\tilde{u}_k = \frac{1}{\sqrt{N}} \sum_n u_n \, e^{-ikna} . \tag{3.28}$$

Note that $\tilde{u}_k$ is in general complex, but that $\tilde{u}_k^* = \tilde{u}_{-k}$. In terms of the $\tilde{u}_k$, the equations of motion take the form

$$\ddot{\tilde{u}}_k = -\frac{2\kappa}{m} \left(1 - \cos(ka) \right) \tilde{u}_k \equiv -\omega_k^2 \, \tilde{u}_k . \tag{3.29}$$

Thus, each $\tilde{u}_k$ is a normal mode, and the normal mode frequencies are

$$\omega_k = 2 \sqrt{\frac{\kappa}{m}} \, \left| \sin\left(\tfrac{1}{2} ka \right) \right| . \tag{3.30}$$

The density of states for this band of phonon excitations is

$$\begin{aligned} g(\varepsilon) &= \int_{-\pi/a}^{\pi/a} \frac{dk}{2\pi} \, \delta(\varepsilon - \hbar\omega_k) \\ &= \frac{2}{\pi a} \big(J^2 - \varepsilon^2 \big)^{-1/2} \, \Theta(\varepsilon) \, \Theta(J - \varepsilon) , \end{aligned} \tag{3.31}$$

where $J = 2\hbar\sqrt{\kappa/m}$ is the phonon bandwidth. The step functions require $0 \leqslant \varepsilon \leqslant J$; outside this range there are no phonon energy levels and the density of states accordingly vanishes.

The entire theory can be quantized, taking $\big[p_n \, , u_{n'} \big] = -i\hbar\delta_{nn'}$. We then define

$$p_n = \frac{1}{\sqrt{N}} \sum_k \tilde{p}_k \, e^{ikna} \qquad , \qquad \tilde{p}_k = \frac{1}{\sqrt{N}} \sum_n p_n \, e^{-ikna} , \tag{3.32}$$

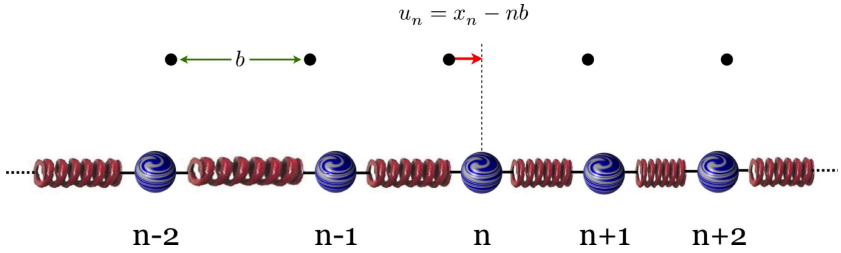


Figure 3.1: A linear chain of masses and springs. The black circles represent the equilibrium positions of the masses. The displacement of mass n relative to its equilibrium value is u_n.

in which case $\left[\tilde{p}_k\,,\tilde{u}_{k'}\right] = -i\hbar\delta_{kk'}$. Note that $\tilde{u}_k^\dagger = \tilde{u}_{-k}$ and $\tilde{p}_k^\dagger = \tilde{p}_{-k}$. We then define the ladder operator

$$a_k = \left(\frac{1}{2m\hbar\omega_k}\right)^{1/2}\tilde{p}_k - i\left(\frac{m\omega_k}{2\hbar}\right)^{1/2}\tilde{u}_k \tag{3.33}$$

and its Hermitean conjugate $a_k^\dagger$, in terms of which the Hamiltonian is

$$\hat{H} = \sum_k \hbar\omega_k\left(a_k^\dagger a_k + \tfrac{1}{2}\right)\,, \tag{3.34}$$

which is a sum over independent harmonic oscillator modes. Note that the sum over k is restricted to an interval of width 2π, *e.g.* $k \in \left[-\frac{\pi}{a},\frac{\pi}{a}\right]$, which is the *first Brillouin zone* for the one-dimensional chain structure. The state at wavevector $k + \frac{2\pi}{a}$ is identical to that at k, as we see from eqn. 3.28.

3.2.2 General theory of lattice vibrations

Consider next the vibrations of a general crystalline lattice in d space dimensions with an r component basis. We define $\boldsymbol{R}$ to be a Bravais lattice vector, *i.e.* a label for a unit cell, and $\boldsymbol{u}_i(\boldsymbol{R})$ to be the displacement of the $i^{\rm th}$ basis ion in the $\boldsymbol{R}$ unit cell. The Hamiltonian is

$$H = \sum_{\boldsymbol{R},i}\frac{\boldsymbol{p}_i^2(\boldsymbol{R})}{2m_i} + \frac{1}{2}\sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{i,j}\sum_{\alpha,\beta} u_i^\alpha(\boldsymbol{R})\,\Phi_{ij}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\,u_j^\beta(\boldsymbol{R}') + \mathcal{O}(u^3) \quad , \tag{3.35}$$

where

$$\Phi_{ij}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}') = \frac{\partial^2 U}{\partial u_i^\alpha(\boldsymbol{R})\,\partial u_j^\beta(\boldsymbol{R}')} \quad . \tag{3.36}$$

Remember that the indices i and j run over the set $\{1,\ldots,r\}$, where r is the number of basis vectors, while α and β are Cartesian vector indices taken from $\{1,2,\ldots,d\}$, where d is the dimension of space.

In the case of molecules, the dynamical matrix is of rank dN. For a molecule with no point group symmetries, this is the dimension of the eigenvalue problem to be solved. In crystals,

by contrast, we may take advantage of translational invariance to reduce the dimension of the eigenvalue problem to dr, *i.e.* to the number of degrees of freedom within a unit cell. This is so even in the case of a triclinic system with no symmetries (*i.e.* point group C_1). Each vibrational state is labeled by a wavevector $\boldsymbol{k}$, and at certain high symmetry points $\boldsymbol{k}$ in the Brillouin zone, crystallographic point group symmetries may be used to group these dr states into multiplets transforming according to point group IRREPs.

Upon Fourier transform,

$$\begin{aligned} u_i^\alpha(\boldsymbol{R}) &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \hat{u}_i^\alpha(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{\delta}_i} \\ p_i^\alpha(\boldsymbol{R}) &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \hat{p}_i^\alpha(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{\delta}_i} \quad , \end{aligned} \tag{3.37}$$

where the sum is over all $\boldsymbol{k}$ within the first Brillouin zone. The Fourier space dynamical matrix is then

$$\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} \Phi_{ij}^{\alpha\beta}(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{\delta}_i}\, e^{i\boldsymbol{k}\cdot\boldsymbol{\delta}_j} \quad . \tag{3.38}$$

The Hamiltonian, to quadratic order, takes the form

$$H = \sum_{\boldsymbol{k},i} \frac{\hat{p}_i^\alpha(\boldsymbol{k})\,\hat{p}_i^\alpha(-\boldsymbol{k})}{2m_i} + \frac{1}{2}\sum_{\boldsymbol{k}}\sum_{i,j}\sum_{\alpha,\beta} \hat{u}_i^\alpha(-\boldsymbol{k})\,\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})\,\hat{u}_j^\beta(\boldsymbol{k}) \quad , \tag{3.39}$$

Note that $\hat{u}_i^\alpha(-\boldsymbol{k}) = \big[\hat{u}_i^\alpha(\boldsymbol{k})\big]^*$ because the displacements $u_i^\alpha(\boldsymbol{R})$ are real; a corresponding relation holds for the momenta. Note also the Poisson bracket relation in crystal momentum space becomes

$$\Big\{u_i^\alpha(\boldsymbol{R})\,,\, p_j^\beta(\boldsymbol{R}')\Big\}_{\text{PB}} = \delta_{\boldsymbol{R}\boldsymbol{R}'}\,\delta_{ij}\,\delta_{\alpha\beta} \quad \Rightarrow \quad \Big\{\hat{u}_i^\alpha(\boldsymbol{k})\,,\,\hat{p}_j^\beta(\boldsymbol{k}')\Big\}_{\text{PB}} = \delta^{\text{P}}_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{0}}\,\delta_{ij}\,\delta_{\alpha\beta} \quad , \tag{3.40}$$

where $\delta^{\text{P}}_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{0}} = \sum_{\boldsymbol{G}} \delta_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{G}}$ requires $\boldsymbol{k}+\boldsymbol{k}' = \boldsymbol{0}$ modulo any reciprocal lattice vector. Note also that

$$\Phi_{ij}^{\alpha\beta}(\boldsymbol{R}) = \Phi_{ji}^{\beta\alpha}(-\boldsymbol{R}) \quad \Rightarrow \quad \hat{\Phi}_{ji}^{\beta\alpha}(\boldsymbol{k}) = \hat{\Phi}_{ij}^{\alpha\beta}(-\boldsymbol{k}) = \big[\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})\big]^* \quad . \tag{3.41}$$

Thus, for each crystal momentum $\boldsymbol{k}$, the dynamical matrix $\hat{\Phi}_{ji}^{\beta\alpha}(\boldsymbol{k})$ is Hermitian, where we take $(i\alpha)$ and $(j\beta)$ as composite indices. We now have the eigensystem

$$\sum_{\beta,j} \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})\,\hat{\mathrm{e}}_{j\lambda}^\beta(\boldsymbol{k}) = m_i\,\omega_\lambda^2(\boldsymbol{k})\,\hat{\mathrm{e}}_{i\lambda}^\alpha(\boldsymbol{k}) \tag{3.42}$$

where $\lambda \in \{1,\ldots,rd\}$ indexes the normal modes, and $S_{i\alpha,\lambda}(\boldsymbol{k}) \equiv \hat{\mathrm{e}}_{i\lambda}^\alpha(\boldsymbol{k}) \equiv m_i^{-1/2}\,U_{i\alpha,\lambda}(\boldsymbol{k})$ diagonalizes the dynamical matrix, with $U_{i\alpha,\lambda}(\boldsymbol{k})$ unitary. We may now write the completeness relation,

$$\sum_{\lambda=1}^{dr} \hat{\mathrm{e}}_{i\lambda}^{\alpha *}(\boldsymbol{k})\,\hat{\mathrm{e}}_{j\lambda}^\beta(\boldsymbol{k}) = \frac{1}{m_i}\,\delta_{ij}\,\delta_{\alpha\beta} \tag{3.43}$$

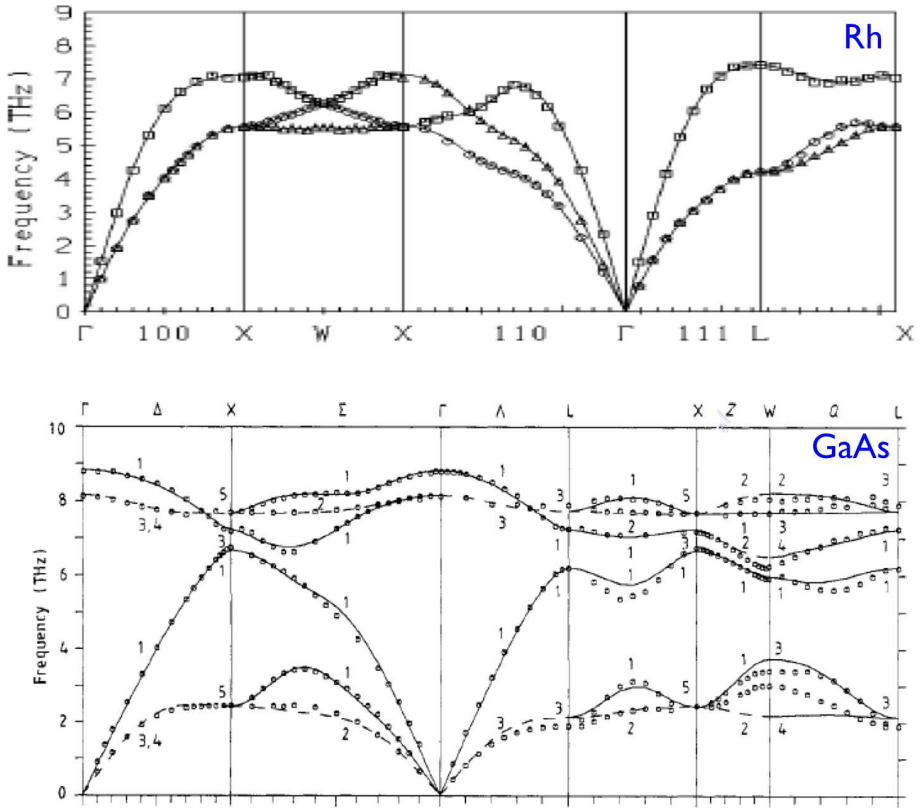


Figure 3.2: Upper panel: phonon spectrum in fcc elemental rhodium (Rh) at $T = 297\,\mathrm{K}$ measured by high precision inelastic neutron scattering (INS) by A. Eichler *et al.*, *Phys. Rev. B* **57**, 324 (1998). Note the three acoustic branches and no optical branches, corresponding to $d = 3$ and $r = 1$. Lower panel: phonon spectrum in gallium arsenide (GaAs) at $T = 12\,\mathrm{K}$, comparing theoretical lattice-dynamical calculations with INS results of D. Strauch and B. Dorner, *J. Phys.: Condens. Matter* **2**, 1457 (1990). Note the three acoustic branches and three optical branches, corresponding to $d = 3$ and $r = 2$. The Greek letters along the x-axis indicate points of high symmetry in the Brillouin zone.

and the orthogonality relation,

$$\sum_{i=1}^{r} \sum_{\alpha=1}^{d} m_i \, \hat{\mathrm{e}}^{\alpha *}_{i\lambda}(\boldsymbol{k}) \, \hat{\mathrm{e}}^{\alpha}_{i\lambda'}(\boldsymbol{k}) = \delta_{\lambda\lambda'} \quad , \tag{3.44}$$

which are the completeness and orthogonality relations, respectively. Since $\hat{\mathrm{e}}^{\alpha *}_{i\lambda}(-\boldsymbol{k})$ and $\hat{\mathrm{e}}^{\alpha}_{i\lambda}(\boldsymbol{k})$ obey the same equation, we have that $\omega_\lambda(-\boldsymbol{k}) = \omega_\lambda(\boldsymbol{k})$. If the phonon eigenmode $|\, \boldsymbol{k}, \lambda \,\rangle$ is nondegenerate, we may choose $\hat{\mathrm{e}}^{\alpha}_{i\lambda}(-\boldsymbol{k}) = \hat{\mathrm{e}}^{\alpha *}_{i\lambda}(\boldsymbol{k})$. Else at best we can conclude $\hat{\mathrm{e}}^{\alpha}_{i\lambda}(-\boldsymbol{k}) = \hat{\mathrm{e}}^{\alpha *}_{i\lambda'}(\boldsymbol{k})\, e^{i\eta}$ where $|\, \boldsymbol{k}, \lambda' \,\rangle$ is another state from the degenerate manifold of phonon states at this wavevector, and $e^{i\eta}$ is a phase.

Expressing $\hat{u}_i^\alpha(\boldsymbol{k})$ and $\hat{p}_i^\alpha(\boldsymbol{k})$ in terms of the normal modes, we write

$$\hat{u}_i^\alpha(\boldsymbol{k}) = \sum_{\lambda=1}^{dr} \hat{\mathrm{e}}_{i\lambda}^\alpha(\boldsymbol{k})\,\hat{q}_\lambda(\boldsymbol{k}) \qquad , \qquad \hat{p}_i^\alpha(\boldsymbol{k}) = m_i \sum_{\lambda=1}^{dr} \hat{\mathrm{e}}_{i\lambda}^{\alpha *}(-\boldsymbol{k})\,\hat{\pi}_\lambda(\boldsymbol{k}) \tag{3.45}$$

where $\big\{\hat{q}_\lambda(\boldsymbol{k}), \hat{\pi}_{\lambda'}(\boldsymbol{k})\big\}_{\text{PB}} = \delta_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{0}}\,\delta_{\lambda\lambda'}$. This entails

$$\hat{q}_\lambda(\boldsymbol{k}) = \sum_{i=1}^{r}\sum_{\alpha=1}^{d} m_i\,\hat{\mathrm{e}}_{i\lambda}^{\alpha *}(\boldsymbol{k})\,\hat{u}_i^\alpha(\boldsymbol{k}) \qquad , \qquad \hat{\pi}_\lambda(\boldsymbol{k}) = \sum_{i=1}^{r}\sum_{\alpha=1}^{d} \hat{\mathrm{e}}_{i\lambda}^{\alpha}(-\boldsymbol{k})\,\hat{p}_i^\alpha(\boldsymbol{k}) \quad . \tag{3.46}$$

The phonon Hamiltonian now takes the diagonalized form

$$H = \sum_{\boldsymbol{k}}\sum_{\lambda=1}^{rd} \Big\{ \tfrac{1}{2}\,\hat{\pi}_\lambda(-\boldsymbol{k})\,\hat{\pi}_\lambda(\boldsymbol{k}) + \tfrac{1}{2}\,\omega_\lambda^2(\boldsymbol{k})\,\hat{q}_\lambda(-\boldsymbol{k})\,\hat{q}_\lambda(\boldsymbol{k}) \Big\} \quad , \tag{3.47}$$

with $\big\{\hat{q}_\lambda(\boldsymbol{k})\,,\,\hat{\pi}_{\lambda'}(\boldsymbol{k}')\big\}_{\text{PB}} = \delta_{\lambda\lambda'}\,\delta^{\text{P}}_{\boldsymbol{k}+\boldsymbol{k}',\boldsymbol{0}}$. To quantize, promote the Poisson brackets to commutators: $\big\{A\,,\,B\big\}_{\text{PB}} \to -i\hbar^{-1}\,\big[A\,,\,B\big]$. Then define the ladder operators,

$$A_\lambda(\boldsymbol{k}) = \left(\frac{\omega_\lambda(\boldsymbol{k})}{2\hbar}\right)^{1/2} \hat{q}_\lambda(\boldsymbol{k}) + i\left(\frac{1}{2\hbar\omega_\lambda(\boldsymbol{k})}\right)^{1/2} \hat{\pi}_\lambda(\boldsymbol{k}) \quad , \tag{3.48}$$

which satisfy $\big[A_\lambda(\boldsymbol{k})\,,\,A^\dagger_{\lambda'}(\boldsymbol{k}')\big] = \delta^{\text{P}}_{\boldsymbol{k}\boldsymbol{k}'}\,\delta_{\lambda\lambda'}$. The quantum phonon Hamiltonian is then

$$\hat{H} = \sum_{\boldsymbol{k}}\sum_{\lambda=1}^{rd} \hbar\omega_\lambda(\boldsymbol{k})\left(A^\dagger_\lambda(\boldsymbol{k})\,A_\lambda(\boldsymbol{k}) + \tfrac{1}{2}\right) \quad . \tag{3.49}$$

Of the dr phonon branches, d are *acoustic*, and behave as $\omega_a(\boldsymbol{k}) = c(\hat{\boldsymbol{k}})\,k$ as $\boldsymbol{k} \to 0$, which is the Γ point in the Brillouin zone. These gapless phonons are the Goldstone bosons of the spontaneously broken translational symmetry which gave rise to the crystalline phase. To each broken generator of translation, there corresponds a Goldstone mode. The remaining $d(r-1)$ modes are called *optical phonons*. Whereas for acoustic modes, all the ions in a given unit cell are moving in phase, for optical modes they are moving out of phase. Hence optical modes are always finite frequency modes. Fig. 3.2 shows the phonon spectra in elemental rhodium (space group $Fm\overline{3}m$, point group O_h), and in gallium arsenide (space group $F\overline{4}3m$, point group T_d). Since Rh forms an fcc Bravais lattice, there are no optical phonon modes. GaAs forms a zincblende structure, *i.e.* two interpenetrating fcc lattices, one for the gallium, the other for the arsenic. Thus $r = 2$ and we expect three acoustic and three optical branches of phonons.

Nota bene : One may choose to define the Fourier transforms above taking the additional phases for the basis elements to all be unity, *viz.*

$$u_i^\alpha(\boldsymbol{R}) = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \hat{u}_i^\alpha(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad p_i^\alpha(\boldsymbol{R}) = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \hat{p}_i^\alpha(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} \Phi_{ij}^{\alpha\beta}(\boldsymbol{R})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{3.50}$$

All the equations starting with Eqn. 3.39 remain the same. Setting the basis phases to unity amounts to a choice of *gauge*. It is somewhat simpler in certain contexts, but it may obscure essential space group symmetries. On the other hand, it should also be noted that the Fourier transforms $\hat{u}_i^\alpha(\boldsymbol{k})$, $\hat{p}_i^\alpha(\boldsymbol{k})$, and $\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})$ are not periodic in the Brillouin zone, but instead satisfy generalized periodic boundary conditions,

$$\begin{aligned}\hat{u}_i^\alpha(\boldsymbol{K}+\boldsymbol{k}) &= e^{-i\boldsymbol{K}\cdot\boldsymbol{\delta}_i}\,\hat{u}_i^\alpha(\boldsymbol{k}) \\ \hat{p}_i^\alpha(\boldsymbol{K}+\boldsymbol{k}) &= e^{-i\boldsymbol{K}\cdot\boldsymbol{\delta}_i}\,\hat{p}_i^\alpha(\boldsymbol{k}) \\ \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{K}+\boldsymbol{k}) &= e^{-i\boldsymbol{K}\cdot(\boldsymbol{\delta}_i-\boldsymbol{\delta}_j)}\,\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k}) \quad ,\end{aligned} \tag{3.51}$$

where $\boldsymbol{K} \in \hat{\mathcal{L}}$ is any reciprocal lattice vector.

3.2.3 Translation and rotation invariance

The potential energy $U\big(\{u_i^\alpha(\boldsymbol{R})\}\big)$ must remain invariant under the operations

$$\begin{aligned}u_i^\alpha(\boldsymbol{R}) &\to u_i^\alpha(\boldsymbol{R}) + d^\alpha \\ u_i^\alpha(\boldsymbol{R}) &\to u_i^\alpha(\boldsymbol{R}) + \epsilon_{\alpha\mu\nu}\,(R^\mu + \delta_i^\mu - \delta_j^\mu)\,d^\nu\end{aligned} \tag{3.52}$$

for an infinitesimal vector $\boldsymbol{d}$. The first equation represents a uniform translation of all lattice sites by $\boldsymbol{d}$. The second represents an infinitesimal rotation about the $j^{\rm th}$ basis ion in the $\boldsymbol{R}=0$ unit cell. We are free to choose any j.

Writing $U(u+\Delta u) = U(u)$, we must have that the linear terms in Δu vanish, hence

$$\begin{aligned}\sum_{\boldsymbol{R},i} \Phi_{ij}^{\alpha\beta}(\boldsymbol{R}) &= \sum_i \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{0}) = 0 \\ \epsilon_{\alpha\mu\nu}\sum_{\boldsymbol{R},i}(R^\mu + \delta_i^\mu - \delta_j^\mu)\,\Phi_{ij}^{\nu\beta}(\boldsymbol{R}) &= i\,\epsilon_{\alpha\mu\nu}\sum_i \frac{\partial\hat{\Phi}_{ij}^{\nu\beta}(\boldsymbol{k})}{\partial k^\mu}\bigg|_{\boldsymbol{k}=0} = 0 \quad .\end{aligned} \tag{3.53}$$

Note that (α,β,j) are free indices in both equations. The first of these equations says that any vector d^β is an eigenvector of the dynamical matrix at $\boldsymbol{k}=0$, with zero eigenvalue. Thus, at $\boldsymbol{k}=0$, there is a three-dimensional space of zero energy modes. These are the Goldstone modes associated with the three broken generators of translation in the crystal.

3.2.4 Phonons in an fcc lattice

When the crystal is a Bravais lattice, there are no basis indices, and the dynamical matrix becomes

$$\hat{\Phi}^{\alpha\beta}(\boldsymbol{k}) = {\sum_{\boldsymbol{R}}}'(1-\cos\boldsymbol{k}\cdot\boldsymbol{R})\,\frac{\partial^2 v(\boldsymbol{R})}{\partial R^\alpha\,\partial R^\beta} \quad , \tag{3.54}$$

where $v(\boldsymbol{r})$ is the inter-ionic potential, and the prime on the sum indicates that $\boldsymbol{R} = 0$ is to be excluded. For central potentials $v(\boldsymbol{R}) = v(R)$,

$$\frac{\partial^2 v(\boldsymbol{R})}{\partial R^\alpha \, \partial R^\beta} = \left(\delta^{\alpha\beta} - \hat{R}^\alpha \, \hat{R}^\beta\right) \frac{v'(R)}{R} + \hat{R}^\alpha \, \hat{R}^\beta \, v''(R) \quad . \tag{3.55}$$

For simplicity, we assume $v(R)$ is negligible beyond the first neighbor. On the fcc lattice, there are twelve first neighbors, lying at $\boldsymbol{\Delta} = \frac{1}{2}a\,(\pm\hat{\boldsymbol{y}} \pm \hat{\boldsymbol{z}})$, $\boldsymbol{\Delta} = \frac{1}{2}a\,(\pm\hat{\boldsymbol{x}} \pm \hat{\boldsymbol{z}})$, and $\boldsymbol{\Delta} = \frac{1}{2}a\,(\pm\hat{\boldsymbol{x}} \pm \hat{\boldsymbol{y}})$. Here a is the side length of the underlying simple cubic lattice, so the fcc lattice constant is $a/\sqrt{2}$. We define

$$A = \frac{\sqrt{2}}{a}\, v'\big(a/\sqrt{2}\big) \quad , \quad B = v''\big(a/\sqrt{2}\big) \quad . \tag{3.56}$$

Along (100), we have $\boldsymbol{k} = k\hat{\boldsymbol{x}}$ and

$$\hat{\Phi}^{\alpha\beta}(\boldsymbol{k}) = 4\sin^2(\tfrac{1}{4}ka) \begin{pmatrix} 2A+2B & 0 & 0 \\ 0 & 3A+B & 0 \\ 0 & 0 & 3A+B \end{pmatrix} \quad , \tag{3.57}$$

which is already diagonal. Thus, the eigenvectors lie along the cubic axes and

$$\omega_{\text{L}} = 2\sqrt{\tfrac{2(A+B)}{m}}\,\big|\sin(ka/4)\big| \qquad , \qquad \omega_{\text{T1}} = \omega_{\text{T2}} = 2\sqrt{\tfrac{3A+B}{m}}\,\big|\sin(ka/4)\big| \quad . \tag{3.58}$$

Along (111), we have $\boldsymbol{k} = \frac{1}{\sqrt{3}}\,k\,(\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}} + \hat{\boldsymbol{z}})$. One finds

$$\hat{\Phi}^{\alpha\beta}(\boldsymbol{k}) = 4\sin^2\big(ka/\sqrt{12}\big) \begin{pmatrix} 4A+2B & B-A & B-A \\ B-A & 4A+2B & B-A \\ B-A & B-A & 4A+2B \end{pmatrix} \quad . \tag{3.59}$$

$$\omega_{\text{L}} = 2\sqrt{\tfrac{A+2B}{m}}\,\big|\sin\big(ka/\sqrt{12}\big)\big| \qquad , \qquad \omega_{\text{T1}} = \omega_{\text{T2}} = 2\sqrt{\tfrac{5A+B}{2m}}\,\big|\sin\big(ka/\sqrt{12}\big)\big| \quad . \tag{3.60}$$

3.2.5 Phonons in the hcp structure

The HCP structure is represented as an underlying simple hexagonal lattice with a two-element basis:

$$\boldsymbol{a}_1 = a\,\hat{\boldsymbol{x}} \quad , \quad \boldsymbol{a}_2 = \tfrac{1}{2}a\,\hat{\boldsymbol{x}} + \tfrac{\sqrt{3}}{2}a\,\hat{\boldsymbol{y}} \quad , \quad \boldsymbol{a}_3 = \sqrt{\tfrac{8}{3}}\,a\,\hat{\boldsymbol{z}}\ . \tag{3.61}$$

Bravais lattice sites are of the form $\boldsymbol{R} = l\boldsymbol{a}_1 + m\boldsymbol{a}_2 + n\boldsymbol{a}_3$. The A sublattice occupies the sites $\{\boldsymbol{R}\}$, while the B sublattice occupies the sites $\{\boldsymbol{R} + \boldsymbol{\delta}\}$, where

$$\boldsymbol{\delta} = \tfrac{1}{2}a\,\hat{\boldsymbol{x}} + \tfrac{1}{2\sqrt{3}}\,a\,\hat{\boldsymbol{y}} + \sqrt{\tfrac{2}{3}}\,a\,\hat{\boldsymbol{z}}\ . \tag{3.62}$$

The nearest neighbor separation is $|\boldsymbol{a}_1| = |\boldsymbol{a}_2| = |\boldsymbol{\delta}| = a$. Note that $\boldsymbol{R}$ can be used to label the unit cells, *i.e.* each unit cell is labeled by the coordinates of its constituent A sublattice site.

Classical energy

The classical energy for the system is the potential energy of the fixed lattice, given by

$$\frac{U_0}{N} = \sum_{\boldsymbol{R}} \Big[v(\boldsymbol{R}) \left(1 - \delta_{\boldsymbol{R},0}\right) + v(\boldsymbol{R} + \boldsymbol{\delta}) \Big] \,, \tag{3.63}$$

where $v(\boldsymbol{r})$ is the interatomic potential.

Dynamical matrix

When phonon fluctuations are included, the positions of the A and B sublattice sites are written

$$\begin{aligned} \boldsymbol{R} &\longrightarrow \boldsymbol{R} + \boldsymbol{u}_{\mathsf{A}}(\boldsymbol{R}) \\ \boldsymbol{R} + \boldsymbol{\delta} &\longrightarrow \boldsymbol{R} + \boldsymbol{\delta} + \boldsymbol{u}_{\mathsf{B}}(\boldsymbol{R}) \,. \end{aligned} \tag{3.64}$$

Then the potential energy is

$$\begin{aligned} U = U_0 &+ \sum_{\boldsymbol{R}} \Big(\boldsymbol{u}_{\mathsf{A}}(\boldsymbol{R}) \cdot \boldsymbol{F}_{\mathsf{A}}(\boldsymbol{R}) + \boldsymbol{u}_{\mathsf{B}}(\boldsymbol{R}) \cdot \boldsymbol{F}_{\mathsf{B}}(\boldsymbol{R}) \Big) \\ &+ \tfrac{1}{2} \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{j,j'} \sum_{\alpha,\alpha'} \Phi_{jj'}^{\alpha\alpha'}(\boldsymbol{R} - \boldsymbol{R}') \, u_j^{\alpha}(\boldsymbol{R}) \, u_{j'}^{\alpha'}(\boldsymbol{R}') + \mathcal{O}(u^3) \,, \end{aligned} \tag{3.65}$$

where

$$\Phi_{jj'}^{\alpha\alpha'}(\boldsymbol{R} - \boldsymbol{R}') = \frac{\partial^2 U}{\partial u_j^{\alpha}(\boldsymbol{R}) \, \partial u_{j'}^{\alpha'}(\boldsymbol{R}')} \,. \tag{3.66}$$

Here $\{\alpha, \alpha'\}$ are spatial indices (x, y, z), and $\{j, j'\}$ are sublattice indices (A, B).

It is convenient to Fourier transform, with

$$\begin{aligned} u_{\mathsf{A}}^{\alpha}(\boldsymbol{R}) &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \hat{u}_{\mathsf{A}}^{\alpha}(\boldsymbol{k}) \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \\ u_{\mathsf{B}}^{\alpha}(\boldsymbol{R}) &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \hat{u}_{\mathsf{B}}^{\alpha}(\boldsymbol{k}) \, e^{i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\delta})} \,, \end{aligned} \tag{3.67}$$

where N is the total number of unit cells. Then

$$U = U_0 + \sum_{\boldsymbol{k}} \sum_{j} \hat{\boldsymbol{u}}_j(\boldsymbol{k}) \cdot \hat{\boldsymbol{F}}_j(-\boldsymbol{k}) + \tfrac{1}{2} \sum_{\boldsymbol{k}} \sum_{j,j'} \sum_{\alpha,\alpha'} \hat{\Phi}_{jj'}^{\alpha\alpha'}(\boldsymbol{k}) \, \hat{u}_j^{\alpha}(\boldsymbol{k}) \, \hat{u}_{j'}^{\alpha'}(-\boldsymbol{k}) + \mathcal{O}(u^3) \,, \tag{3.68}$$

where the dynamical matrix is

$$\hat{\Phi}_{jj'}^{\alpha\alpha'}(\boldsymbol{k}) = \begin{pmatrix} \hat{\Phi}_{11}^{\alpha\alpha'}(\boldsymbol{k}) & \hat{\Phi}_{12}^{\alpha\alpha'}(\boldsymbol{k}) \\ \\ \hat{\Phi}_{21}^{\alpha\alpha'}(\boldsymbol{k}) & \hat{\Phi}_{22}^{\alpha\alpha'}(\boldsymbol{k}) \end{pmatrix} \,. \tag{3.69}$$

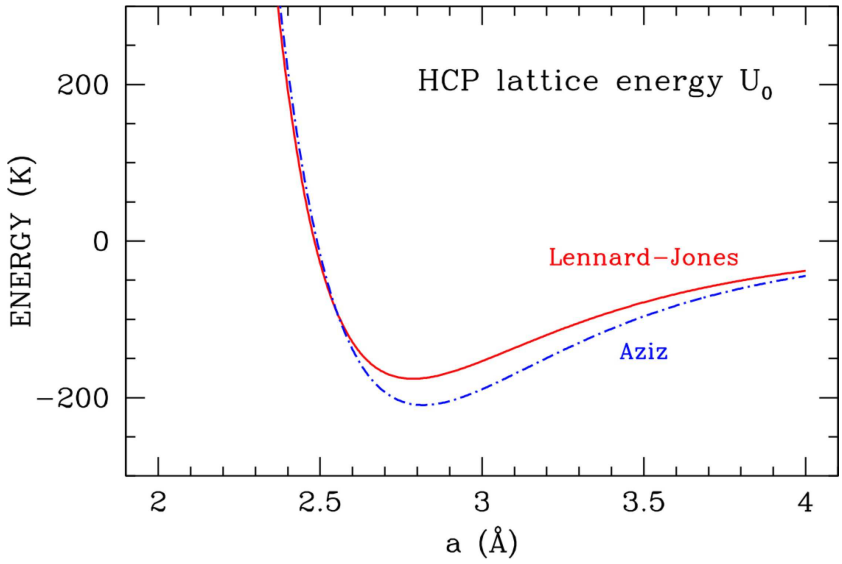


Figure 3.3: Classical lattice energy for hcp ^{4}He as a function of nearest neighbor separation a for the Lennard-Jones potential (red) and the Aziz potential (blue).

where

$$\begin{aligned}\hat\Phi^{\alpha\beta}_{11}(\boldsymbol{k}) &= {\sum_{\boldsymbol{R}}}'(1-\cos\boldsymbol{k}\cdot\boldsymbol{R})\,\frac{\partial^2 v(\boldsymbol{R})}{\partial R^\alpha\,\partial R^\beta} + \sum_{\boldsymbol{R}}\partial_\alpha\partial_\beta\,v(\boldsymbol{R}+\boldsymbol{\delta}) \\ \hat\Phi^{\alpha\beta}_{12}(\boldsymbol{k}) &= -\sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\delta})}\,\frac{\partial^2 v(\boldsymbol{R}+\boldsymbol{\delta})}{\partial R^\alpha\,\partial R^\beta}\end{aligned} \tag{3.70}$$

Note that $\hat\Phi^{\alpha\beta}_{21}(\boldsymbol{k}) = \big[\hat\Phi^{\alpha\beta}_{12}(\boldsymbol{k})\big]^*$. Note also that if $v(\boldsymbol{R}) = v(R)$ is a central potential, then

$$\frac{\partial^2 v(R)}{\partial R^\alpha\,\partial R^\beta} = \big(\delta^{\alpha\beta} - \hat R^\alpha\hat R^\beta\big)\,\frac{v'(R)}{R} + \hat R^\alpha\hat R^\beta\,v''(R)\;, \tag{3.71}$$

where $\hat R^\alpha = R^\alpha/|\boldsymbol{R}|$.

Lennard-Jones potential

The Lennard-Jones potential is given by

$$v(r) = 4\varepsilon_0\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{3.72}$$

where

$$\varepsilon_0 = 10.22\,\mathrm{K} \qquad , \qquad \sigma = 2.556\,\text{Å}\;. \tag{3.73}$$

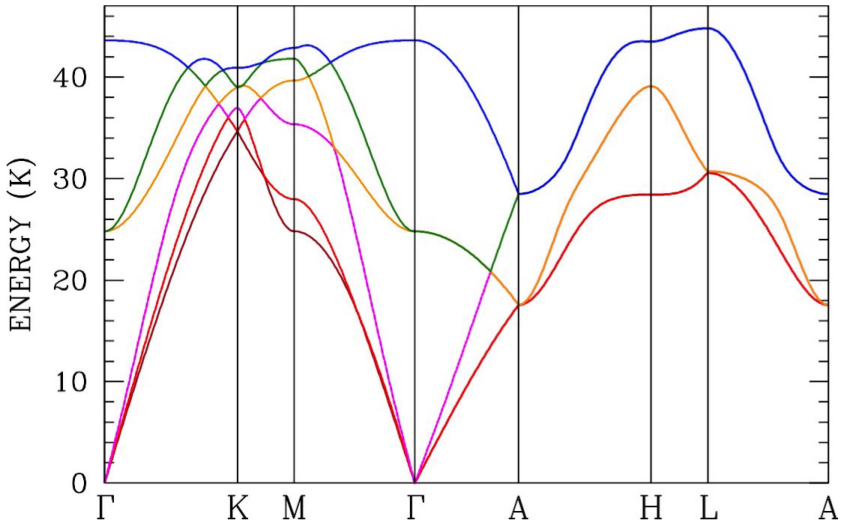


Figure 3.4: Phonon dispersions along high-symmetry directions in the Brillouin zone for hcp ^{4}He at molar volume $v_0 = 12\,\mathrm{cm}^3/\mathrm{mol}$, using the Lennard-Jones potential.

Aziz potential

The Aziz potential is given by

$$v(r) = \varepsilon_0 \left\{ A\, e^{-\alpha r/b} - \left[C_6 \left(\frac{b}{r}\right)^6 + C_8 \left(\frac{b}{r}\right)^8 + C_{10} \left(\frac{b}{r}\right)^{10} \right] F(r) \right\} , \tag{3.74}$$

where

$$F(r) = \begin{cases} e^{-\left(\frac{Db}{r}-1\right)^2} & \text{if } r \leqslant Db \\ 1 & \text{if } r > Db\,, \end{cases} \tag{3.75}$$

with

$$\varepsilon = 10.8\,\mathrm{K} \quad , \quad b = 2.9763\,\text{Å} \quad , \quad A = 5.448504 \times 10^5 \quad , \quad \alpha = 13.353384 \tag{3.76}$$

and

$$C_6 = 1.37732412 \quad , \quad C_8 = 0.4253785 \quad , \quad C_{10} = 0.171800 \quad , \quad D = 1.231314\,. \tag{3.77}$$

The mass of the helium-4 atom is $m = 6.65 \times 10^{-24}\,\mathrm{g}$.

3.2.6 Phonon density of states

For a crystalline lattice with an r-element basis, there are then $d \cdot r$ phonon modes for each wavevector $\boldsymbol{k}$ lying in the first Brillouin zone. If we impose periodic boundary conditions, then the $\boldsymbol{k}$ points within the first Brillouin zone are themselves quantized, as in the $d = 1$ case where

we found $k = 2\pi n/N$. There are N distinct $\boldsymbol{k}$ points in the first Brillouin zone – one for every direct lattice site. The total number of modes is than $d \cdot r \cdot N$, which is the total number of translational degrees of freedom in our system: rN total atoms (N unit cells each with an r atom basis) each free to vibrate in d dimensions. Of the $d \cdot r$ branches of phonon excitations, d of them will be *acoustic modes* whose frequency vanishes as $\boldsymbol{k} \to 0$. The remaining $d(r-1)$ branches are *optical modes* and oscillate at finite frequencies. Basically, in an acoustic mode, for $\boldsymbol{k}$ close to the (Brillouin) zone center $\boldsymbol{k} = 0$, all the atoms in each unit cell move together in the same direction at any moment of time. In an optical mode, the different basis atoms move in different directions.

There is no number conservation law for phonons – they may be freely created or destroyed in anharmonic processes, where two photons with wavevectors $\boldsymbol{k}$ and $\boldsymbol{q}$ can combine into a single phonon with wavevector $\boldsymbol{k} + \boldsymbol{q}$, and *vice versa*. Therefore the chemical potential for phonons is $\mu = 0$. We define the density of states $g_s(\omega)$ for the $s^{\rm th}$ phonon mode as

$$g_s(\omega) = \frac{1}{N}\sum_{\boldsymbol{k}} \delta\big(\omega - \omega_s(\boldsymbol{k})\big) = \Omega \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d}\, \delta\big(\omega - \omega_s(\boldsymbol{k})\big)\ , \tag{3.78}$$

where N is the number of unit cells, Ω is the unit cell volume of the direct lattice, and the $\boldsymbol{k}$ sum and integral are over the first Brillouin zone only. Note that ω here has dimensions of frequency. The functions $g_a(\omega)$ is normalized to unity:

$$\int\limits_0^\infty d\omega\, g_s(\omega) = 1\ . \tag{3.79}$$

The total phonon density of states per unit cell is given by[5] $g(\omega) = \sum_{s=1}^{dr} g_s(\omega)$.

The grand potential for the phonon gas is

$$\begin{aligned}\Omega(T,V) &= -k_{\rm B}T \ln \prod_{\boldsymbol{k},s} \sum_{n_a(\boldsymbol{k})=0}^{\infty} e^{-\beta\hbar\omega_s(\boldsymbol{k})\left(n_s(\boldsymbol{k})+\frac{1}{2}\right)} \\ &= k_{\rm B}T \sum_{\boldsymbol{k},s} \ln\left[2\sinh\left(\frac{\hbar\omega_s(\boldsymbol{k})}{2k_{\rm B}T}\right)\right] = Nk_{\rm B}T \int\limits_0^\infty d\omega\, g(\omega)\, \ln\left[2\sinh\left(\frac{\hbar\omega}{2k_{\rm B}T}\right)\right]\ .\end{aligned} \tag{3.80}$$

Note that $V = N\mathcal{V}_0$ since there are N unit cells, each of volume $\mathcal{V}_0$. The entropy is given by $S = -\left(\frac{\partial\Omega}{\partial T}\right)_V$ and thus the heat capacity is

$$C_V = -T\,\frac{\partial^2\Omega}{\partial T^2} = Nk_{\rm B} \int\limits_0^\infty d\omega\, g(\omega) \left(\frac{\hbar\omega}{2k_{\rm B}T}\right)^2 {\rm csch}^2\left(\frac{\hbar\omega}{2k_{\rm B}T}\right) \tag{3.81}$$

[5]Note the dimensions of $g(\omega)$ are $(\text{frequency})^{-1}$. By contrast, the dimensions of $g(\varepsilon)$ are $(\text{energy})^{-1} \cdot (\text{volume})^{-1}$. The difference lies in the a factor of $\mathcal{V}_0 \cdot \hbar$, where $\mathcal{V}_0$ is the unit cell volume.

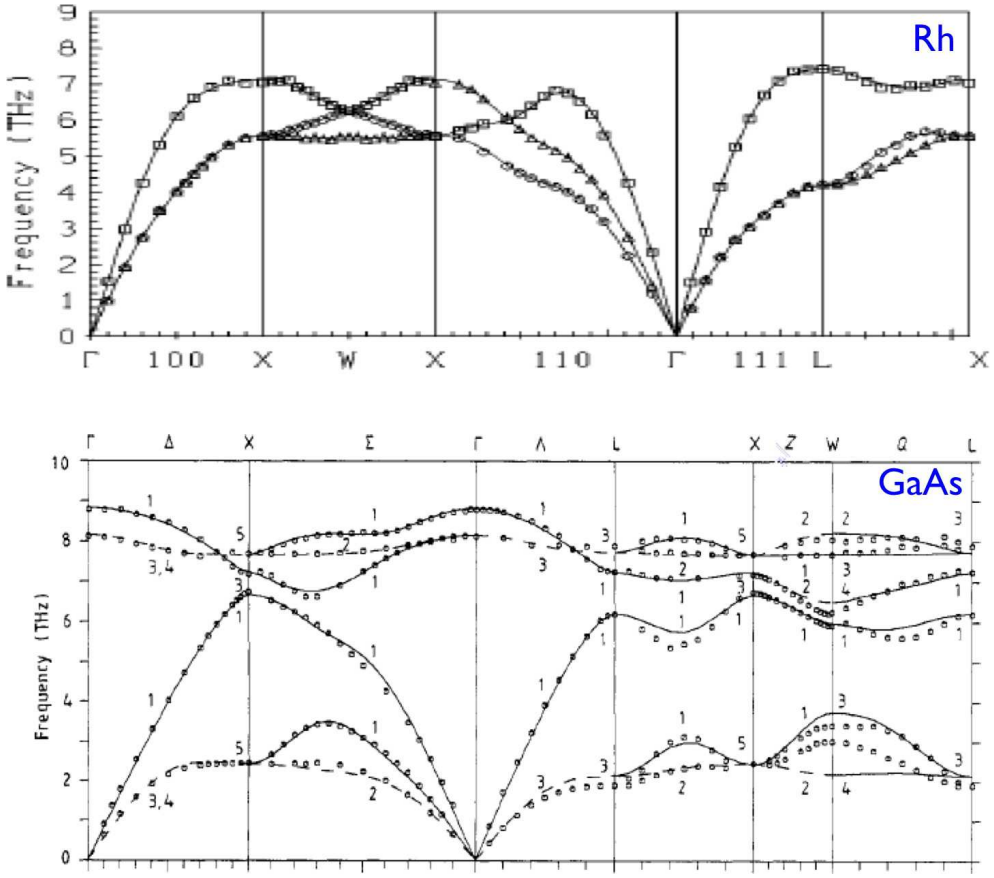


Figure 3.5: Upper panel: phonon spectrum in elemental rhodium (Rh) at $T = 297\,\mathrm{K}$ measured by high precision inelastic neutron scattering (INS) by A. Eichler *et al.*, *Phys. Rev. B* **57**, 324 (1998). Note the three acoustic branches and no optical branches, corresponding to $d = 3$ and $r = 1$. Lower panel: phonon spectrum in gallium arsenide (GaAs) at $T = 12\,\mathrm{K}$, comparing theoretical lattice-dynamical calculations with INS results of D. Strauch and B. Dorner, *J. Phys.: Condens. Matter* **2**, 1457 (1990). Note the three acoustic branches and three optical branches, corresponding to $d = 3$ and $r = 2$. The Greek letters along the x-axis indicate points of high symmetry in the Brillouin zone.

Note that as $T \to \infty$ we have $\operatorname{csch}\left(\frac{\hbar\omega}{2k_{\mathrm{B}}T}\right) \to \frac{2k_{\mathrm{B}}T}{\hbar\omega}$, and therefore

$$\lim_{T\to\infty} C_V(T) = Nk_{\mathrm{B}} \int_0^\infty d\omega\, g(\omega) = rdNk_{\mathrm{B}}\,. \tag{3.82}$$

This is the classical Dulong-Petit limit of $\frac{1}{2}k_{\mathrm{B}}$ per quadratic degree of freedom; there are rN atoms moving in d dimensions, hence $d \cdot rN$ positions and an equal number of momenta, resulting in a high temperature limit of $C_V = rdNk_{\mathrm{B}}$.

3.2.7 Einstein and Debye models

HIstorically, two models of lattice vibrations have received wide attention. First is the so-called *Einstein model*, in which there is no dispersion to the individual phonon modes. We approximate $g_a(\omega) \approx \delta(\omega - \omega_a)$, in which case

$$C_V(T) = N k_{\rm B} \sum_s \left(\frac{\hbar\omega_s}{2k_{\rm B}T} \right)^2 \mathrm{csch}^2\left(\frac{\hbar\omega_s}{2k_{\rm B}T} \right) . \tag{3.83}$$

At low temperatures, the contribution from each branch vanishes exponentially, because $\mathrm{csch}^2\big(\frac{\hbar\omega_s}{2k_{\rm B}T}\big) \simeq 4\, e^{-\hbar\omega_s/k_{\rm B}T} \to 0$. Real solids don't behave this way.

A more realistic model. due to Debye, accounts for the low-lying acoustic phonon branches. Since the acoustic phonon dispersion vanishes linearly with $|\boldsymbol{k}|$ as $\boldsymbol{k} \to 0$, there is no temperature at which the acoustic phonons 'freeze out' exponentially, as in the case of Einstein phonons. Indeed, the Einstein model is appropriate in describing the $d\,(r-1)$ optical phonon branches, though it fails miserably for the acoustic branches.

In the vicinity of the zone center $\boldsymbol{k} = 0$ (also called Γ in crystallographic notation) the d acoustic modes obey a linear dispersion, with $\omega_s(\boldsymbol{k}) = c_s(\hat{\boldsymbol{k}})\, k$. This results in an acoustic phonon density of states in $d = 3$ dimensions of

$$\begin{aligned} \tilde{g}(\omega) &= \frac{\mathcal{V}_0\, \omega^2}{2\pi^2} \sum_s \int \frac{d\hat{\boldsymbol{k}}}{4\pi}\, \frac{1}{c_s^3(\boldsymbol{k})}\, \Theta(\omega_{\rm D} - \omega) \\ &= \frac{3\mathcal{V}_0}{2\pi^2 \bar{c}^3}\, \omega^2\, \Theta(\omega_{\rm D} - \omega) \ , \end{aligned} \tag{3.84}$$

where $\bar{c}$ is an average acoustic phonon velocity (*i.e.* speed of sound) defined by

$$\frac{3}{\bar{c}^3} = \sum_s \int \frac{d\hat{\boldsymbol{k}}}{4\pi}\, \frac{1}{c_s^3(\boldsymbol{k})} \tag{3.85}$$

and $\omega_{\rm D}$ is a cutoff known as the *Debye frequency*. The cutoff is necessary because the phonon branch does not extend forever, but only to the boundaries of the Brillouin zone. Thus, $\omega_{\rm D}$ should roughly be equal to the energy of a zone boundary phonon. Alternatively, we can define $\omega_{\rm D}$ by the normalization condition

$$\int_0^\infty d\omega\, \tilde{g}(\omega) = 3 \quad \Longrightarrow \quad \omega_{\rm D} = (6\pi^2/\mathcal{V}_0)^{1/3}\, \bar{c}\, . \tag{3.86}$$

This allows us to write $\tilde{g}(\omega) = \big(9\omega^2/\omega_{\rm D}^3\big)\, \Theta(\omega_{\rm D} - \omega)$.

The specific heat due to the acoustic phonons is then

$$\begin{aligned} C_V(T) &= \frac{9Nk_{\rm B}}{\omega_{\rm D}^3}\int\limits_0^{\omega_{\rm D}} d\omega\,\omega^2\left(\frac{\hbar\omega}{2k_{\rm B}T}\right)^2 {\rm csch}^2\left(\frac{\hbar\omega}{2k_{\rm B}T}\right) \\ &= 9Nk_{\rm B}\left(\frac{2T}{\Theta_{\rm D}}\right)^3 \phi\big(\Theta_{\rm D}/2T\big)\ , \end{aligned} \tag{3.87}$$

where $\Theta_{\rm D} = \hbar\omega_{\rm D}/k_{\rm B}$ is the *Debye temperature* and

$$\phi(x) = \int\limits_0^x dt\,t^4\,{\rm csch}^2 t = \begin{cases} \frac{1}{3}x^3 & x \to 0 \\ \\ \frac{\pi^4}{30} & x \to \infty\ . \end{cases} \tag{3.88}$$

Therefore,

$$C_V(T) = \begin{cases} \frac{12\pi^4}{5}\,Nk_{\rm B}\left(\frac{T}{\Theta_{\rm D}}\right)^3 & T \ll \Theta_{\rm D} \\ \\ 3Nk_{\rm B} & T \gg \Theta_{\rm D}\ . \end{cases} \tag{3.89}$$

Thus, the heat capacity due to acoustic phonons obeys the Dulong-Petit rule in that $C_V(T \to \infty) = 3Nk_{\rm B}$, corresponding to the three acoustic degrees of freedom per unit cell. The remaining contribution of $3(r-1)Nk_{\rm B}$ to the high temperature heat capacity comes from the optical modes not considered in the Debye model. The low temperature T^3 behavior of the heat capacity of crystalline solids is a generic feature, and its detailed description is a triumph of the Debye model.

3.2.8 Phenomenological theory of melting

Atomic fluctuations in a crystal

For the one-dimensional chain, eqn. 3.33 gives

$$\tilde{u}_k = i\left(\frac{\hbar}{2m\omega_k}\right)^{1/2}\big(a_k - a_{-k}^\dagger\big)\ . \tag{3.90}$$

Therefore the RMS fluctuations at each site are given by

$$\langle u_n^2\rangle = \frac{1}{N}\sum_k \langle \tilde{u}_k\,\tilde{u}_{-k}\rangle = \frac{1}{N}\sum_k \frac{\hbar}{m\omega_k}\left(n(k) + \tfrac{1}{2}\right)\ , \tag{3.91}$$

where $n(k,T) = \big[\exp(\hbar\omega_k/k_{\rm B}T) - 1\big]^{-1}$ is the Bose occupancy function.

Element	Ag	Al	Au	C	Cd	Cr	Cu	Fe	Mn
$\Theta_{\rm D}$ (K)	227	433	162	2250	210	606	347	477	409
$T_{\rm melt}$ (K)	962	660	1064	3500	321	1857	1083	1535	1245
Element	Ni	Pb	Pt	Si	Sn	Ta	Ti	W	Zn
$\Theta_{\rm D}$ (K)	477	105	237	645	199	246	420	383	329
$T_{\rm melt}$ (K)	1453	327	1772	1410	232	2996	1660	3410	420

Table 3.2: Debye temperatures (at $T = 0$) and melting points for some common elements (carbon is assumed to be diamond and not graphite). (*Source: the internet!*)

Let us now generalize this expression to the case of a d-dimensional solid. The appropriate expression for the RMS position fluctuations of the $i^{\rm th}$ basis atom in each unit cell is

$$\langle \boldsymbol{u}_l^2(\boldsymbol{R}) \rangle = \frac{1}{N} \sum_{\boldsymbol{k}} \sum_{s=1}^{dr} \frac{\hbar}{\mathcal{M}_{ls}(\boldsymbol{k})\,\omega_s(\boldsymbol{k})} \left(n_s(\boldsymbol{k}) + \tfrac{1}{2} \right) . \tag{3.92}$$

Here we sum over all wavevectors $\boldsymbol{k}$ in the first Brilliouin zone, and over all normal modes a. There are dr normal modes per unit cell *i.e.* d branches of the phonon dispersion $\omega_s(\boldsymbol{k})$. (For the one-dimensional chain with $d = 1$ and $r = 1$ there was only one such branch to consider). Note also the quantity $M_{is}(\boldsymbol{k})$, which has units of mass and is defined in terms of the polarization vectors $\mathsf{e}^\alpha_{ls}(\boldsymbol{k})$ as

$$\frac{1}{\mathcal{M}_{ls}(\boldsymbol{k})} = \sum_{\mu=1}^{d} \left| \mathsf{e}^\mu_{ls}(\boldsymbol{k}) \right|^2 . \tag{3.93}$$

The dimensions of the polarization vector are $[\text{mass}]^{-1/2}$, since the generalized orthonormality condition on the normal modes is

$$\sum_{l,\mu} M_l\, \mathsf{e}^{\mu\,*}_{ls}(\boldsymbol{k})\, \mathsf{e}^\mu_{ls'}(\boldsymbol{k}) = \delta_{ss'} , \tag{3.94}$$

where M_i is the mass of the atom of species i within the unit cell ($i \in \{1, \ldots, r\}$). For our purposes we can replace $\mathcal{M}_{is}(\boldsymbol{k})$ by an appropriately averaged quantity which we call $\mathcal{M}_i$; this 'effective mass' is then independent of the mode index a as well as the wavevector $\boldsymbol{k}$. We may then write

$$\langle\, \boldsymbol{u}_l^2 \,\rangle \approx \int_0^\infty d\omega\; g(\omega)\, \frac{\hbar}{\mathcal{M}_l\,\omega} \cdot \left\{ \frac{1}{e^{\hbar\omega/k_{\rm B}T} - 1} + \frac{1}{2} \right\} , \tag{3.95}$$

where we have dropped the site label $\boldsymbol{R}$ since translational invariance guarantees that the fluctuations are the same from one unit cell to the next. Note that the fluctuations $\langle\, \boldsymbol{u}_i^2 \,\rangle$ can be divided into a temperature-dependent part $\langle\, \boldsymbol{u}_i^2 \,\rangle_{\rm th}$ and a temperature-independent quantum

contribution $\langle\, \boldsymbol{u}_i^2 \,\rangle_{\rm qu}$, where

$$\begin{aligned}\langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm th} &= \frac{\hbar}{\mathcal{M}_l} \int\limits_0^\infty d\omega\, \frac{g(\omega)}{\omega} \cdot \frac{1}{e^{\hbar\omega/k_{\rm B}T} - 1} \\ \langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm qu} &= \frac{\hbar}{2\mathcal{M}_l} \int\limits_0^\infty d\omega\, \frac{g(\omega)}{\omega}\ .\end{aligned} \tag{3.96}$$

Let's evaluate these contributions within the Debye model, where we replace $g(\omega)$ by

$$\bar{g}(\omega) = \frac{d^2\,\omega^{d-1}}{\omega_{\rm D}^d}\,\Theta(\omega_{\rm D} - \omega)\ . \tag{3.97}$$

We then find

$$\begin{aligned}\langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm th} &= \frac{d^2\hbar}{\mathcal{M}_l\,\omega_{\rm D}} \left(\frac{k_{\rm B}T}{\hbar\omega_{\rm D}}\right)^{d-1} F_d(\hbar\omega_{\rm D}/k_{\rm B}T) \\ \langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm qu} &= \frac{d^2}{d-1} \cdot \frac{\hbar}{2\mathcal{M}_l\,\omega_{\rm D}}\ ,\end{aligned} \tag{3.98}$$

where

$$F_d(x) = \int\limits_0^x dy\, \frac{y^{d-2}}{e^y - 1} = \begin{cases} \frac{x^{d-2}}{d-2} & x \to 0 \\ \\ \zeta(d-1) & x \to \infty \end{cases}\ . \tag{3.99}$$

We can now extract from these expressions several important conclusions:

1) The $T = 0$ contribution to the the fluctuations, $\langle \boldsymbol{u}_l^2 \rangle_{\rm qu}$, diverges in $d = 1$ dimensions. *Therefore there are no one-dimensional quantum solids.*

2) The thermal contribution to the fluctuations, $\langle \boldsymbol{u}_l^2 \rangle_{\rm th}$, diverges for any $T > 0$ whenever $d \leqslant 2$. This is because the integrand of $F_d(x)$ goes as y^{d-3} as $y \to 0$. *Therefore, there are no two-dimensional classical solids.*

3) Both the above conclusions are valid in the thermodynamic limit. Finite size imposes a cutoff on the frequency integrals, because there is a smallest wavevector $k_{\rm min} \sim 2\pi/L$, where L is the (finite) linear dimension of the system. This leads to a low frequency cutoff $\omega_{\rm min} = 2\pi\bar{c}/L$, where $\bar{c}$ is the appropriately averaged acoustic phonon velocity from eqn. 3.85, which mitigates any divergences.

Lindemann melting criterion

An old phenomenological theory of melting due to Lindemann says that a crystalline solid melts when the RMS fluctuations in the atomic positions exceeds a certain fraction η of the lattice constant a. We therefore define the ratios

$$\begin{aligned} x_{l,\rm th}^2 &\equiv \frac{\langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm th}}{a^2} = d^2 \cdot \left(\frac{\hbar^2}{\mathcal{M}_l\, a^2\, k_{\rm B}}\right) \cdot \frac{T^{d-1}}{\Theta_{\rm D}^d} \cdot F(\Theta_{\rm D}/T) \\ x_{l,\rm qu}^2 &\equiv \frac{\langle\, \boldsymbol{u}_l^2 \,\rangle_{\rm qu}}{a^2} = \frac{d^2}{2(d-1)} \cdot \left(\frac{\hbar^2}{\mathcal{M}_l\, a^2\, k_{\rm B}}\right) \cdot \frac{1}{\Theta_{\rm D}} \quad , \end{aligned} \tag{3.100}$$

with $x_l = \sqrt{x_{l,\rm th}^2 + x_{l,\rm qu}^2} = \sqrt{\langle\, \boldsymbol{u}_l^2 \,\rangle}\Big/a$.

Let's now work through an example of a three-dimensional solid. We'll assume a single element basis ($r = 1$). We have that

$$\frac{9\hbar^2/4k_{\rm B}}{1\,\mathtt{amu}\,\text{Å}^2} = 109\,\text{K}\,. \tag{3.101}$$

According to table 3.2, the melting temperature always exceeds the Debye temperature, and often by a great amount. We therefore assume $T \gg \Theta_{\rm D}$, which puts us in the small x limit of $F_d(x)$. We then find

$$x_{\rm qu}^2 = \frac{\Theta^\star}{\Theta_{\rm D}} \qquad , \qquad x_{\rm th}^2 = \frac{\Theta^\star}{\Theta_{\rm D}} \cdot \frac{4T}{\Theta_{\rm D}} \qquad , \qquad x = \sqrt{\left(1 + \frac{4T}{\Theta_{\rm D}}\right)\frac{\Theta^\star}{\Theta_{\rm D}}}\,. \tag{3.102}$$

where

$$\Theta^* = \frac{109\,\text{K}}{M[\mathtt{amu}] \cdot \big(a[\text{Å}]\big)^2}\,. \tag{3.103}$$

The total position fluctuation is of course the sum $x_l^2 = x_{l,\rm th}^2 + x_{l,\rm qu}^2$. Consider for example the case of copper, with $M = 56\,\mathtt{amu}$ and $a = 2.87\,\text{Å}$. The Debye temperature is $\Theta_{\rm D} = 347\,\text{K}$. From this we find $x_{\rm qu} = 0.026$, which says that at $T = 0$ the RMS fluctuations of the atomic positions are not quite three percent of the lattice spacing (*i.e.* the distance between neighboring copper atoms). At room temperature, $T = 293\,\text{K}$, one finds $x_{\rm th} = 0.048$, which is about twice as large as the quantum contribution. How big are the atomic position fluctuations at the melting point? According to our table, $T_{\rm melt} = 1083\,\text{K}$ for copper, and from our formulae we obtain $x_{\rm melt} = 0.096$. The *Lindemann criterion* says that solids melt when $x(T) \approx 0.1$.

We were very lucky to hit the magic number $x_{\rm melt} = 0.1$ with copper. Let's try another example. Lead has $M = 208\,\mathtt{amu}$ and $a = 4.95\,\text{Å}$. The Debye temperature is $\Theta_{\rm D} = 105\,\text{K}$ ('soft phonons'), and the melting point is $T_{\rm melt} = 327\,\text{K}$. From these data we obtain $x(T = 0) = 0.014$, $x(293\,\text{K}) = 0.050$ and $x(T = 327\,\text{K}) = 0.053$. Same ballpark.

We can turn the analysis around and predict a melting temperature based on the Lindemann criterion $x(T_{\rm melt}) = \eta$, where $\eta \approx 0.1$. We obtain

$$T_{\rm L} = \left(\frac{\eta^2\,\Theta_{\rm D}}{\Theta^\star} - 1 \right) \cdot \frac{\Theta_{\rm D}}{4} \ . \tag{3.104}$$

We call $T_{\rm L}$ the *Lindemann temperature*. Most treatments of the Lindemann criterion ignore the quantum correction, which gives the -1 contribution inside the above parentheses. But if we are more careful and include it, we see that it may be possible to have $T_{\rm L} < 0$. This occurs for any crystal where $\Theta_{\rm D} < \Theta^\star/\eta^2$.

Consider for example the case of ^{4}He, which at atmospheric pressure condenses into a liquid at $T_{\rm c} = 4.2\,{\rm K}$ and remains in the liquid state down to absolute zero. At $p = 1\,{\rm atm}$, it never solidifies! Why? The number density of liquid ^{4}He at $p = 1$ atm and $T = 0\,{\rm K}$ is $2.2 \times 10^{22}\,{\rm cm}^{-3}$. Let's say the Helium atoms want to form a crystalline lattice. We don't know *a priori* what the lattice structure will be, so let's for the sake of simplicity assume a simple cubic lattice. From the number density we obtain a lattice spacing of $a = 3.57\,{\rm Å}$. OK now what do we take for the Debye temperature? Theoretically this should depend on the microscopic force constants which enter the small oscillations problem (*i.e.* the spring constants between pairs of helium atoms in equilibrium). We'll use the expression we derived for the Debye frequency, $\omega_{\rm D} = (6\pi^2/\mathcal{V}_0)^{1/3}\bar{c}$, where $\mathcal{V}_0$ is the unit cell volume. We'll take $\bar{c} = 238\,{\rm m/s}$, which is the speed of sound in liquid helium at $T = 0$. This gives $\Theta_{\rm D} = 19.8\,{\rm K}$. We find $\Theta^\star = 2.13\,{\rm K}$, and if we take $\eta = 0.1$ this gives $\Theta^\star/\eta^2 = 213\,{\rm K}$, which significantly exceeds $\Theta_{\rm D}$. Thus, the solid should melt because the RMS fluctuations in the atomic positions at absolute zero are huge: $x_{\rm qu} = (\Theta^\star/\Theta_{\rm D})^{1/2} = 0.33$. By applying pressure, one can get ^{4}He to crystallize above $p_{\rm c} = 25\,{\rm atm}$ (at absolute zero). Under pressure, the unit cell volume $\mathcal{V}_0$ decreases and the phonon velocity $\bar{c}$ increases, so the Debye temperature itself increases.

It is important to recognize that the Lindemann criterion does not provide us with a theory of melting *per se*. Rather it provides us with a heuristic which allows us to predict roughly when a solid should melt.

3.2.9 Goldstone bosons

The vanishing of the acoustic phonon dispersion at $\boldsymbol{k} = 0$ is a consequence of *Goldstone's theorem* which says that associated with every *broken generator* of a *continuous symmetry* there is an associated bosonic gapless excitation (*i.e.* one whose frequency ω vanishes in the long wavelength limit). In the case of phonons, the 'broken generators' are the symmetries under spatial translation in the x, y, and z directions. The crystal selects a particular location for its center-of-mass, which breaks this symmetry. There are, accordingly, three gapless acoustic phonons.

Magnetic materials support another branch of elementary excitations known as spin waves, or *magnons*. In *isotropic* magnets, there is a global symmetry associated with rotations in internal

spin space, described by the group $\mathrm{SU}(2)$. If the system spontaneously magnetizes, meaning there is long-ranged ferromagnetic order ($\uparrow\uparrow\uparrow\ \cdots$), or long-ranged antiferromagnetic order ($\uparrow\downarrow\uparrow\downarrow\ \cdots$), then global spin rotation symmetry is broken. Typically a particular direction is chosen for the magnetic moment (or staggered moment, in the case of an antiferromagnet). Symmetry under rotations about this axis is then preserved, but rotations which do not preserve the selected axis are 'broken'. In the most straightforward case, that of the antiferromagnet, there are two such rotations for $\mathrm{SU}(2)$, and concomitantly two gapless magnon branches, with linearly vanishing dispersions $\omega_a(\boldsymbol{k})$. The situation is more subtle in the case of ferromagnets, because the total magnetization is conserved by the dynamics (unlike the total staggered magnetization in the case of antiferromagnets). Another wrinkle arises if there are long-ranged interactions present.

For our purposes, we can safely ignore the deep physical reasons underlying the gaplessness of Goldstone bosons and simply posit a gapless dispersion relation of the form $\omega(\boldsymbol{k}) = A\,|\boldsymbol{k}|^\sigma$. The density of states for this excitation branch is then

$$g(\omega) = \mathcal{C}\,\omega^{\frac{d}{\sigma}-1}\,\Theta(\omega_{\rm c}-\omega)\ , \tag{3.105}$$

where $\mathcal{C}$ is a constant and $\omega_{\rm c}$ is the cutoff, which is the bandwidth for this excitation branch.[6] Normalizing the density of states for this branch results in the identification $\omega_{\rm c} = (d/\sigma\mathcal{C})^{\sigma/d}$.

The heat capacity is then found to be

$$\begin{aligned} C_V &= N k_{\rm B}\,\mathcal{C}\int\limits_0^{\omega_{\rm c}} d\omega\,\omega^{\frac{d}{\sigma}-1}\left(\frac{\hbar\omega}{k_{\rm B}T}\right)^2 \mathrm{csch}^2\!\left(\frac{\hbar\omega}{2k_{\rm B}T}\right) \\ &= \frac{d}{\sigma}\,N k_{\rm B}\left(\frac{2T}{\Theta}\right)^{d/\sigma}\phi\big(\Theta/2T\big)\ , \end{aligned} \tag{3.106}$$

where $\Theta = \hbar\omega_{\rm c}/k_{\rm B}$ and

$$\phi(x) = \int\limits_0^x dt\,t^{\frac{d}{\sigma}+1}\,\mathrm{csch}^2 t = \begin{cases} \frac{\sigma}{d}\,x^{d/\sigma} & x\to 0 \\ \\ 2^{-d/\sigma}\,\Gamma\big(2+\frac{d}{\sigma}\big)\,\zeta\big(2+\frac{d}{\sigma}\big) & x\to\infty\ , \end{cases} \tag{3.107}$$

which is a generalization of our earlier results. Once again, we recover Dulong-Petit for $k_{\rm B}T \gg \hbar\omega_{\rm c}$, with $C_V(T \gg \hbar\omega_{\rm c}/k_{\rm B}) = N k_{\rm B}$.

In an isotropic ferromagnet, *i.e.*a ferromagnetic material where there is full SU(2) symmetry in internal 'spin' space, the magnons have a k^2 dispersion. Thus, a bulk three-dimensional isotropic ferromagnet will exhibit a heat capacity due to spin waves which behaves as $T^{3/2}$ at low temperatures. For sufficiently low temperatures this will overwhelm the phonon contribution, which behaves as T^3.

[6] If $\omega(\boldsymbol{k}) = Ak^\sigma$, then $\mathcal{C} = 2^{1-d}\,\pi^{-\frac{d}{2}}\,\sigma^{-1}\,A^{-\frac{d}{\sigma}}\,\mathsf{g}\,/\,\Gamma(d/2)$.

3.2.10 Elasticity theory redux : Bravais lattices

In a Bravais lattice, we have $\hat{\Phi}^{\alpha\beta}(\mathbf{0}) = 0$ from translational invariance. The potential energy may then be written in the form

$$U = U_0 - \frac{1}{4}\sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{\alpha,\beta}\left[u^\alpha(\boldsymbol{R}) - u^\alpha(\boldsymbol{R}')\right]\Phi^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\left[u^\beta(\boldsymbol{R}) - u^\beta(\boldsymbol{R}')\right] \quad . \tag{3.108}$$

We now assume a very long wavelength disturbance, and write

$$u^\alpha(\boldsymbol{R}) - u^\alpha(\boldsymbol{R}') = (R^\mu - R'^\mu)\left.\frac{\partial u^\alpha}{\partial x^\mu}\right|_{\boldsymbol{R}} + \ldots \quad . \tag{3.109}$$

Thus,

$$U = U_0 - \frac{1}{4}\sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{\alpha,\beta}\sum_{\mu,\nu}\left.\frac{\partial u^\alpha}{\partial x^\mu}\right|_{\boldsymbol{R}}\left.\frac{\partial u^\beta}{\partial x^\nu}\right|_{\boldsymbol{R}}(R^\mu - R'^\mu)\,(R^\nu - R'^\nu)\,\Phi^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}') \quad . \tag{3.110}$$

We may symmetrize with respect to Cartesian indices[7] to obtain the elastic tensor

$$C_{\alpha\beta\mu\nu} \equiv -\frac{1}{8\Omega}\sum_{\boldsymbol{R}}\Big(R^\mu R^\nu\,\Phi^{\alpha\beta}(\boldsymbol{R}) + R^\mu R^\beta\,\Phi^{\alpha\nu}(\boldsymbol{R}) + R^\alpha R^\nu\,\Phi^{\mu\beta}(\boldsymbol{R}) + R^\alpha R^\beta\,\Phi^{\mu\nu}(\boldsymbol{R})\Big) \quad . \tag{3.111}$$

Note that

$$C_{\alpha\beta\mu\nu} = C_{\beta\alpha\mu\nu} = C_{\alpha\beta\nu\mu} = C_{\mu\nu\alpha\beta} \quad , \tag{3.112}$$

where Ω is the Wigner-Seitz cell volume.

Elasticity in solids

Recall from §3.1.2 that we may regard the rank four tensor $C_{\alpha\beta\mu\nu}$ as a symmetric 6×6 matrix C_{ab}, where we replace $(\alpha\beta)\to a$ and $(\mu\nu)\to b$ according to the scheme from Tab. 3.1. In cubic crystals, for example, we have

$$\begin{aligned} C_{11} &= C_{xxxx} = C_{yyyy} = C_{zzzz} \\ C_{12} &= C_{xxyy} = C_{xxzz} = C_{yyzz} \\ C_{44} &= C_{xyxy} = C_{xzxz} = C_{yzyz} \quad . \end{aligned} \tag{3.113}$$

Typical values of C_{ab} in solids are on the order of gigapascals, *i.e.* 10^9 Pa:

The *bulk modulus* of a solid is defined as $\mathcal{B} = V\,\partial^2 F/\partial V^2$. We consider a uniform dilation, which is described by $\boldsymbol{R}\to(1+\zeta)\boldsymbol{R}$ at each lattice site. Thus the displacement field is $\boldsymbol{u}(\boldsymbol{r}) = \zeta\boldsymbol{r}$.

[7] Symmetrization is valid because the antisymmetric combination $\left(\frac{\partial u^\alpha}{\partial x^\beta} - \frac{\partial u^\beta}{\partial x^\alpha}\right)$ corresponds to a rotation.

element	C_{11}	C_{12}	C_{44}
^{4}He	0.031	0.028	0.022
Cu	16	8	12
Al	108	62	28.3
Pb	48.8	41.4	14.8
C (diamond)	1040	170	550

Table 3.3: Elastic moduli for various solids (in GPa).

This leads to a volume change of $\delta V = 3\zeta V$, hence $\zeta = \delta V/3V$. The strain tensor is $\varepsilon_{\alpha\beta} = \zeta\,\delta_{\alpha\beta}$, hence

$$\delta F = \frac{(\delta V)^2}{18V}\sum_{\alpha,\beta} C_{\alpha\alpha\beta\beta} = \tfrac{1}{9}\sum_{a,b=1}^{3} C_{ab} \quad . \tag{3.114}$$

Thus, for cubic materials, $\mathcal{B} = \frac{1}{3}C_{11} + \frac{2}{3}C_{12}$.

Elastic waves

The Lagrangian of an elastic medium is be written as

$$L = \int d^dr\,\mathcal{L} = \int d^dr\left\{\tfrac{1}{2}\,\rho\left(\frac{\partial u^\alpha}{\partial t}\right)^2 - \tfrac{1}{2}\,C_{\alpha\beta\mu\nu}\,\frac{\partial u^\alpha}{\partial x^\beta}\,\frac{\partial u^\mu}{\partial x^\nu}\right\} \quad , \tag{3.115}$$

where ρ is the overall mass density of the crystal, *i.e.* $\rho = m/\Omega$. The Euler-Lagrange equations of motion are then

$$\begin{aligned} 0 &= \frac{\partial}{\partial t}\,\frac{\partial\mathcal{L}}{\partial(\partial_t u^\alpha)} + \frac{\partial}{\partial x^\beta}\,\frac{\partial\mathcal{L}}{\partial(\partial_\beta u^\alpha)} \\ &= \rho\,\frac{\partial^2 u^\alpha}{\partial t^2} - C_{\alpha\beta\mu\nu}\,\frac{\partial^2 u^\nu}{\partial x^\beta\,\partial x^\mu} \quad . \end{aligned} \tag{3.116}$$

The solutions are *elastic waves*, with $\boldsymbol{u}(\boldsymbol{x},t) = \hat{\mathrm{e}}(\boldsymbol{k})\,e^{i(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)}$ where

$$\rho\omega^2\,\mathrm{e}^\alpha(\boldsymbol{k}) = C_{\alpha\beta\mu\nu}\,k^\beta\,k^\mu\,\mathrm{e}^\nu(\boldsymbol{k}) \quad . \tag{3.117}$$

Thus, the dispersion is $\omega_a(\boldsymbol{k}) = c_a(\hat{\boldsymbol{k}})\,k$, where

$$\mathsf{det}\left[\rho\,c^2(\hat{\boldsymbol{k}})\,\delta_{\alpha\nu} - C_{\alpha\beta\mu\nu}\,\hat{k}^\beta\,\hat{k}^\mu\right] = 0 \tag{3.118}$$

is the equation to be solved for the speeds of sound $c_a(\hat{\boldsymbol{k}})$ in each elastic wave branch a.

For isotropic solids, $C_{12} \equiv \lambda$, $C_{44} \equiv \mu$, and $C_{11} = C_{12} + 2C_{14} = \lambda + 2\mu$, where λ and μ are the Lamé coefficients. The free energy density is discussed in §3.1 and is given by

$$f = \tfrac{1}{2}\lambda\,(\partial_i u_i)^2 + \tfrac{1}{2}\mu\,(\partial_i u_j)\,(\partial_i u_j) + \tfrac{1}{2}\mu\,(\partial_i u_j)\,(\partial_j u_i) \quad , \tag{3.119}$$

which results in the Euler-Lagrange equations of motion

$$\rho\,\ddot{\boldsymbol{u}} = (\lambda+\mu)\,\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\boldsymbol{u}) + \mu\,\nabla^2\boldsymbol{u} \quad . \tag{3.120}$$

Writing $\boldsymbol{u}(\boldsymbol{r},t) = u_0\,\hat{\mathbf{e}}(\boldsymbol{k})\,e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)}$, where $\hat{\mathbf{e}}$ is a polarization unit vector, we obtain a longitudinal mode when $\hat{\mathbf{e}}(\boldsymbol{k})\cdot\hat{\boldsymbol{k}} = 1$ with $\omega_{\text{L}}(\boldsymbol{k}) = c_{\text{L}}\,|\boldsymbol{k}|$ and $c_{\text{L}} = \sqrt{(\lambda+2\mu)/\rho}$, and two transverse modes when $\hat{\mathbf{e}}(\boldsymbol{k})\cdot\hat{\boldsymbol{k}} = 0$ with $\omega_{\text{T}}(\boldsymbol{k}) = c_{\text{T}}\,|\boldsymbol{k}|$ and $c_{\text{T}} = \sqrt{\mu/\rho}$.

In cubic crystals, there are three independent elastic moduli, C_{11}, C_{12}, and C_{14}. We then have

$$\begin{aligned}
\rho\,c^2(\hat{\boldsymbol{k}})\,\mathrm{e}^x &= \Big[C_{11}\,\hat{k}_x^2 + C_{44}\big(\hat{k}_y^2+\hat{k}_z^2\big)\Big]\hat{\mathrm{e}}^x + (C_{12}+C_{44})\big(\hat{k}_x\,\hat{k}_y\,\hat{\mathrm{e}}^y + \hat{k}_x\,\hat{k}_z\,\hat{\mathrm{e}}^z\big) \\
\rho\,c^2(\hat{\boldsymbol{k}})\,\mathrm{e}^y &= \Big[C_{11}\,\hat{k}_y^2 + C_{44}\big(\hat{k}_x^2+\hat{k}_z^2\big)\Big]\hat{\mathrm{e}}^y + (C_{12}+C_{44})\big(\hat{k}_x\,\hat{k}_y\,\hat{\mathrm{e}}^x + \hat{k}_y\,\hat{k}_z\,\hat{\mathrm{e}}^z\big) \\
\rho\,c^2(\hat{\boldsymbol{k}})\,\mathrm{e}^z &= \Big[C_{11}\,\hat{k}_z^2 + C_{44}\big(\hat{k}_x^2+\hat{k}_y^2\big)\Big]\hat{\mathrm{e}}^z + (C_{12}+C_{44})\big(\hat{k}_x\,\hat{k}_z\,\hat{\mathrm{e}}^x + \hat{k}_y\,\hat{k}_z\,\hat{\mathrm{e}}^y\big) \quad .
\end{aligned} \tag{3.121}$$

This still yields a cubic equation, but it can be simplified by looking along a high symmetry direction in the Brillouin zone.

Along the (100) direction $\boldsymbol{k} = k\,\hat{\boldsymbol{x}}$, we have

$$\hat{\mathbf{e}}_{\text{L}} = \hat{\boldsymbol{x}} \qquad c_{\text{L}} = \sqrt{C_{11}/\rho} \tag{3.122}$$

$$\hat{\mathbf{e}}_{\text{T1}} = \hat{\boldsymbol{y}} \qquad c_{\text{T1}} = \sqrt{C_{44}/\rho} \tag{3.123}$$

$$\hat{\mathbf{e}}_{\text{T2}} = \hat{\boldsymbol{z}} \qquad c_{\text{T2}} = \sqrt{C_{44}/\rho} \quad . \tag{3.124}$$

Along the (110) direction, we have $\boldsymbol{k} = \frac{1}{\sqrt{2}}\,k\,(\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}})$. In this case

$$\hat{\mathbf{e}}_{\text{L}} = \tfrac{1}{\sqrt{2}}(\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}}) \qquad c_{\text{L}} = \sqrt{(C_{11}+2C_{12}+4C_{44})/3\rho} \tag{3.125}$$

$$\hat{\mathbf{e}}_{\text{T1}} = \tfrac{1}{\sqrt{2}}(\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}}) \qquad c_{\text{T1}} = \sqrt{(C_{11}-C_{12})/2\rho} \tag{3.126}$$

$$\hat{\mathbf{e}}_{\text{T2}} = \hat{\boldsymbol{z}} \qquad c_{\text{T2}} = \sqrt{C_{44}/\rho} \quad . \tag{3.127}$$

Along the (111) direction, we have $\boldsymbol{k} = \frac{1}{\sqrt{3}}\,k\,(\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}}+\hat{\boldsymbol{z}})$. In this case

$$\hat{\mathbf{e}}_{\text{L}} = \tfrac{1}{\sqrt{3}}(\hat{\boldsymbol{x}}+\hat{\boldsymbol{y}}+\hat{\boldsymbol{z}}) \qquad c_{\text{L}} = \sqrt{(C_{11}+C_{12}+2C_{44})/2\rho} \tag{3.128}$$

$$\hat{\mathbf{e}}_{\text{T1}} = \tfrac{1}{\sqrt{6}}(2\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}}-\hat{\boldsymbol{z}}) \qquad c_{\text{T1}} = \sqrt{(C_{11}-C_{12})/3\rho} \tag{3.129}$$

$$\hat{\mathbf{e}}_{\text{T2}} = \tfrac{1}{\sqrt{2}}(\hat{\boldsymbol{y}}-\hat{\boldsymbol{z}}) \qquad c_{\text{T2}} = \sqrt{(C_{11}-C_{12})/3\rho} \quad . \tag{3.130}$$

3.2.11 Elasticity theory in cases with bases

The derivation of the elastic tensor $C_{\alpha\beta\mu\nu}$ is significantly complicated by the presence of a basis. Sadly, translational invariance if of no direct avail because

$$U \neq U_0 - \frac{1}{4} \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{\alpha,\beta} \sum_{i,j} \left[u_i^\alpha(\boldsymbol{R}) - u_i^\alpha(\boldsymbol{R}') \right] \Phi_{ij}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}') \left[u_j^\beta(\boldsymbol{R}) - u_j^\beta(\boldsymbol{R}') \right] \quad . \tag{3.131}$$

The student should understand why the above relation is not an equality.

Rather than work with the energy, we will work with the eigenvalue equation 3.42,

$$\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})\, \hat{\mathrm{e}}_{j\lambda}^{\beta}(\boldsymbol{k}) = m_i\, \omega_\lambda^2(\boldsymbol{k})\, \hat{\mathrm{e}}_{i\lambda}^{\alpha}(\boldsymbol{k}) \quad ,$$

and expand in powers of $\boldsymbol{k}$. Accordingly, we write

$$\begin{aligned} \hat{\mathrm{e}}_{i\lambda}^{\alpha}(\boldsymbol{k}) &= d_i^\alpha + k^\sigma f_{i\sigma}^\alpha + \tfrac{1}{2}\, k^\sigma k^\tau g_{i\sigma\tau}^\beta + \mathcal{O}(k^3) \\ \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k}) &= \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{0}) + k^\mu \left.\frac{\partial \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\mu}\right|_{\boldsymbol{0}} + \tfrac{1}{2}\, k^\mu\, k^\nu \left.\frac{\partial^2 \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\mu\, \partial k^\nu}\right|_{\boldsymbol{0}} + \mathcal{O}(k^3) \quad . \end{aligned} \tag{3.132}$$

We retain the basis index i on d_i^α *even though it is independent of* i because we will use it to make clear certain necessary sums on the basis index within the Einstein convention. We then have

$$\begin{aligned} m_i\, \omega^2 \left\{ d_i^\alpha + k^\sigma f_{i\sigma}^\alpha + \ldots \right\} = & \\ \left\{ \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{0}) + k^\mu \left.\frac{\partial \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\mu}\right|_{\boldsymbol{0}} + \tfrac{1}{2}\, k^\mu\, k^\nu \left.\frac{\partial^2 \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\mu\, \partial k^\nu}\right|_{\boldsymbol{0}} + \ldots \right\} & \left\{ d_j^\beta + k^\tau f_{j\tau}^\beta + \ldots \right\} \quad , \end{aligned} \tag{3.133}$$

where there is no implied sum on i on the LHS. We now work order by order in $\boldsymbol{k}$. To start, note that $\omega^2(\boldsymbol{k}) = c^2(\hat{\boldsymbol{k}})\, k^2$ is already second order. On the RHS, we have $\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{0})\, d_j^\beta = 0$ to zeroth order in k. At first order, we must have

$$\hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{0})\, f_{j\sigma}^\beta + \left.\frac{\partial \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\sigma}\right|_{\boldsymbol{0}} d_j^\beta = 0 \quad , \tag{3.134}$$

and defining the matrix inverse $\hat{\Upsilon}_{li}^{\gamma\alpha}(\boldsymbol{k})$ by the relation

$$\hat{\Upsilon}_{li}^{\gamma\alpha}(\boldsymbol{k})\, \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k}) = \delta^{\gamma\beta}\, \delta_{lj} \quad , \tag{3.135}$$

we have

$$f_{l\sigma}^\gamma = -\hat{\Upsilon}_{li}^{\gamma\alpha}(\boldsymbol{0}) \left.\frac{\partial \hat{\Phi}_{ij}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\sigma}\right|_{\boldsymbol{0}} d_j^\beta \tag{3.136}$$

Finally, we obtain the eigenvalue equation for the elastic waves,

$$m_i\,\omega^2\,d_i^\alpha = \left[\frac{1}{2}\,\frac{\partial^2\hat\Phi_{il}^{\alpha\beta}(\boldsymbol{k})}{\partial k^\mu\,\partial k^\nu}\bigg|_{\mathbf{0}} - \frac{\partial\hat\Phi_{ij}^{\alpha\sigma}(\boldsymbol{k})}{\partial k^\mu}\bigg|_{\mathbf{0}}\,\hat\Upsilon_{jm}^{\sigma\gamma}(\mathbf{0})\,\frac{\partial\hat\Phi_{ml}^{\gamma\beta}(\boldsymbol{k})}{\partial k^\nu}\bigg|_{\mathbf{0}}\right]k^\mu\,k^\nu\,d_l^\beta \quad . \tag{3.137}$$

Remember that d_i^α is independent of the basis index i. We have dropped the mode index λ here for notational convenience. Note that the quadratic coefficient $g_{i\sigma\tau}^\beta$ never entered our calculation because it leads to an inhomogeneous term in the eigenvalue equation, and therefore must be dropped. We do not report here the explicit form for the elastic tensor, which may be derived from the above eigenvalue equation.

3.3 Neutron diffraction

3.3.1 Inelastic differential scattering cross-section

Elastic X-ray scattering yields a measure of the static structure factor of a crystal,

$$S(\boldsymbol{q}) = \frac{1}{N}\sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{i,j} e^{i\boldsymbol{q}\cdot(\boldsymbol{R}'+\boldsymbol{\delta}_j-\boldsymbol{R}-\boldsymbol{\delta}_i)} \quad . \tag{3.138}$$

The wavevector transfer is $\boldsymbol{q} = \boldsymbol{k}_f - \boldsymbol{k}_i$. Now consider an inelastic process between states

$$|\,\Psi_i\,\rangle \equiv \Big|\,\boldsymbol{k}_i\,,\,\{n_s^i(\boldsymbol{k})\}\,\Big\rangle \to \Big|\,\boldsymbol{k}_f\,,\,\{n_s^f(\boldsymbol{k})\}\,\Big\rangle \equiv |\,\Psi_f\,\rangle \quad . \tag{3.139}$$

The initial and final energies are given by

$$\begin{aligned} E_i &= \frac{\hbar^2\boldsymbol{k}_i^2}{2m_{\mathsf{n}}} + \sum_{\boldsymbol{k},s}\hbar\omega_s(\boldsymbol{k})\left(n_s^i(\boldsymbol{k})+\tfrac{1}{2}\right) \\ E_f &= \frac{\hbar^2\boldsymbol{k}_i^2}{2m_{\mathsf{n}}} + \sum_{\boldsymbol{k},s}\hbar\omega_s(\boldsymbol{k})\left(n_s^f(\boldsymbol{k})+\tfrac{1}{2}\right) \quad . \end{aligned} \tag{3.140}$$

Energy conservation requires $E_f = E_i$, and we define the energy transfer to the lattice to be

$$\hbar\omega \equiv \frac{\hbar^2}{2m_{\mathsf{n}}}\big(\boldsymbol{k}_i^2 - \boldsymbol{k}_f^2\big) = \sum_{\boldsymbol{k},s}\hbar\omega_s(\boldsymbol{k})\left(n_s^f(\boldsymbol{k}) - n_s^i(\boldsymbol{k})\right) \equiv \mathcal{E}_f - \mathcal{E}_i \quad , \tag{3.141}$$

where $\mathcal{E} = \sum_{\boldsymbol{k},s}\hbar\omega_s(\boldsymbol{k})\left(n_s(\boldsymbol{k})+\frac{1}{2}\right)$ is the energy of the lattice vibrations.

The scattering rate from $|\,\Psi_i\,\rangle$ to $|\,\Psi_f\,\rangle$ is given by Fermi's Golden Rule, *viz.*

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}\left|\langle\,f\,|\,V\,|\,i\,\rangle\right|^2\delta(E_f - E_i) \quad , \tag{3.142}$$

from which we derive the differential scattering cross section

$$\frac{\partial^2\sigma}{\partial\Omega\,\partial\omega} = \frac{\hbar}{4\pi v_i}\sum_{i,f} P_i\,\Gamma_{i\to f}\,g(k_f) \quad , \tag{3.143}$$

where P_i is the Boltzmann weight for the lattice state $|\,i\,\rangle = |\,\{n_s^i(\boldsymbol{k})\}\,\rangle$, and where $g(k_f)$ is the density of states,

$$g(k_f) = \int \frac{d^3k}{(2\pi)^3}\,\delta\!\left(\frac{\hbar^2\boldsymbol{k}^2}{2m_{\mathsf{n}}} - \frac{\hbar^2\boldsymbol{k}_f^2}{2m_{\mathsf{n}}}\right) = \frac{m_{\mathsf{n}}k_f}{2\pi^2\hbar^2} \quad . \tag{3.144}$$

Thus,

$$\frac{\partial^2\sigma}{\partial\Omega\,\partial\omega} = \frac{m_{\mathsf{n}}^2\,k_i}{4\pi^2\hbar^3\,k_f}\sum_{i,f} P_i\,\big|\langle\,\boldsymbol{k}_f, f\,|\,V\,|\,\boldsymbol{k}_i, i\,\rangle\big|^2\delta(E_f - E_i) \quad . \tag{3.145}$$

We may approximate the potential $V(\boldsymbol{r})$ as

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{R},l}\frac{2\pi a_l\hbar^2}{m_{\mathsf{n}}}\,\delta\big(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\delta}_l - \boldsymbol{u}_l(\boldsymbol{R})\big) \quad , \tag{3.146}$$

where a_l is the effective s-wave scattering length for ions of species l. The matrix element is thus

$$\langle\,\boldsymbol{k}_f, f\,|\,V\,|\,\boldsymbol{k}_i, i\,\rangle = \Big\langle\,f\,\Big|\sum_{\boldsymbol{R},l}\frac{2\pi\hbar^2 a_l}{m_{\mathsf{n}}}\,e^{-i\boldsymbol{q}\cdot(\boldsymbol{R}+\boldsymbol{\delta}_l)}\,e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R})}\,\Big|\,i\,\Big\rangle \quad , \tag{3.147}$$

which is an approximation of the more correct form

$$\begin{aligned} V(\boldsymbol{r}) &= \sum_{\boldsymbol{R},l} v_l\big(\boldsymbol{r} - \boldsymbol{R} - \boldsymbol{\delta}_l - \boldsymbol{u}_l(\boldsymbol{R})\big) \\ \langle\,\boldsymbol{k}_f\,|\,V\,|\,\boldsymbol{k}_i\,\rangle &= \sum_{\boldsymbol{R},l}\hat{v}(\boldsymbol{q})\,e^{-i\boldsymbol{q}\cdot(\boldsymbol{R}+\boldsymbol{\delta}_l)}\,e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R})} \quad . \end{aligned} \tag{3.148}$$

We now have

$$\frac{\partial^2\sigma}{\partial\Omega\,\partial\omega} = \frac{\hbar k_f}{k_i}\sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{l,l'} a_l\,a_{l'}^*\,e^{i\boldsymbol{q}\cdot(\boldsymbol{R}'-\boldsymbol{R}+\boldsymbol{\delta}_{l'}-\boldsymbol{\delta}_l)}\sum_{i,f} P_i\,\langle\,i\,|\,e^{i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{R}')}\,|\,f\,\rangle\,\langle\,f\,|\,e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R})}\,|\,i\,\rangle\,\delta(\hbar\omega + \mathcal{E}_i - \mathcal{E}_f) \quad . \tag{3.149}$$

Writing

$$\delta(\hbar\omega + \mathcal{E}_i - \mathcal{E}_f) = \int_0^\infty \frac{dt}{2\pi\hbar}\,e^{i\omega t}\,e^{i(\mathcal{E}_i-\mathcal{E}_f)t/\hbar} \quad , \tag{3.150}$$

we have,

$$\frac{\partial^2\sigma}{\partial\Omega\,\partial\omega} = N\,\frac{k_f}{k_i}\sum_{l,l'} a_l\,a_{l'}^*\,e^{i\boldsymbol{q}\cdot(\boldsymbol{\delta}_{l'}-\boldsymbol{\delta}_l)}\sum_{\boldsymbol{R}} e^{-i\boldsymbol{q}\cdot\boldsymbol{R}}\int_{-\infty}^{\infty}\frac{dt}{2\pi}\,e^{i\omega t}\,\Big\langle e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R},t)}\,e^{+i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{0},0)}\Big\rangle_T \quad , \tag{3.151}$$

where $\langle \mathcal{O} \rangle_T = \sum_i P_i \langle\, i \,|\, \mathcal{O} \,|\, i \,\rangle$ is the thermodynamic average. We define the *dynamic structure factor* (dsf),

$$S_{ll'}(\boldsymbol{q},\omega) = \int\limits_{-\infty}^{\infty} \frac{dt}{2\pi}\, e^{i\omega t} \sum_{\boldsymbol{R}} e^{-i\boldsymbol{q}\cdot\boldsymbol{R}} \left\langle e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R},t)}\, e^{+i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{0},0)} \right\rangle_T \quad . \tag{3.152}$$

3.3.2 Evaluation of $S_{ll'}(\boldsymbol{q},\omega)$

To evaluate the dsf, it is convenient to express the displacements $\boldsymbol{u}_a(\boldsymbol{R},t)$ in terms of the ladder operators, *viz.*

$$u_l^\alpha(\boldsymbol{R},t) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k},s} \left(\frac{\hbar}{2\omega_s(\boldsymbol{k})} \right)^{1/2} \mathrm{e}_{ls}^\alpha(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \left(A_s(\boldsymbol{k})\, e^{-i\omega_s(\boldsymbol{k})\,t} + A_s^\dagger(-\boldsymbol{k})\, e^{+i\omega_s(\boldsymbol{k})\,t} \right) \quad . \tag{3.153}$$

Thus,

$$\begin{aligned} i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R},t) &= \sum_{\boldsymbol{k},s} \left(X_{\boldsymbol{k},s} A_{\boldsymbol{k},s} - X^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s} \right) \\ i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{0},0) &= \sum_{\boldsymbol{k},s} \left(Y_{\boldsymbol{k},s} A_{\boldsymbol{k},s} - Y^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s} \right) \quad , \end{aligned} \tag{3.154}$$

where

$$\begin{aligned} X_{\boldsymbol{k},s} &= \frac{i}{\sqrt{N}} \left(\frac{\hbar}{2\omega_s(\boldsymbol{k})} \right)^{1/2} \boldsymbol{q}\cdot\mathbf{e}_{ls}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{-i\omega_s(\boldsymbol{k})\,t} \\ Y_{\boldsymbol{k},s} &= \frac{i}{\sqrt{N}} \left(\frac{\hbar}{2\omega_s(\boldsymbol{k})} \right)^{1/2} \boldsymbol{q}\cdot\mathbf{e}_{l's}(\boldsymbol{k}) \quad . \end{aligned} \tag{3.155}$$

Thus, we may write

$$\left\langle e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R},t)}\, e^{+i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{0},0)} \right\rangle_T = \prod_{\boldsymbol{k},s} \left\langle \exp\!\left(X^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s} - X_{\boldsymbol{k},s} A_{\boldsymbol{k},s}\right) \exp\!\left(Y_{\boldsymbol{k},s} A_{\boldsymbol{k},s} - Y^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s}\right) \right\rangle \quad , \tag{3.156}$$

where we have invoked the fact that $\left[A_{\boldsymbol{k},s}, A^\dagger_{\boldsymbol{k}',s'}\right] = \delta_{\boldsymbol{k}\boldsymbol{k}'}\,\delta_{ss'}$. To evaluate this expression, we appeal to the Baker-Campbell-Hausdorff equality,

$$e^A\, e^B = e^{A+B}\, e^{\frac{1}{2}[A,B]} \quad , \tag{3.157}$$

valid when both A and B commute with their commutator $[A,B]$. We may then write, for each $(\boldsymbol{k},s)$ pair,

$$\begin{aligned} \exp\!\left(X^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s} - X_{\boldsymbol{k},s} A_{\boldsymbol{k},s}\right) \exp\!\left(Y_{\boldsymbol{k},s} A_{\boldsymbol{k},s} - Y^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s}\right) = \\ \exp\!\left[\tfrac{1}{2}\left(X_{\boldsymbol{k},s} Y^*_{\boldsymbol{k},s} - X^*_{\boldsymbol{k},s} Y_{\boldsymbol{k},s}\right)\right] \exp\!\left(Z_{\boldsymbol{k},s} A_{\boldsymbol{k},s} - Z^*_{\boldsymbol{k},s} A^\dagger_{\boldsymbol{k},s}\right) \quad , \end{aligned} \tag{3.158}$$

where $Z_{\boldsymbol{k},s} = Y_{\boldsymbol{k},s} - X_{\boldsymbol{k},s}$. Now consider a single harmonic oscillator with Hamiltonian $H = \hbar\omega\,(a^\dagger a + \frac{1}{2})$ and define $g(x,y) = \langle e^{xa}\,e^{ya^\dagger}\rangle_T$. Then from the cyclic property of the trace, we have[8]

$$\begin{aligned} g(x,y) &\equiv \left\langle e^{xa}\,e^{ya^\dagger}\right\rangle_T \\ &= Z^{-1}\,\mathsf{Tr}\left(e^{-\beta H}e^{xa}\,e^{ya^\dagger}\right) \\ &= Z^{-1}\,e^{xy}\,\mathsf{Tr}\left(e^{-\beta H}e^{ya^\dagger}e^{xa}\right) \\ &= Z^{-1}\,e^{xy}\,\mathsf{Tr}\left(e^{-\beta H}e^{\beta H}e^{xa}e^{-\beta H}e^{ya^\dagger}\right) \\ &= Z^{-1}\,e^{xy}\,\mathsf{Tr}\left(e^{-\beta H}e^{x\exp(-\beta\hbar\omega)\,a}e^{ya^\dagger}\right) \\ &= e^{xy}\,g\big(x\exp(-\beta\hbar\omega),y\big) \\ &= e^{xy}\,e^{x\exp(-\beta\hbar\omega)\,y}\,g\big(x\exp(-2\beta\hbar\omega),y\big) \\ &= \exp\left(\sum_{n=0}^{\infty} xy\,e^{-n\beta\hbar\omega}\right)g(0,y) \\ &= \exp\left(\frac{xy}{1-\exp(-\beta\hbar\omega)}\right) \quad , \end{aligned} \tag{3.159}$$

since $g(0,y) = \langle e^{ya^\dagger}\rangle_T = 1$. We therefore find

$$\begin{aligned} \left\langle\,\exp\big(Z_{\boldsymbol{k},s}A_{\boldsymbol{k},s} - Z^*_{\boldsymbol{k},s}A^\dagger_{\boldsymbol{k},s}\big)\right\rangle &= \exp\big(\tfrac{1}{2}Z^*_{\boldsymbol{k},s}Z_{\boldsymbol{k},s}\big)\left\langle e^{Z_{\boldsymbol{k},s}A_{\boldsymbol{k},s}}\,e^{-Z^*_{\boldsymbol{k},s}A^\dagger_{\boldsymbol{k},s}}\right\rangle \\ &= \exp\Big(-\big(n_{\boldsymbol{k},s}+\tfrac{1}{2}\big)Z^*_{\boldsymbol{k},s}Z_{\boldsymbol{k},s}\Big) \quad , \end{aligned} \tag{3.160}$$

where

$$n_{\boldsymbol{k},s} = \frac{1}{e^{\beta\hbar\omega_s(\boldsymbol{k})}-1} \tag{3.161}$$

is the Bose function. Finally, we have

$$\begin{aligned} \left\langle e^{-i\boldsymbol{q}\cdot\boldsymbol{u}_l(\boldsymbol{R},t)}\,e^{+i\boldsymbol{q}\cdot\boldsymbol{u}_{l'}(\boldsymbol{0},0)}\right\rangle_T &= \exp\left\{-\sum_{\boldsymbol{k},s}\big(|X_{\boldsymbol{k},s}|^2+|Y_{\boldsymbol{k},s}|^2\big)\big(n_s(\boldsymbol{k})+\tfrac{1}{2}\big)\right\} \\ &\quad\times\exp\left\{\sum_{\boldsymbol{k},s}\Big[X_{\boldsymbol{k},s}Y^*_{\boldsymbol{k},s}\,\big(n_s(\boldsymbol{k})+1\big)+X^*_{\boldsymbol{k},s}Y_{\boldsymbol{k},s}\,n_s(\boldsymbol{k})\Big]\right\} \quad . \end{aligned} \tag{3.162}$$

3.3.3 Dynamic structure factor for Bravais lattices

For the case of Bravais lattices, we have $r = 1$ and

$$S(\boldsymbol{q},\omega) = e^{-2W(\boldsymbol{q})}\int_{-\infty}^{\infty}\frac{dt}{2\pi}\,\sum_{\boldsymbol{R}}e^{-i\boldsymbol{q}\cdot\boldsymbol{R}}\,e^{i\omega t}\,e^{\Gamma(\boldsymbol{q},\boldsymbol{R},t)} \quad , \tag{3.163}$$

[8]See N. D. Mermin, *J. Math. Phys.* **7**, 1038 (1966).

where

$$W(\boldsymbol{q}) = \tfrac{1}{2}\Omega \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \sum_{s=1}^{d} \hbar \omega_s^{-1}(\boldsymbol{k}) \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 \operatorname{ctnh}\left(\frac{\hbar \omega_s(\boldsymbol{k})}{2k_{\rm B}T} \right) \tag{3.164}$$

and

$$\Gamma(\boldsymbol{q}, \boldsymbol{R}, t) = \tfrac{1}{2}\Omega \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \sum_{s=1}^{d} \frac{\hbar}{\omega_s(\boldsymbol{k})} \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 \left\{ \big(n_s(\boldsymbol{k}) + 1 \big) e^{i\boldsymbol{k}\cdot\boldsymbol{R}} e^{-i\omega_s(\boldsymbol{k})\, t} + n_s(\boldsymbol{k})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{i\omega_s(\boldsymbol{k})\, t} \right\} \quad . \tag{3.165}$$

Expanding $e^{\Gamma} = 1 + \Gamma + \frac{1}{2}\Gamma^2 + \ldots$ in a power series, we have

$$S(\boldsymbol{q}, \omega) = e^{-2W(\boldsymbol{q})} \left\{ \overbrace{\hat{\Omega}\, \delta(\omega) \sum_{\boldsymbol{G}} \delta(\boldsymbol{q} - \boldsymbol{G})}^{\text{zero phonons}} + \tfrac{1}{2} \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \sum_{s=1}^{d} \frac{\hbar}{\omega_s(\boldsymbol{k})} \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 \times \right. \tag{3.166}$$

$$\left. \left[\overbrace{n_s(\boldsymbol{k})\, \delta\big(\omega + \omega_s(\boldsymbol{k})\big) \sum_{\boldsymbol{G}} \delta(\boldsymbol{q} + \boldsymbol{k} - \boldsymbol{G})}^{\text{single phonon absorption}} + \overbrace{\big(n_s(\boldsymbol{k}) + 1 \big)\, \delta\big(\omega - \omega_s(\boldsymbol{k})\big) \sum_{\boldsymbol{G}} \delta(\boldsymbol{q} - \boldsymbol{k} - \boldsymbol{G})}^{\text{single phonon emission}} \right] + \ldots \right\} \quad .$$

Here we have labeled the terms corresponding to zero phonon processes, in which the entire lattice recoils elastically, and single phonon absorption and emission processes. The ellipses contain terms corresponding to multiphonon processes. The fact that processes in which a phonon is created (emitted) are proportional to $n_s(\boldsymbol{k}) + 1$, while processes in which a phonon is destroyed (absorbed) are proportional to $n_s(\boldsymbol{k})$ is a consequence of detailed balance. Satisfying the Dirac delta functions for the single phonon processes, we may write

$$S(\boldsymbol{q}, \omega) = e^{-2W(\boldsymbol{q})} \left\{ \overbrace{\hat{\Omega}\, \delta(\omega) \sum_{\boldsymbol{G}} \delta(\boldsymbol{q} - \boldsymbol{G})}^{\text{zero phonons}} + \sum_{s=1}^{d} \frac{\hbar}{2\omega_s(\boldsymbol{q})} \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{q}) \right|^2 \left[\overbrace{n_s(\boldsymbol{q})\, \delta\big(\omega + \omega_s(\boldsymbol{q})\big)}^{\text{single phonon absorption}} \right. \right.$$

$$\left. \left. + \overbrace{\big(n_s(\boldsymbol{q}) + 1 \big)\, \delta\big(\omega - \omega_s(\boldsymbol{k})\big)}^{\text{single phonon emission}} \right] + \ldots \right\} \quad . \tag{3.167}$$

3.3.4 Debye-Waller Factor

The term $e^{-2W(\boldsymbol{q})}$ is called the Debye-Waller factor. Note that

$$2W(\boldsymbol{q}) = \big\langle \big(\boldsymbol{q} \cdot \boldsymbol{u}(\boldsymbol{R}) \big)^2 \big\rangle_T = \Omega \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \sum_{s=1}^{d} \frac{\hbar}{\omega_s(\boldsymbol{k})} \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 \operatorname{ctnh}\left(\frac{\hbar \omega_s(\boldsymbol{k})}{2k_{\rm B}T} \right) \quad . \tag{3.168}$$

We may approximate the angular integral

$$\int \frac{d\hat{\boldsymbol{k}}}{4\pi} \left| \boldsymbol{q} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 \approx \tfrac{1}{3} \boldsymbol{q}^2 \left| \hat{\boldsymbol{q}} \cdot \mathbf{e}_s(\boldsymbol{k}) \right|^2 = \frac{\boldsymbol{q}^2}{3M} \quad , \tag{3.169}$$

where M is the ionic mass. We then obtain

$$2W(\boldsymbol{q}) \approx \frac{\boldsymbol{q}^2}{3M} \int\limits_0^\infty d\omega \, g(\omega) \, \frac{\hbar}{2\omega} \, \mathrm{ctnh}\left(\frac{\hbar\omega}{2k_{\rm B}T} \right) \quad , \tag{3.170}$$

where $g(\omega)$ is the total phonon density of states. Within the Debye model in $d = 3$ dimensions,

$$g_{\rm Debye}(\omega) = \frac{9\omega^2}{\omega_{\rm D}^2} \, \Theta(\omega_{\rm D} - \omega) \quad , \tag{3.171}$$

For $k_{\rm B}T \ll \hbar\omega_{\rm D}$, we find

$$2W(\boldsymbol{q}) = \frac{3\hbar\boldsymbol{q}^2}{4M\omega_{\rm D}} \left\{ 1 + \frac{2\pi^2}{3} \left(\frac{k_{\rm B}T}{\hbar\omega_{\rm D}} \right)^2 + \ldots \right\} \quad , \tag{3.172}$$

while for $k_{\rm B}T \gg \hbar\omega_{\rm D}$ we obtain

$$2W(\boldsymbol{q}) = \frac{3\hbar\boldsymbol{q}^2}{M\omega_{\rm D}} \cdot \frac{k_{\rm B}T}{\hbar\omega_{\rm D}} \quad . \tag{3.173}$$

We see that $W(\boldsymbol{q})$ increases linearly with T, and that as $T \to 0$ it approaches a constant, given by $W_{T=0}(\boldsymbol{q}) = 3\hbar\boldsymbol{q}^2/8M\omega_{\rm D}$. The $\boldsymbol{q}$-dependence has the effect of reducing the intensity of large $|\boldsymbol{q}|$ processes relative to small $|\boldsymbol{q}|$ processes. One noteworthy feature is that finite temperature fluctuations *do not* smooth out the Bragg peaks in $\sum_{\boldsymbol{G}} \delta(\boldsymbol{q} - \boldsymbol{G})$. Rather, the Bragg peaks at each reciprocal lattice vector $\boldsymbol{G}$ are simply reduced in intensity by the Debye-Waller factor $e^{-2W(\boldsymbol{G})}$. Note that $W(\boldsymbol{q})$ does not vanish at $T = 0$, due to *quantum fluctuations* of the ionic positions. In a one-dimensional lattice, these fluctuations are strong enough to melt the lattice and destroy long-ranged positional order.

3.3.5 The Mössbauer effect

Suppose a stationary ion (or atom) of mass M radiates and decays from an excited state with energy $E = E_1$ to its ground state at $E = E_0$. A photon of energy $\varepsilon = h\nu$ and momentum $p = h\nu/c$ is emitted in the process. This results in a recoil of the ion with energy $R = (h\nu/c)^2/2m$. Thus, energy conservation requires

$$E_1 = E_0 + h\nu + R \quad \Rightarrow \quad h\nu = \Delta E - \frac{(h\nu)^2}{2Mc^2} \simeq \Delta E - \frac{(\Delta E)^2}{2Mc^2} \quad , \tag{3.174}$$

where $\Delta E = E_1 - E_0$, assuming $Mc^2 \gg \Delta E$. In ^{57}Fe, for example, we have $\Delta E = 14.4\,\text{keV}$ and $Mc^2 = 53.3\,\text{GeV}$, and the recoil energy is $R \approx 1.96\,\text{meV}$. If the photon is reabsorbed by another ^{57}Fe nucleus in its ground state, then the energy change is

$$\Delta E' = \Delta E - \frac{(h\nu)^2}{2Mc^2} \simeq \Delta E - \frac{(\Delta E)^2}{Mc^2} \quad , \tag{3.175}$$

and we have $\Delta E - \Delta E' = 2R = 3.91\,\text{meV}$. This is small, but still much greater than the natural linewidth of the transition, which is $\Gamma = 4.6 \times 10^{-9}\,\text{eV}$. So the resonant absorption should never happen. Except that it is indeed observed, as first shown by Mössbauer in 1958 in an experiment for which he won the Nobel Prize in Physics in 1961[9]. The reason is that in a crystal there are zero phonon processes in which the entire lattice recoils, hence $R = (h\nu)^2/2NMc^2$, where N is the number of unit cells of the crystal, which is thermodynamically large, *i.e.* $R = 0$ for all intents and purposes. In addition, by using a moving source, the Doppler shift may be used to probe the structure of the absorption line. The Doppler shifted frequency is

$$\omega' = \frac{\omega - \boldsymbol{v}\cdot\boldsymbol{k}}{\sqrt{1-\frac{v^2}{c^2}}} = \omega - \boldsymbol{v}\cdot\boldsymbol{k} + \mathcal{O}(v^2/c^2) \quad . \tag{3.176}$$

If we take $\boldsymbol{v} \parallel \boldsymbol{k}$ then we have $\delta\omega = vk$ and $\hbar\,\delta\omega = (v/c)\cdot\hbar ck \approx (v/c)\Delta E$. Setting $\hbar\,\delta\omega = \Gamma$, we obtain $v = 0.01\,\text{cm/s}$ – a very small velocity compared with c indeed! Since the Debye-Waller factor involves the ratio $\hbar\boldsymbol{q}^2/2M\omega_{\text{D}} = (h\nu)^2/2Mc^2\hbar\omega_{\text{D}}$, where the Debye energy $\hbar\omega_{\text{D}}$ is on the order of millivolts, only low energy atomic γ-transitions yield observable Mössbauer effects.

[9] Mössbauer used a ^{191}Os source and a ^{191}Ir absorber.

Chapter 4

Electronic Band Structure of Crystals

4.1 Energy Bands in Solids

4.1.1 Bloch's theorem

The Hamiltonian for an electron in a crystal is

$$H = -\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2 + V(\boldsymbol{r}) \quad , \tag{4.1}$$

where $V(\boldsymbol{r}) = V(\boldsymbol{r}+\boldsymbol{R})$ for all $\boldsymbol{R} \in \mathcal{L}$, where $\mathcal{L}$ is the direct Bravais lattice underlying the crystal structure. The potential $V(\boldsymbol{r})$ describes the crystalline potential due to the ions, plus the average (Hartree) potential of the other electrons. The lattice translation operator is $t(\boldsymbol{R}) = \exp(i\boldsymbol{R}\cdot\boldsymbol{p}/\hbar) = \exp(\boldsymbol{R}\cdot\boldsymbol{\nabla})$. Acting on any function of $\boldsymbol{r}$, we have

$$t(\boldsymbol{R})\,f(\boldsymbol{r}) = f(\boldsymbol{r}+\boldsymbol{R}) \quad . \tag{4.2}$$

Note that lattice translations are unitary, *i.e.* $t^\dagger(\boldsymbol{R}) = t^{-1}(\boldsymbol{R}) = t(-\boldsymbol{R})$, and they satisfy the composition rule $t(\boldsymbol{R}_1)\,t(\boldsymbol{R}_2) = t(\boldsymbol{R}_1+\boldsymbol{R}_2)$. Since $\big[t(\boldsymbol{R}), H\big] = 0$ for all Bravais lattice vectors $\boldsymbol{R}$, the Hamiltonian H and all lattice translations $t(\boldsymbol{R})$ may be simultaneously diagonalized. Let $\psi(\boldsymbol{r})$ be such a common eigenfunction. Since $t(\boldsymbol{R})$ is unitary, its eigenvalue must be a phase $\exp(i\theta_{\boldsymbol{R}})$, and as a consequence of the composition rule, we must have $\theta_{\boldsymbol{R}_1+\boldsymbol{R}_2} = \theta_{\boldsymbol{R}_1} + \theta_{\boldsymbol{R}_2}$. This requires that $\theta_{\boldsymbol{R}}$ be linear in each of the components of $\boldsymbol{R}$, *i.e.* $\theta_{\boldsymbol{R}} = \boldsymbol{k}\cdot\boldsymbol{R}$, where $\boldsymbol{k}$ is called the *wavevector*. Since $\exp(i\,\boldsymbol{G}\cdot\boldsymbol{R}) = 1$ for any $\boldsymbol{G} \in \hat{\mathcal{L}}$, *i.e.* for any reciprocal lattice vector (RLV) $\boldsymbol{G}$ in the reciprocal lattice $\widehat{\mathcal{L}}$, the wavevector is only defined modulo $\boldsymbol{G}$, which means that $\boldsymbol{k}$ may be restricted to the first Brillouin zone of the reciprocal lattice. The quantity $\hbar\boldsymbol{k}$ is called the *crystal momentum*. Unlike ordinary momentum $\boldsymbol{p}$, crystal momentum is only conserved modulo $\hbar\boldsymbol{G}$.

The energy E will in general depend on $\boldsymbol{k}$, but there may be several distinct energy eigenstates with the same value of $\boldsymbol{k}$. We label these different energy states by a discrete index n, called the

band index. Thus, eigenstates of H are labeled by the pair $(n, \boldsymbol{k})$, with

$$H\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = E_n(\boldsymbol{k})\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \qquad , \qquad t(\boldsymbol{R})\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{4.3}$$

This is the content of *Bloch's theorem*. Note that the *cell function* $u_{n\boldsymbol{k}}(\boldsymbol{r}) \equiv \psi_{n\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$ is periodic in the direct lattice, with $u_{n\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R}) = u_{n\boldsymbol{k}}(\boldsymbol{r})$. Thus, each Bloch function $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ may be written as the product of a plane wave and a cell function, *viz.*

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) \equiv \langle\,\boldsymbol{r}\,|\,n\boldsymbol{k}\,\rangle = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad . \tag{4.4}$$

We may always choose the Bloch functions to be periodic in the reciprocal lattice, *i.e.* $\psi_{n,\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{r}) = \psi_{\boldsymbol{k}}(\boldsymbol{r})$. This choice entails the condition $u_{n,\boldsymbol{k}+\boldsymbol{G}}(\boldsymbol{r}) = u_{n\boldsymbol{k}}(\boldsymbol{r})\,e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}$. However, there is no guarantee that $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ is continuous as a function of $\boldsymbol{k} \in \mathbb{T}^d$. As an example, consider the one-dimensional Bloch function $\psi_{nk}(x) = L^{-1/2}\,e^{i(G_n+k)x}$, where n labels the reciprocal lattice vector $G_n = 2\pi n/a$. If $\psi_{nk}(x)$ is taken to be continuous as a function of k, then clearly $\psi_{n,k+G_m}(x) = \psi_{n,k}(x)\,e^{iG_m x} = \psi_{n+m,k}(x) \neq \psi_{nk}(x)$.

The Bloch states, being eigenstates of a Hermitian operator, satisfy the conditions of completeness,

$$\sum_{n,\boldsymbol{k}} |\,n\boldsymbol{k}\,\rangle\langle\,n\boldsymbol{k}\,| = 1 \quad , \tag{4.5}$$

and orthonormality,

$$\langle\,n\boldsymbol{k}\,|\,n'\boldsymbol{k}'\,\rangle = \delta_{nn'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \tag{4.6}$$

Here we have assumed quantization of $\boldsymbol{k}$ in a large box of dimensions $L_1 \times L_2 \times L_3$. Each allowed wavevector then takes the form $\boldsymbol{k} = \left(\frac{2\pi n_1}{L_1}, \frac{2\pi n_2}{L_2}, \frac{2\pi n_3}{L_3}\right)$, where $n_{1,2,3} \in \mathbb{Z}$. In the thermodynamic limit, where $L_{1,2,3} \to \infty$, we have

$$\sum_{\boldsymbol{k}} \longrightarrow N v_0 \int \frac{d^d\boldsymbol{k}}{(2\pi)^d} \quad , \tag{4.7}$$

where d is the dimension of space ($d = 3$ unless otherwise noted), v_0 is the unit cell volume in real space, *i.e.* the volume of the Wigner-Seitz (WS) cell[1], and $N = L_1 \cdots L_d/v_0$ is the number of unit cells in the system, and is assumed to be thermodynamically large. Thus, we have from Eqn. 4.5,

$$\delta(\boldsymbol{r}-\boldsymbol{r}') = N v_0 \sum_n \int \frac{d^d\boldsymbol{k}}{(2\pi)^d}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r})\,\psi^*_{n\boldsymbol{k}}(\boldsymbol{r}') \quad . \tag{4.8}$$

One can see how the above equation is true in the simple case where $u_{n\boldsymbol{k}}(\boldsymbol{r}) = 1$ and $\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = V^{-1/2}\,e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, with $V = N v_0$.

[1]Elsewhere in these notes we denote the WS cell volume by Ω and the Brillouin zone volume by $\hat{\Omega}$.

4.1.2 Schrödinger equation

The potential $V(\boldsymbol{r})$ has a discrete Fourier representation as

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \quad , \tag{4.9}$$

where the sum is over all reciprocal lattice vectors $\boldsymbol{G} \in \widehat{\mathcal{L}}$. Since $V(\boldsymbol{r})$ is a real function, we must have $V_{-\boldsymbol{G}} = V_{\boldsymbol{G}}^*$. Any Bloch function $\psi_{\boldsymbol{k}}(\boldsymbol{r})$ may also be written as a Fourier sum, *viz.*

$$\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C_{\boldsymbol{G}}(\boldsymbol{k})\, e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}} \quad . \tag{4.10}$$

If we choose the Bloch functions to be periodic in the reciprocal lattice, then $C_{\boldsymbol{G}}(\boldsymbol{k}) = C(\boldsymbol{G}+\boldsymbol{k})$ is a function of $\boldsymbol{G}+\boldsymbol{k}$. Here, we have suppressed the band index n, and the wavevector $\boldsymbol{k} \in \hat{\Omega}$ must lie within the first Brillouin zone. The Schrödinger equation $H\psi_{\boldsymbol{k}}(\boldsymbol{r}) = E(\boldsymbol{k})\,\psi_{\boldsymbol{k}}(\boldsymbol{r})$ then takes the form

$$\frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m}\, C_{\boldsymbol{G}}(\boldsymbol{k}) + \sum_{\boldsymbol{G}'} V_{\boldsymbol{G}-\boldsymbol{G}'}\, C_{\boldsymbol{G}'}(\boldsymbol{k}) = E(\boldsymbol{k})\, C_{\boldsymbol{G}}(\boldsymbol{k}) \quad . \tag{4.11}$$

Note that we have one such equation for each wavevector $\boldsymbol{k} \in \hat{\Omega}$. This equation can be written in matrix form, as

$$\sum_{\boldsymbol{G}'} H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})\, C_{\boldsymbol{G}'}(\boldsymbol{k}) = E(\boldsymbol{k})\, C_{\boldsymbol{G}}(\boldsymbol{k}) \quad , \tag{4.12}$$

where, for each $\boldsymbol{k}$, $H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})$ is an infinite rank matrix,

$$H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}) = \frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m}\,\delta_{\boldsymbol{G}\boldsymbol{G}'} + V_{\boldsymbol{G}-\boldsymbol{G}'} \quad , \tag{4.13}$$

whose rows and columns are indexed by reciprocal lattice vectors $\boldsymbol{G}$ and $\boldsymbol{G}'$, respectively. The solutions, for any fixed value of $\boldsymbol{k}$, are then labeled by a band index n, hence

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \overbrace{\left(\sum_{\boldsymbol{G}} C_{\boldsymbol{G}}^{(n)}(\boldsymbol{k})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}}\right)}^{\text{cell function } u_{n\boldsymbol{k}}(\boldsymbol{r})} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \quad . \tag{4.14}$$

Note how the cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is explicitly periodic under direct lattice translations $\boldsymbol{r} \to \boldsymbol{r}+\boldsymbol{R}$. Note also that $u_{n\boldsymbol{k}}(\boldsymbol{r})$ is an eigenfunction of the unitarily transformed Hamiltonian

$$\begin{aligned} H(\boldsymbol{k}) \equiv e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\, H\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} &= \frac{(\boldsymbol{p}+\hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r}) \\ H(\boldsymbol{k})\, u_{n\boldsymbol{k}}(\boldsymbol{r}) &= E_n(\boldsymbol{k})\, u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad . \end{aligned} \tag{4.15}$$

4.1.3 $V = 0$: empty lattice

Consider the case of $d = 1$ with $V = 0$, *i.e.* an empty lattice. We can read off the eigenvalues of $H_{GG'}$ from Eqn. 4.13: $E_{nk} = \hbar^2(G+k)^2/2m$, where the band index n identifies the reciprocal lattice vector $G = 2\pi n/a$, where a is the lattice spacing[2]. The first Brillouin zone $\hat{\Omega}$ is the region $k \in \left[-\frac{\pi}{a}\,,\,\frac{\pi}{a}\right]$. Fig. 4.1 shows how the usual ballistic dispersion $E(q) = \hbar^2 q^2/2m$ is "folded" into the first Brillouin zone by translating sections by integer multiples of the primary reciprocal lattice vector $b \equiv 2\pi/a$.

4.1.4 Perturbation theory

Let's consider the case where the potential $V(\boldsymbol{r})$ is weak. This is known as the *nearly free electron* (NFE) model. The matrix form of the Hamiltonian $H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k})$ is given by

$$H_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}) = \begin{pmatrix} \frac{\hbar^2(\boldsymbol{G}_1+\boldsymbol{k})^2}{2m} + V_0 & V_{\boldsymbol{G}_1-\boldsymbol{G}_2} & V_{\boldsymbol{G}_1-\boldsymbol{G}_3} & \cdots \\ V^*_{\boldsymbol{G}_1-\boldsymbol{G}_2} & \frac{\hbar^2(\boldsymbol{G}_2+\boldsymbol{k})^2}{2m} + V_0 & V_{\boldsymbol{G}_2-\boldsymbol{G}_3} & \cdots \\ V^*_{\boldsymbol{G}_1-\boldsymbol{G}_3} & V^*_{\boldsymbol{G}_2-\boldsymbol{G}_3} & \frac{\hbar^2(\boldsymbol{G}_3+\boldsymbol{k})^2}{2m} + V_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad . \tag{4.16}$$

Suppose we perturb in the off-diagonal elements, going to second order in $V_{\boldsymbol{G}-\boldsymbol{G}'}$. We then obtain

$$E_{\boldsymbol{G}}(\boldsymbol{k}) = E^0_{\boldsymbol{G}}(\boldsymbol{k}) + V_0 + \sum_{\boldsymbol{G}'(\neq\boldsymbol{G})} \frac{\left|V_{\boldsymbol{G}-\boldsymbol{G}'}\right|^2}{E^0_{\boldsymbol{G}}(\boldsymbol{k}) - E^0_{\boldsymbol{G}'}(\boldsymbol{k})} + \mathcal{O}(V^3) \quad , \tag{4.17}$$

where the unperturbed eigenvalues are $E^0_{\boldsymbol{G}}(\boldsymbol{k}) = \hbar^2(\boldsymbol{G}+\boldsymbol{k})^2/2m$. Note that the term with $\boldsymbol{G}' = \boldsymbol{G}$ is excluded from the sum. Here and henceforth, we shall set $V_0 \equiv 0$. The denominator in the above sum can vanish if a $\boldsymbol{G}'$ can be found such that $E^0_{\boldsymbol{G}}(\boldsymbol{k}) = E^0_{\boldsymbol{G}'}(\boldsymbol{k})$. In this case, the calculation fails, and we must use degenerate perturbation theory.

[2]Since the lattice is empty, we can use any value for a we please. The eigenspectrum will be identical, although the labeling of the eigenstates will depend on a since this defines the size of the Brillouin zone.

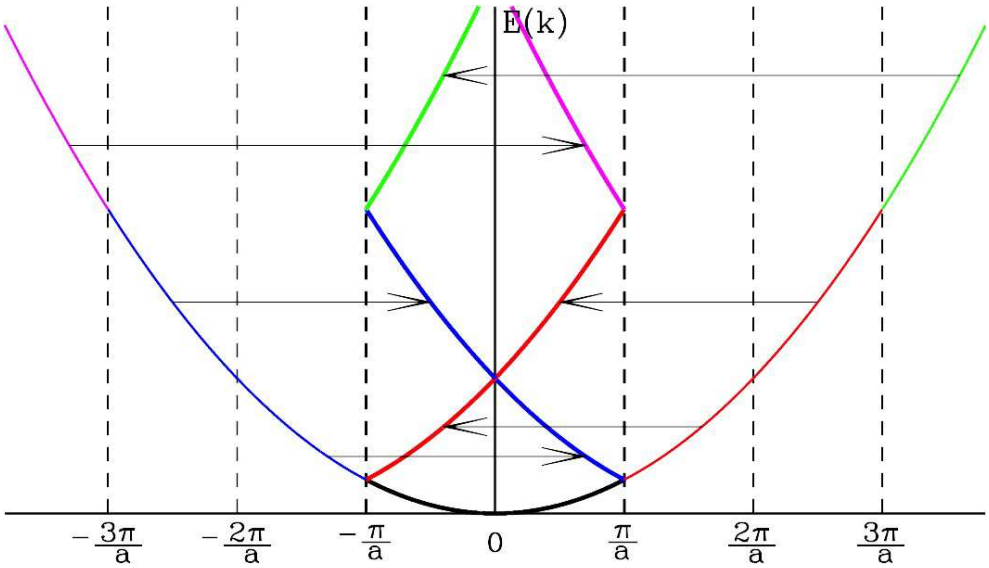


Figure 4.1: Band structure for an empty one-dimensional lattice, showing how the quadratic dispersion is "folded" from the extended zone picture into the first Brillouin zone.

Zone center (Γ), $\boldsymbol{G} = 0$

Let's first look in the vicinity of the zone center, labeled Γ, *i.e.* $\boldsymbol{k} \approx 0$. For the band associated with $\boldsymbol{G} = 0$, we have

$$\begin{aligned} E_0(\boldsymbol{k}) &= \frac{\hbar^2 \boldsymbol{k}^2}{2m} - \frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{\boldsymbol{G}^2 + 2\boldsymbol{G}\cdot\boldsymbol{k}} + \mathcal{O}(V^3) \\ &= \frac{\hbar^2 \boldsymbol{k}^2}{2m} - \frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^2} \left\{ 1 - \frac{2\,\boldsymbol{G}\cdot\boldsymbol{k}}{|\boldsymbol{G}|^2} + \frac{4(\boldsymbol{G}\cdot\boldsymbol{k})^2}{|\boldsymbol{G}|^4} + \ldots \right\} + \mathcal{O}(V^3) \quad . \end{aligned} \tag{4.18}$$

Since $V_{-\boldsymbol{G}} = V_{\boldsymbol{G}}^*$, the second term inside the curly bracket vanishes upon summation, so we have

$$E_0(\boldsymbol{k}) = \Delta + \tfrac{1}{2}\hbar^2 (m^*)^{-1}_{\mu\nu}\, k^\mu\, k^\nu + \ldots \quad , \tag{4.19}$$

where

$$\begin{aligned} \Delta &= -\frac{2m}{\hbar^2} \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^2} \\ (m^*)^{-1}_{\mu\nu} &= \frac{1}{m}\,\delta_{\mu\nu} - \left(\frac{4\sqrt{m}}{\hbar^2}\right)^2 \sum_{\boldsymbol{G}(\neq 0)} \frac{|V_{\boldsymbol{G}}|^2}{|\boldsymbol{G}|^6}\, G^\mu\, G^\nu + \ldots \quad . \end{aligned} \tag{4.20}$$

Here Δ is the *band offset* relative to the unperturbed case, and $(m^*)^{-1}_{\mu\nu}$ are the components of the inverse *effective mass tensor*. Note that the dispersion in general is no longer isotropic. Rather,

the effective mass tensor $m^*_{\mu\nu}$ transforms according to a tensor representation of the crystallographic point group. For three-dimensional systems with cubic symmetry, m^* is a multiple of the identity, for the same reason that the inertia tensor of a cube is $I = \frac{1}{6}Ma^2\,\mathsf{diag}(1,1,1)$. But for a crystal with a tetragonal symmetry, in which one of the cubic axes is shortened or lengthened, the effective mass tensor along principal axes takes the general form $m^* = \mathsf{diag}(m_x, m_x, m_z)$, with $m_x \neq m_z$ in general.

Zone center (Γ), $\boldsymbol{G} = \pm\boldsymbol{b}_{1,2,3}$

Consider the cubic lattice with primitive direct lattice vectors $\boldsymbol{a}_j = a\,\hat{\mathbf{e}}_j$ and primitive reciprocal lattice vectors $\boldsymbol{b}_j = \frac{2\pi}{a}\,\hat{\mathbf{e}}_j$. For $\boldsymbol{k} = 0$, the six bands corresponding to $\boldsymbol{G} = \pm\boldsymbol{b}_j$ with $j \in \{1,2,3\}$ are degenerate, with $E^0_{\boldsymbol{G}} = 2\pi^2\hbar^2/ma^2$. Focusing only on these rows and columns, we obtain a 6×6 effective Hamiltonian,

$$H_{6\times6} = \begin{pmatrix} \frac{\hbar^2}{2m}(\boldsymbol{b}_1+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_1} & V_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V_{\boldsymbol{b}_1+\boldsymbol{b}_3} \\ V^*_{2\boldsymbol{b}_1} & \frac{\hbar^2}{2m}(\boldsymbol{b}_1-\boldsymbol{k})^2 & V^*_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V^*_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V^*_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_1-\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V_{\boldsymbol{b}_1+\boldsymbol{b}_2} & \frac{\hbar^2}{2m}(\boldsymbol{b}_2+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_2} & V_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V_{\boldsymbol{b}_2+\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1+\boldsymbol{b}_2} & V_{\boldsymbol{b}_1-\boldsymbol{b}_2} & V^*_{2\boldsymbol{b}_2} & \frac{\hbar^2}{2m}(\boldsymbol{b}_2-\boldsymbol{k})^2 & V^*_{\boldsymbol{b}_2+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2-\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V_{\boldsymbol{b}_2+\boldsymbol{b}_3} & \frac{\hbar^2}{2m}(\boldsymbol{b}_3+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_3} \\ V^*_{\boldsymbol{b}_1+\boldsymbol{b}_3} & V_{\boldsymbol{b}_1-\boldsymbol{b}_3} & V^*_{\boldsymbol{b}_2+\boldsymbol{b}_3} & V_{\boldsymbol{b}_2-\boldsymbol{b}_3} & V^*_{2\boldsymbol{b}_3} & \frac{\hbar^2}{2m}(\boldsymbol{b}_3-\boldsymbol{k})^2 \end{pmatrix} . \tag{4.21}$$

To simplify matters, suppose that the only significant Fourier components $V_{\boldsymbol{G}}$ are those with $\boldsymbol{G} = \pm 2\boldsymbol{b}_j$. In this case, the above 6×6 matrix becomes block diagonal, *i.e.* a direct sum of 2×2 blocks, each of which resembles

$$H_{2\times2}(\Gamma) = \begin{pmatrix} \frac{\hbar^2}{2m}(\boldsymbol{b}_j+\boldsymbol{k})^2 & V_{2\boldsymbol{b}_j} \\ V^*_{2\boldsymbol{b}_j} & \frac{\hbar^2}{2m}(\boldsymbol{b}_j-\boldsymbol{k})^2 \end{pmatrix} . \tag{4.22}$$

Diagonalizing, we obtain

$$E_{j,\pm}(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{b}_j^2}{2m} + \frac{\hbar^2\boldsymbol{k}^2}{2m} \pm \sqrt{\left(\frac{\hbar^2}{m}\,\boldsymbol{b}_j\cdot\boldsymbol{k}\right)^2 + |V_{2\boldsymbol{b}_j}|^2} \quad . \tag{4.23}$$

Assuming cubic symmetry with $V_{\boldsymbol{b}_1} = V_{\boldsymbol{b}_2} = V_{\boldsymbol{b}_3} = V$, we obtain six bands,

$$E_{j,\pm}(\boldsymbol{k}) = \frac{2\pi^2\hbar^2}{ma^2} + \frac{\hbar^2\boldsymbol{k}^2}{2m} \pm \sqrt{\left(\frac{2\pi\hbar^2}{ma}\,k_j\right)^2 + |V|^2} \quad . \tag{4.24}$$

The band gap at $\boldsymbol{k} = 0$ is then $2\,|V|$.

Zone edge (X), $\boldsymbol{G} = 0$

Consider now the case $\boldsymbol{k} = \frac{1}{2}\boldsymbol{b} + \boldsymbol{q}$ with $|\boldsymbol{q}a| \ll 1$, and the band $\boldsymbol{G} = 0$. This state is nearly degenerate with one in the band with $\boldsymbol{G} = -\boldsymbol{b}$. Isolating these contributions to $H_{\boldsymbol{G}\boldsymbol{G}'}$, we obtain the 2×2 matrix

$$H_{2\times 2}(\mathrm{X}) = \begin{pmatrix} \frac{\hbar^2}{2m}\big(\frac{1}{2}\boldsymbol{b} + \boldsymbol{q}\big)^2 & V_{-\boldsymbol{b}} \\ V_{\boldsymbol{b}} & \frac{\hbar^2}{2m}\big(-\frac{1}{2}\boldsymbol{b} + \boldsymbol{q}\big)^2 \end{pmatrix} \quad , \tag{4.25}$$

with dispersion

$$E_\pm(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{b}^2}{8m} + \frac{\hbar^2\boldsymbol{q}^2}{2m} \pm \sqrt{\left(\frac{\hbar^2}{2m}\,\boldsymbol{b}\cdot\boldsymbol{q}\right)^2 + |V_{\boldsymbol{b}}|^2} \quad . \tag{4.26}$$

The band gap is again $2\,|V_{\boldsymbol{b}}|$.

4.1.5 Solvable model : one-dimensional Dirac comb

Consider the one-dimensional periodic potential,

$$V(x) = -W_0 \sum_{n=-\infty}^{\infty} \delta(x - na) \tag{4.27}$$

with $W_0 > 0$. Define $W_0 \equiv \hbar^2/2m\sigma$, where σ is the *scattering length*. The Hamiltonian is then

$$H = -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m\sigma}\sum_n \delta(x - na) \quad . \tag{4.28}$$

The eigenstates of H must satisfy Bloch's theorem: $\psi_{nk}(x+a) = e^{ika}\,\psi_{nk}(x)$. Thus, we may write

$$\begin{aligned} x \in [-a\,,\,0] \quad &: \quad \psi_{nk}(x) = A\,e^{iqx} + B\,e^{-iqx} \\ x \in [0\,,\,+a] \quad &: \quad \psi_{nk}(x) = e^{ika}\,\psi_{nk}(x-a) \\ &\qquad\quad = A\,e^{i(k-q)a}\,e^{iqx} + B\,e^{i(k+q)a}\,e^{-iqx} \quad . \end{aligned} \tag{4.29}$$

Continuity at $x = 0$ requires $\psi_{nk}(0^-) = \psi_{nk}(0^+)$, or

$$A + B = A\,e^{i(k-q)a} + B\,e^{i(k+q)a} \quad . \tag{4.30}$$

A second equation follows from integrating the Schrödinger equation from $x = 0^-$ to $x = 0^+$:

$$\begin{aligned} \int_{0^-}^{0^+} dx\; H\,\psi_{nk}(x) &= \int_{0^-}^{0^+} dx\,\left\{-\frac{\hbar^2}{2m}\,\frac{d^2\psi_{nk}}{dx^2} - \frac{\hbar^2}{2m\sigma}\,\psi_{nk}(x)\,\delta(x)\right\} \\ &= \frac{\hbar^2}{2m}\Big[\psi'_{nk}(0^-) - \psi'_{nk}(0^+)\Big] - \frac{\hbar^2}{2m\sigma}\,\psi_{nk}(0) \quad . \end{aligned} \tag{4.31}$$

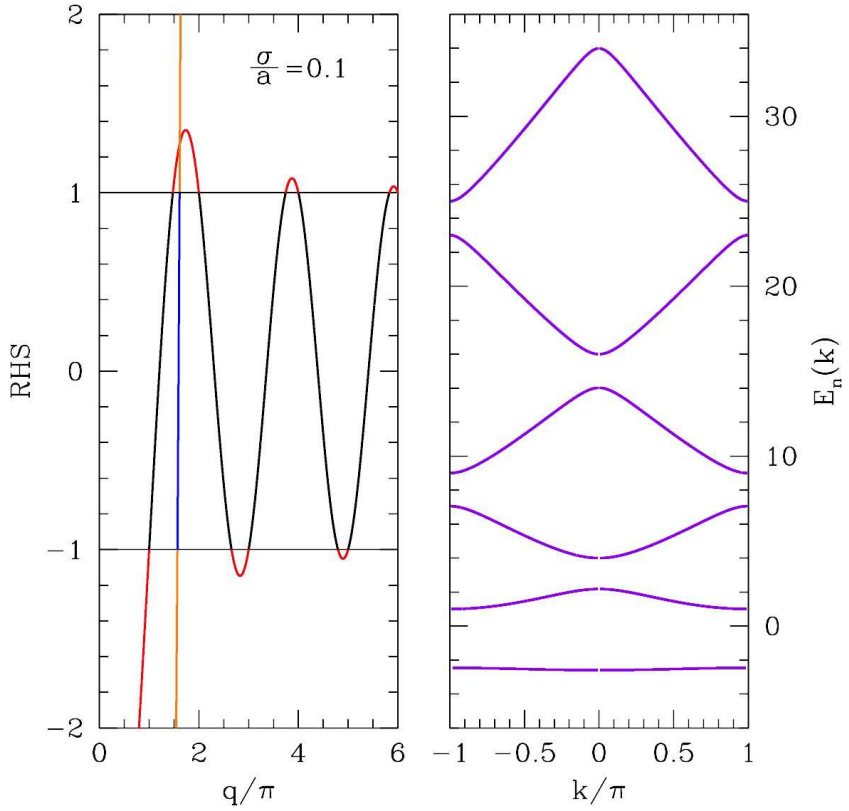


Figure 4.2: Left: Plots of the RHS of Eqns. 4.35 (black and red) and 4.36 (blue and orange) for the Dirac comb potential with scattering length $\sigma = 0.1\,a$. Allowed solutions (black and blue portions) must satisfy $\mathrm{RHS} \in [-1, 1]$. Right: Corresponding energy band structure.

Since $H\psi_{nk}(x) = E_n(k)\psi_{nk}(x)$, the LHS of the above equation is infinitesimal. Thus,

$$\psi'_{nk}(0^-) - \psi'_{nk}(0^+) = \frac{1}{\sigma}\,\psi_{nk}(0) \quad , \tag{4.32}$$

or

$$A - B - A\,e^{i(k-q)a} + Be^{i(k+q)a} = \frac{A+B}{iq\sigma} \quad . \tag{4.33}$$

The two independent equations we have derived can be combined in the form

$$\begin{pmatrix} e^{i(k-q)a} - 1 & e^{i(k+q)a} - 1 \\ e^{i(k-q)a} - 1 - \frac{i}{q\sigma} & 1 - e^{i(k+q)a} - \frac{i}{q\sigma} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = 0 \quad . \tag{4.34}$$

In order that the solution be nontrivial, we set the determinant to zero, which yields the condition

$$\cos(ka) = \cos(qa) - \frac{a}{2\sigma} \cdot \frac{\sin(qa)}{qa} \quad . \tag{4.35}$$

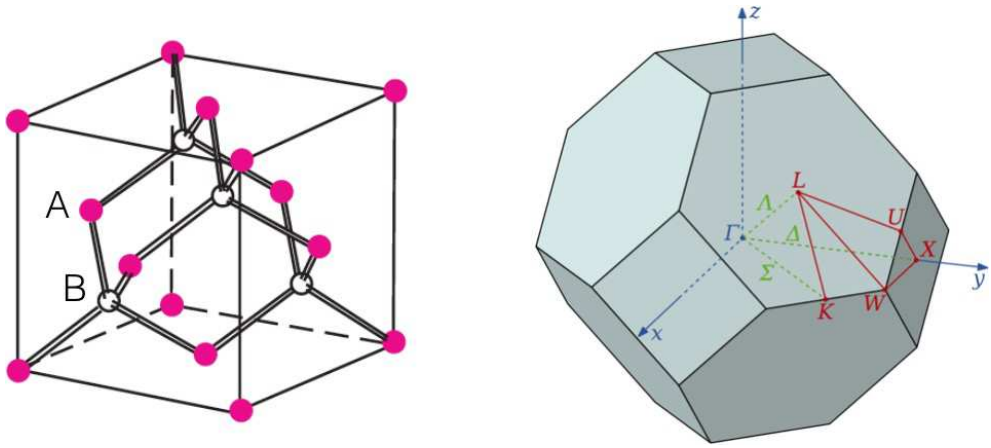


Figure 4.3: Left: The zincblende structure consists of two interpenetrating fcc lattices. Right: First Brillouin zone for the fcc lattice, with high symmetry points identified. From Wikipedia.

This is to be regarded as an equation for $q(k)$, parameterized by the dimensionless quantity σ/a. The energy eigenvalue is $E_n(k) = \hbar^2 q^2/2m$. If we set $q \equiv iQ$, the above equation becomes

$$\cos(ka) = \cosh(Qa) - \frac{a}{2\sigma} \cdot \frac{\sinh(Qa)}{Qa} \quad . \tag{4.36}$$

Here we solve for $Q(k)$, and the energy eigenvalue is $E_n(k) = -\hbar^2 Q^2/2m$. In each case, there is a discrete infinity of solutions indexed by the band index n. Results for the case $\sigma = 0.1\,a$ are shown in Fig. 4.2. In the limit $\sigma \to 0$, the solutions to Eqn. 4.35 are $q = k + \frac{2\pi n}{a}$, and we recover the free electron bands in the reduced zone scheme. When $\sigma = 0$, the only solution to Eqn. 4.36 is $Q = 0$ for the case $k = 0$.

As $q \to \infty$, the second term on the RHS of Eqn. 4.35 becomes small, and the solution for the $n^{\rm th}$ band ($n \in \mathbb{Z}_+$) tends to $q = k + 2\pi n/a$. The band gaps at $k = 0$ and $k = \pi$ become smaller and smaller with increasing band index. Clearly Eqn. 4.36 has no solutions for sufficiently large Q, since the RHS increases exponentially. Note that there is one band in the right panel of Fig. 4.2 with negative energy. This is because we have taken the potential $V(x)$ to be attractive. Recall that the potential $V(x) = -W_0\,\delta(x)$ has a single bound state $\psi_0(x) = \frac{1}{2\sqrt{\sigma}}\,e^{-|x|/2\sigma}$, again with $\sigma \equiv \hbar^2/2mW_0$. For the Dirac comb, the bound states in different unit cells overlap, which leads to dispersion. If $W_0 < 0$, the potential is purely repulsive, and all energy eigenvalues are positive. (There is no solution to Eqn. 4.36 when $\sigma < 0$.)

4.1.6 Diamond lattice bands

In dimensions $d > 1$, the essential physics is similar to what was discussed in the case of the NFE model, but the labeling of the bands and the wavevectors is more complicated than in the $d = 1$ case. Consider the zincblende structure depicted in the left panel of Fig. 4.3. Zincblende

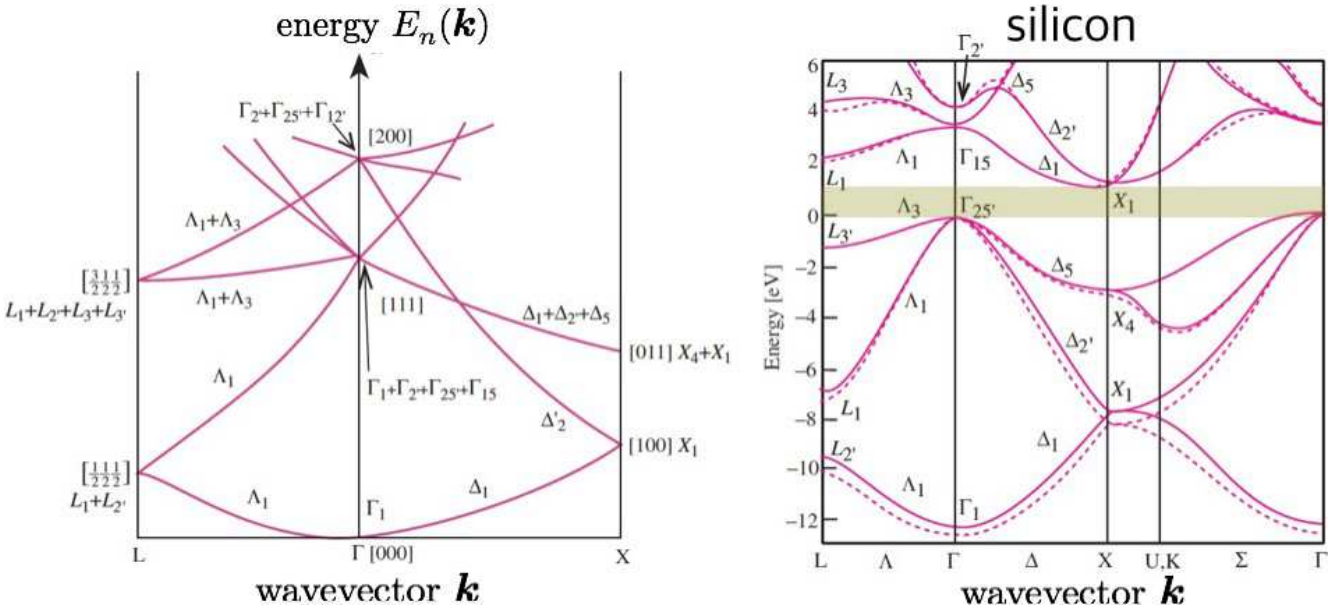


Figure 4.4: Left: Empty lattice s-orbital bands for the diamond structure. Right: Electronic band structure of Si based on local (dashed) and nonlocal (solid) pseudopotential calculations. The shaded region contains no electronic eigenstates, and reflects a (indirect) *band gap*. From ch. 2 of P. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, 1996).

consists of two interpenetrating face centered cubic lattices (labeled A and B in the figure), and is commonly found in nature (*e.g.*, in GaAs, InP, ZnTe, ZnS, HgTe, CdTe, *etc.*). Diamond is a homonuclear form of zincblende in which the ions on the two fcc sublattices are identical; the most familiar examples are C (carbon diamond) and Si. For both zincblende and diamond, the underlying Bravais lattice is fcc; the first Brillouin zone of the fcc lattice is depicted in the right panel of Fig. 4.3.

The left panel of Fig. 4.4 depicts the empty lattice (free electron) energy bands for the diamond structure along linear segments $\mathrm{L}\Gamma$ and $\Gamma\mathrm{X}$ (see Fig. 4.3 for the letter labels of high symmetry points). Energy levels at high symmetry points are labeled by reciprocal lattice vectors (square brackets) in the extended zone scheme, all in units of $2\pi/a$, where a is the size of the unit cube in Fig. 4.3. The other labels denote group representations under which the electronic eigenstates transform. In the extended zone scheme, the dispersion is $E(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$, so all the branches of the dispersion in the reduced zone scheme correspond to displacements of sections of this paraboloid by RLVs.

The right panel of Fig. 4.4 depicts the energy bands of crystalline silicon (Si), which has the diamond lattice structure. Notice how the lowest L-point energy levels are no longer degenerate. A gap has opened, as we saw in our analysis of the NFE model. Indeed, between $E = 0$ and $E = 1.12\,\mathrm{eV}$, there are *no* electronic energy eigenstates. This is the 'band gap' of silicon. Note that the $1.12\,\mathrm{eV}$ gap is *indirect* – it is between states at the Γ and X points. If the minimum en-

material	gap (eV)	type	material	gap	type	material	gap	type
C	5.47	indirect	Si	1.14	indirect	h-BN	5.96	direct
Ge	0.67	indirect	Sn	$\lesssim 0.08$	indirect	AlN	6.28	direct
GaN	3.44	direct	InN	0.7	direct	ZnO	3.37	direct
GaAs	1.43	direct	InP	1.35	direct	ZnSe	2.7	direct
GaP	2.26	indirect	InAs	0.36	direct	ZnS	3.54	direct
GaSb	0.726	direct	InSb	0.17	direct	ZnTe	2.25	direct
CdS	2.42	direct	CdTe	1.49	direct	Cu_2S	1.2	indirect

Table 4.1: Common semiconductors and their band gaps. From Wikipedia.

ergy gap occurs between levels at the same wavevector, the gap is said to be *direct*. In intrinsic semiconductors and insulators, *transport* measurements typically can provide information on indirect gaps. *Optical* measurements, however, reveal direct gaps. The reason is that the speed of light is very large, and momentum conservation requires optical transitions to be essentially vertical in $(\boldsymbol{k}, E)$ space.

4.2 Metals and Insulators

4.2.1 Density of states

In addition to energy eigenstates being labeled by band index ν and (crystal) wavevector $\boldsymbol{k}$, they are also labeled by spin polarization $\sigma = \pm 1$ relative to some fixed axis in internal space (typically $\hat{\boldsymbol{z}}$)[3]. The component of the spin angular momentum along $\hat{\boldsymbol{z}}$ is then $S^z = \frac{1}{2}\hbar\sigma = \pm\frac{1}{2}\hbar$. Typically, $E_\nu(\boldsymbol{k}, \sigma)$ is independent of the spin polarization σ, but there are many examples where this is not the case[4].

The *density of states* (DOS) per unit energy per unit volume, $g(\varepsilon)$, is given by

$$g(\varepsilon) = \frac{1}{V}{\sum_{\nu,\boldsymbol{k},\sigma}}' \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \overset{V\to\infty}{=} \sum_\nu \sum_\sigma \int\limits_{\hat{\Omega}} \frac{d^dk}{(2\pi)^d}\, \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \quad . \tag{4.37}$$

Here we assume box quantization with $\boldsymbol{k} = \left(\frac{2\pi j_1}{L_1}, \ldots, \frac{2\pi j_d}{L_d}\right)$, where j_1 *etc.* are all integers. The volume associated with each point in $\boldsymbol{k}$ space is then $\Delta V = (2\pi/L_1)\cdots(2\pi/L_d) = (2\pi)^d/V$,

[3]Here we denote the band index as ν, to obviate confusion with the occupancy n below.
[4]If there is an external magnetic field H, for example, the energy levels will be spin polarization dependent.

which establishes the above equality in the thermodynamic limit. We can also restrict our attention to a particular band ν and spin polarization σ, and define

$$g_{\nu\sigma}(\varepsilon) = \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d}\, \delta\big(\varepsilon - E_\nu(\boldsymbol{k},\sigma)\big) \quad . \tag{4.38}$$

Finally, we may multiply by the real space unit cell volume v_0 to obtain $\overline{g}(\varepsilon) \equiv v_0\, g(\varepsilon)$, which has dimensions of inverse energy, and gives the number of levels per unit energy per unit cell.

Examples

Consider the case of a one-dimensional band with dispersion $E(k) = -2t\cos(ka)$. The density of states per unit cell is

$$\overline{g}(\varepsilon) = a\int\limits_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{dk}{2\pi}\, \delta(\varepsilon + 2t\cos ka) = \frac{1}{\pi}\left(B^2 - \varepsilon^2\right)^{-1/2} \Theta(B^2 - \varepsilon^2) \quad , \tag{4.39}$$

where $B = 2t$ is half the bandwidth. *I.e.* the allowed energies are $\varepsilon \in [-B, +B]$. Note the square root singularity in $\overline{g}_{d=1}(\varepsilon)$ at the band edges.

Now let's jump to d space dimensions, and the dispersion $E(\boldsymbol{k}) = -2t\sum_{i=1}^d \cos(k_i a)$. The DOS per unit cell is then

$$\overline{g}_d(\varepsilon) = \int\limits_{-\pi}^{\pi} \frac{d\theta_1}{2\pi} \cdots \int\limits_{-\pi}^{\pi} \frac{d\theta_d}{2\pi}\, \delta\big(\varepsilon + 2t\cos\theta_1 + \cdots + 2t\cos\theta_d\big) = \frac{1}{\pi}\int\limits_0^\infty du\, \cos(\varepsilon u)\, \big[J_0(2tu)\big]^d \quad , \tag{4.40}$$

where each $\theta_j = k_j a$, and we have invoked an integral representation of the Dirac δ-function. Here $J_0(x)$ is the ordinary Bessel function of the first kind. Since $\int_{-\infty}^{\infty} d\varepsilon\, \cos(\varepsilon u) = 2\pi\,\delta(u)$, it is easy to see that $\int_{-\infty}^{\infty} d\varepsilon\, \overline{g}(\varepsilon) = 1$, *i.e.* that the DOS is correctly normalized. For $d = 2$, the integral may be performed to yield

$$\overline{g}_2(\varepsilon) = \frac{2}{\pi^2 B}\, \mathsf{K}\Big(\sqrt{1 - (\varepsilon/B)^2}\Big)\, \Theta(B^2 - \varepsilon^2) \quad , \tag{4.41}$$

where

$$\mathsf{K}(k) = \int\limits_0^{\pi/2} \frac{d\theta}{\sqrt{1 - k^2\sin^2\theta}} \tag{4.42}$$

is the complete elliptic integral of the first kind[5], and $B = 4t$ is the half bandwidth. The function $\overline{g}_2(\varepsilon)$ has a logarithmic singularity at the band center $\varepsilon = 0$, called a *van Hove singularity*.

[5]There is an unfortunate notational variation in some sources, which write $K(m)$ in place of $\mathsf{K}(k)$, where $m = k^2$.

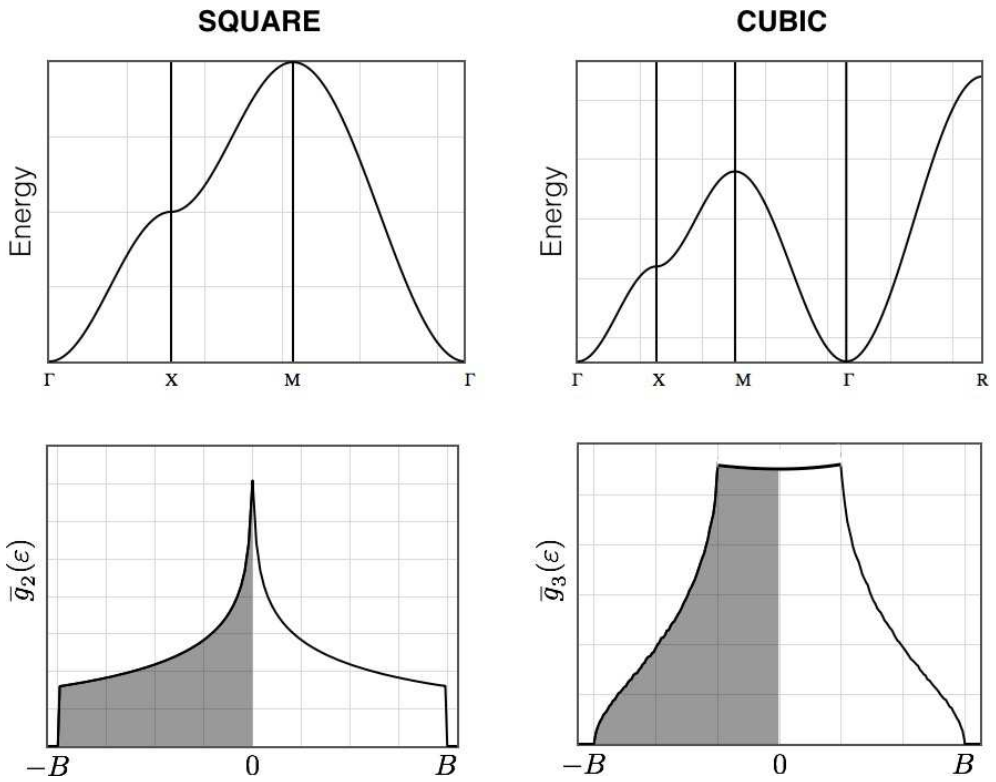


Figure 4.5: Upper left: Two-dimensional dispersion $E(k_x, k_y)$ along high symmetry lines in the 2D square lattice first Brillouin zone. Lower left: Corresponding density of states $\overline{g}_2(\varepsilon)$. Upper right: Three-dimensional dispersion $E(k_x, k_y, k_z)$ along high symmetry lines in the 2D cubic lattice first Brillouin zone. Lower right: Corresponding density of states $\overline{g}_3(\varepsilon)$. Shaded regions show occupied states for a lattice of s-orbitals with one electron per site. Figures from http://lampx.tugraz.at/∼hadley/ss1/bands/tbtable/tbtable.html .

The results for $d = 2$ and $d = 3$ are plotted in Fig. 4.5. The logarithmic van Hove singularity at $\varepsilon = 0$ is apparent in $\overline{g}_2(\varepsilon)$. The function $\overline{g}_3(\varepsilon)$ has van Hove singularities at $\varepsilon = \pm\frac{1}{3}B$, where the derivative $g_3'(\varepsilon)$ is discontinuous. In the limit $d \to \infty$, we can use the fact that $J_0(x) = 1 - \frac{1}{4}x^2 + \ldots$ to extract

$$\overline{g}_{d\gg1}(\varepsilon) = \sqrt{\frac{d}{\pi B^2}}\, e^{-d\varepsilon^2/B^2} = (4\pi dt^2)^{-1/2}\, \exp\bigl(-\varepsilon^2/4dt^2\bigr) \quad . \tag{4.43}$$

We recognize this result as the Central Limit Theorem in action. With $E(\boldsymbol{k}) = -2t\sum_{i=1}^{d}\cos\theta_i$ and θ_i uniformly distributed along $[-\pi, \pi]$, the standard deviation σ is given by

$$\sigma^2 = (2t)^2 \times d \times \langle\cos^2\theta\rangle = 2dt^2 \quad , \tag{4.44}$$

exactly as in Eqn. 4.43

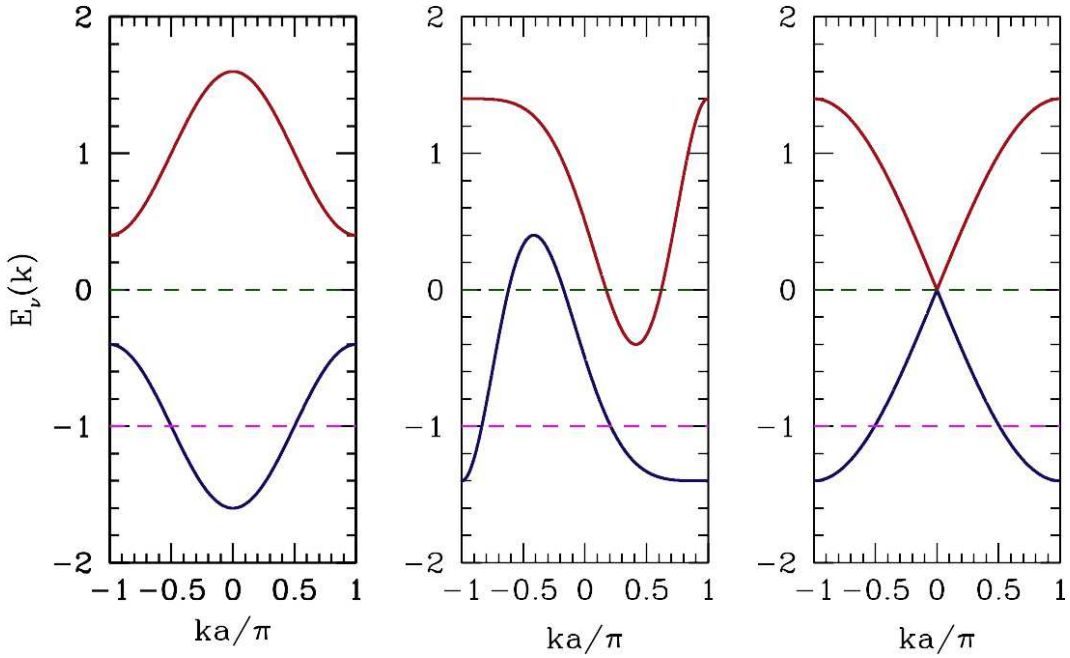


Figure 4.6: Three one-dimensional band structures. Valence bands are shown in dark blue, and conduction bands in dark red. Left: Non-overlapping bands with forbidden region $E \in [-0.4, 0.4]$. Center: At each point there is a direct gap, but the indirect gap is negative and there is no forbidden region. Right: Linear crossing leading to cusp-like band touching. In all cases, when $\varepsilon_{\rm F} = -1$ (dashed magenta line), the Fermi energy cuts through the bottom band and the system is a metal. When $\varepsilon_{\rm F} = 0$ (dashed green line), the system at the left is an insulator, with $g(\varepsilon_{\rm F}) = 0$. The system in the middle is a metal, with $g(\varepsilon_{\rm F}) > 0$ The system on the right also has a finite density of states at $\varepsilon_{\rm F}$, but in two space dimensions such a diabolical point in the dispersion, where $\varepsilon(\boldsymbol{k}) = \pm\hbar v_{\rm F}|\boldsymbol{k}|$ results in a continuously vanishing density of states $g(E)$ as $E \to \varepsilon_{\rm F} = 0$. Such a system is called a *semimetal*.

Band edge behavior

In the vicinity of a quadratic band minimum, along principal axes of the effective mass tensor, we have

$$E(\boldsymbol{k}) = \Delta + \sum_{i=1}^{d} \frac{\hbar^2 k_i^2}{2m_i^*} \quad , \tag{4.45}$$

and the density of states is

$$\begin{aligned} \overline{g}(\varepsilon) &= v_0 \int \frac{dk_1}{2\pi} \cdots \int \frac{dk_d}{2\pi}\, \delta\left(\varepsilon - \Delta - \sum_i \frac{\hbar^2 k_i^2}{2m_i^*}\right) \\ &= \frac{v_0}{2} \cdot \frac{\sqrt{2m_1^*}}{h} \cdots \frac{\sqrt{2m_d^*}}{h}\, \Omega_d\, (\varepsilon - \Delta)^{\frac{d}{2}-1} \quad , \end{aligned} \tag{4.46}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in d space dimensions. Consistent with Fig. 4.5, $g_{d=2}(\varepsilon)$ tends to a constant at the band edges, and then discontinuously drops to zero as one exits the band. In $d = 3$, $\overline{g}_{d=3}(\varepsilon)$ vanishes as $(\varepsilon - \Delta)^{1/2}$ at a band edge.

4.2.2 Fermi statistics

If we assume the electrons are noninteracting[6], the energy of the state for which the occupancy of state $|\nu \boldsymbol{k}\sigma\rangle$ is $n_{\nu\boldsymbol{k}\sigma}$ is

$$E\big[\{n_{\nu\boldsymbol{k}\sigma}\}\big] = \sum_\nu \sum_\sigma {\sum_{\boldsymbol{k}}}' E_\nu(\boldsymbol{k},\sigma)\, n_{\nu\boldsymbol{k}\sigma} \quad . \tag{4.47}$$

The Pauli exclusion principle tells us that a given electronic energy level can accommodate at zero or one fermion, which means each $n_{\nu\boldsymbol{k}\sigma}$ is either 0 or 1. At zero temperature, the N electron ground state is obtained by filling up all the energy levels starting from the bottom of the spectrum, with one electron per level, until the lowest N such levels have been filled. In $E_\nu(\boldsymbol{k},\sigma)$ is independent of σ, then there will be a twofold *Kramers degeneracy* whenever N is odd, as the last level filled can either have $\sigma = +1$ or $\sigma = -1$[7].

At finite temperature $T > 0$, the thermodynamic average of $n_{\nu\boldsymbol{k}\sigma}$ is given, within the grand canonical ensemble, by

$$\big\langle n_{\nu\boldsymbol{k}\sigma}\big\rangle = \frac{1}{\exp\Big(\frac{E_\nu(\boldsymbol{k},\sigma)-\mu}{k_{\rm B}T}\Big)+1} \equiv f\big(E_\nu(\boldsymbol{k},\sigma)-\mu\big) \quad , \tag{4.48}$$

where $f(x)$ is the *Fermi function*,

$$f(x) = \frac{1}{e^{x/k_{\rm B}T}+1} \quad . \tag{4.49}$$

The total electron number density is then

$$n(T,\mu) = \frac{N}{V} = \int\limits_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, f(\varepsilon-\mu) = \sum_\nu \sum_\sigma \int \frac{d^dk}{(2\pi)^d} \frac{1}{\exp\Big(\frac{E_\nu(\boldsymbol{k},\sigma)-\mu}{k_{\rm B}T}\Big)+1} \quad . \tag{4.50}$$

This is a Gibbs-Duhem relation, involving the three intensive quantities (n, T, μ). In principle it can be inverted to yield the chemical potential $\mu(n,T)$ as a function of number density and temperature. When $T = 0$, we write $\mu(n, T=0) \equiv \varepsilon_{\rm F}$, which is the *Fermi energy*. Since the Fermi function becomes $f(x) = \Theta(-x)$ at zero temperature, we have

$$n(\varepsilon_{\rm F}) = \int\limits_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\, g(\varepsilon) \quad . \tag{4.51}$$

[6]Other, that is, than the mean "Hartree" contribution to the potential $V(\boldsymbol{r})$.

[7]There can be additional degeneracies. For example, in $d = 1$ if, suppressing the band index, $E(k) = E(-k)$, then each level with $k \neq 0$ and $k \neq \pi/a$ is fourfold degenerate: $(k\uparrow, k\downarrow, -k\uparrow, -k\uparrow)$.

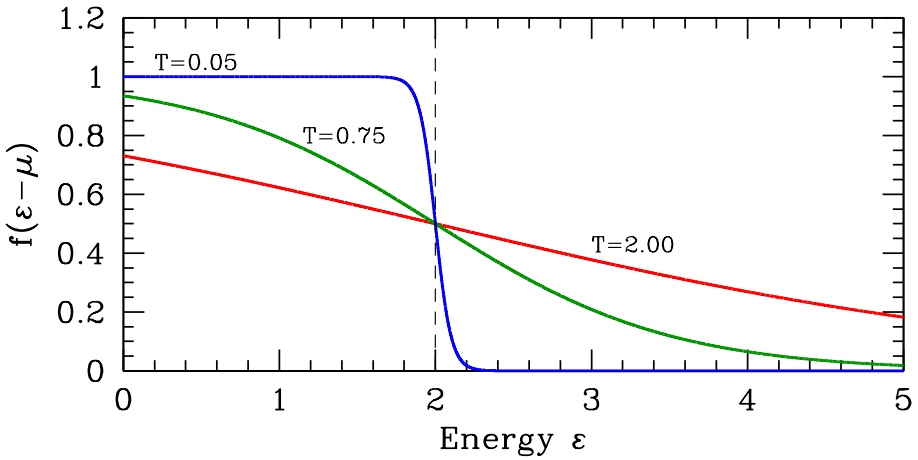


Figure 4.7: The Fermi distribution, $f(x) = \big[\exp(x/k_{\rm B}T) + 1\big]^{-1}$. Here we have set $k_{\rm B} = 1$ and taken $\mu = 2$, with $T = \frac{1}{20}$ (blue), $T = \frac{3}{4}$ (green), and $T = 2$ (red). In the $T \to 0$ limit, $f(x)$ approaches a step function $\Theta(-x)$.

This is to be inverted to obtain $\mu(n, T = 0) = \varepsilon_{\rm F}(n)$.

4.2.3 Metals and insulators at $T = 0$

At $T = 0$, the ground state is formed by filling up each single particle state $|\, n\boldsymbol{k}\sigma \,\rangle$ until the source of electrons (*i.e.* the atoms) is exhausted. Suppose there are $N_{\rm e}$ electrons in total. If there is a finite gap Δ between the $N_{\rm e}^{\rm th}$ and $(N_{\rm e}+1)^{\rm th}$ energy states, the material is an *insulator*. If the gap is zero, the material is a *metal* or possibly a *semimetal*. For a metal, $g(\varepsilon_{\rm F}) > 0$, whereas for a semimetal, $g(\varepsilon_{\rm F}) = 0$ but $g(\varepsilon) \sim |\varepsilon - \varepsilon_{\rm F}|^\alpha$ as $\varepsilon \to \varepsilon_{\rm F}$, where $\alpha > 0$.

Under periodic boundary conditions, there are N quantized wavevectors $\boldsymbol{k}$ in each Brillouin zone, where N is the number of unit cells in the crystal. Since, for a given band index n and wavevector $\boldsymbol{k}$ we can accommodate a maximum of two electrons, one with spin $\uparrow$ and the second with spin $\downarrow$, each band can accommodate a total of $2N$ electrons. Thus, if the number of electrons per cell $N_{\rm e}/N$ is not a precise multiple of two, then *necessarily* at least one of the bands will be partially filled, which means the material is a metal. Typically we only speak of valence and conduction electrons, since the core bands are all fully occupied and the high energy bands are all completely empty. Then we can define the *electron filling factor* $\nu = \widetilde{N}_{\rm e}/N$, where $\widetilde{N}_{\rm e}$ is the total number of valence plus conduction electrons. As we have just noted, if $\nu \neq 2k$ for some $k \in \mathbb{Z}$, the material is a metal.

Is the converse also the case, *i.e.* if $\nu = 2k$ is the material always an insulator? It ain't necessarily so! As the middle panel of Fig. 4.6 shows, it is at least in principle possible to have an arrangement of several partially filled bands such that the total number of electrons per site is an even integer. This is certainly a nongeneric state of affairs, but it is not completely ruled out.

4.3 Tight Binding Model

4.3.1 Bands from atomic orbitals

A crystal is a regular assembly of atoms, which are bound in the crystalline state due to the physics of electrostatics and quantum mechanics. Consider for the sake of simplicity a homonuclear Bravais lattice, *i.e.* a crystalline lattice in which there is the same type of atom at every lattice site, and in which all lattice sites are equivalent under translation. As the lattice constant a tends to infinity, the electronic energy spectrum of the crystal is the same as that of each atom, with an extensive degeneracy of N, the number of unit cells in the lattice. For finite a, the atomic orbitals on different lattice sites will overlap. Initially we will assume a Bravais lattice, but further below we shall generalize this to include the possibility of a basis.

Let $|\, n\boldsymbol{R}\,)$ denote an atomic orbital at Bravais lattice site $\boldsymbol{R}$, where $n \in \{1\mathrm{s}, 2\mathrm{s}, 2\mathrm{p}, \ldots\}$. The atomic wavefunctions[8],

$$\varphi_{n\boldsymbol{R}}(\boldsymbol{r}) = (\, \boldsymbol{r} \,|\, n\boldsymbol{R}\,) = \varphi_n(\boldsymbol{r} - \boldsymbol{R}) \quad , \tag{4.52}$$

Atomic orbitals on the same site form an orthonormal basis: $(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}\,) = \delta_{nn'}$. However, orbitals on different lattice sites are not orthogonal, and satisfy

$$(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}'\,) = \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R})\, \varphi_{n'}(\boldsymbol{r} - \boldsymbol{R}') \equiv S_{nn'}(\boldsymbol{R} - \boldsymbol{R}') \quad , \tag{4.53}$$

where $S_{nn'}(\boldsymbol{R}-\boldsymbol{R}')$ is the *overlap matrix*. Note that $S_{nn'}(0) = \delta_{nn'}$. If we expand the wavefunction $|\,\psi\,\rangle = \sum_{n,\boldsymbol{R}} C_{n\boldsymbol{R}}\,|\, n\boldsymbol{R}\,)$ in atomic orbitals, the Schrödinger equation takes the form

$$\sum_{n',\boldsymbol{R}'} \Big\{ \overbrace{(\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}'\,)}^{H_{nn'}(\boldsymbol{R}-\boldsymbol{R}')} - E\, \overbrace{(\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}'\,)}^{S_{nn'}(\boldsymbol{R}-\boldsymbol{R}')} \Big\}\, C_{n'\boldsymbol{R}'} = 0 \quad , \tag{4.54}$$

where

$$\begin{aligned} H_{nn'}(\boldsymbol{R} - \boldsymbol{R}') &= (\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}'\,) \\ &= \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R}) \left\{ -\frac{\hbar^2}{2m}\nabla^2 + V(\boldsymbol{r}) \right\} \varphi_{n'}(\boldsymbol{r} - \boldsymbol{R}') \quad , \end{aligned} \tag{4.55}$$

where $V(\boldsymbol{r}) = \sum_{\boldsymbol{R}} v(\boldsymbol{r} - \boldsymbol{R})$ is the lattice potential. Note that

$$\begin{aligned} H_{nn'}(\boldsymbol{R}) = \tfrac{1}{2}\Big(E_n^{\rm at} + E_{n'}^{\rm at}\Big)\, S_{nn'}(\boldsymbol{R}) + \tfrac{1}{2}\int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R}) \Big\{ v(\boldsymbol{r}) + v(\boldsymbol{r} - \boldsymbol{R}) \Big\} \varphi_{n'}(\boldsymbol{r}) \\ + \sum_{\substack{\boldsymbol{R}' \\ (\neq 0, \boldsymbol{R})}} \int d^d r\; \varphi_n^*(\boldsymbol{r} - \boldsymbol{R})\, v(\boldsymbol{r} - \boldsymbol{R}')\, \varphi_{n'}(\boldsymbol{r}) \quad . \end{aligned} \tag{4.56}$$

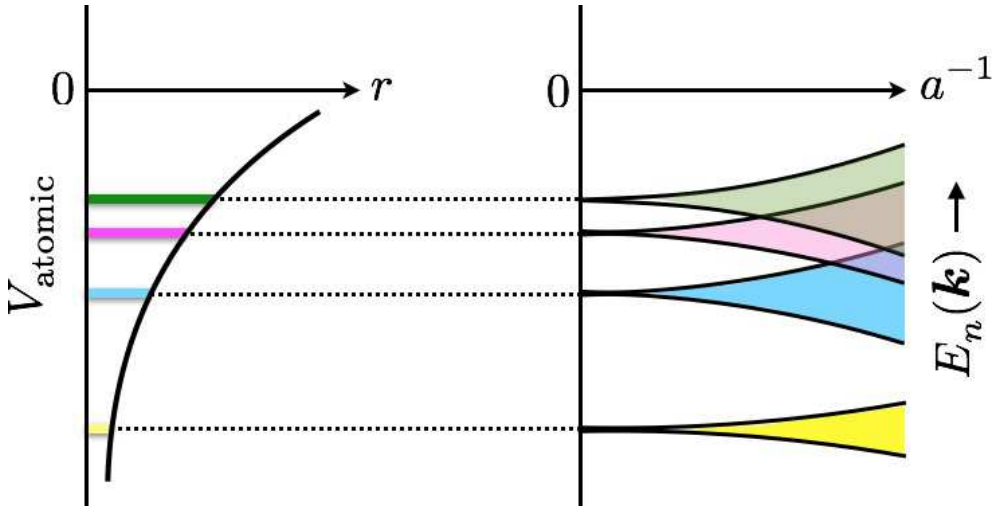


Figure 4.8: Left: Atomic energy levels. Right: Dispersion of crystalline energy bands as a function of interatomic separation.

We can simplify Eqn. 4.54 a bit by utilizing the translational invariance of the Hamiltonian and overlap matrices. We write $C_{n\boldsymbol{R}} = C_n(\boldsymbol{k})\,\exp(i\boldsymbol{k}\cdot\boldsymbol{R})$, as well as

$$\widehat{S}_{nn'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, S_{nn'}(\boldsymbol{R}) \quad , \quad \widehat{H}_{nn'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, H_{nn'}(\boldsymbol{R}) \quad . \tag{4.57}$$

The Schrödinger equation then separates for each $\boldsymbol{k}$ value, *viz.*

$$\sum_{n'} \Big\{ \widehat{H}_{nn'}(\boldsymbol{k}) - E(\boldsymbol{k})\,\widehat{S}_{nn'}(\boldsymbol{k}) \Big\}\, C_{n'}(\boldsymbol{k}) \quad . \tag{4.58}$$

Note that

$$\widehat{H}^*_{n'n}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, H^*_{n'n}(\boldsymbol{R}) = \sum_{\boldsymbol{R}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, H_{nn'}(-\boldsymbol{R}) = \widehat{H}_{nn'}(\boldsymbol{k}) \quad , \tag{4.59}$$

hence for each $\boldsymbol{k}$, the matrices $\widehat{H}_{nn'}(\boldsymbol{k})$ and $\widehat{S}_{nn'}(\boldsymbol{k})$ are Hermitian.

Suppose we ignore the overlap between different bands. We can then suppress the band index, and write[9]

$$\begin{aligned} S(\boldsymbol{R}) &= \int d^dr\, \varphi^*(\boldsymbol{r}-\boldsymbol{R})\,\varphi(\boldsymbol{r}) \\ H(\boldsymbol{R}) &= E^{\rm at}\, S(\boldsymbol{R}) + \sum_{\boldsymbol{R}'\neq 0} \int d^dr\, \varphi^*(\boldsymbol{r}-\boldsymbol{R})\, v(\boldsymbol{r}-\boldsymbol{R}')\,\varphi(\boldsymbol{r}) \quad ; \end{aligned} \tag{4.60}$$

[8] In our notation, $|\,\boldsymbol{r}\,) = |\,\boldsymbol{r}\,\rangle$, so $(\,\boldsymbol{r}\,|\,\boldsymbol{r}'\,) = \delta(\boldsymbol{r}-\boldsymbol{r}')$.

[9] The student should derive the formulae in Eqn. 4.60. In so doing, is it necessary to presume that the atomic wavefunctions are each of a definite parity?

Note that $S(0) = 1$. The tight binding dispersion is then

$$E(\boldsymbol{k}) = \frac{\widehat{H}(\boldsymbol{k})}{\widehat{S}(\boldsymbol{k})} = E^{\rm at} - \frac{\sum_{\boldsymbol{R}} t(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}}{\sum_{\boldsymbol{R}} S(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}} \quad , \tag{4.61}$$

where

$$t(\boldsymbol{R}) = E^{\rm at} S(\boldsymbol{R}) - H(\boldsymbol{R}) = -\int d^d r\, \varphi^*(\boldsymbol{r}-\boldsymbol{R}) \left\{ \sum_{\boldsymbol{R}'\neq 0} v(\boldsymbol{r}-\boldsymbol{R}') \right\} \varphi(\boldsymbol{r}) \quad . \tag{4.62}$$

Let's examine this result for a d-dimensional cubic lattice. To simplify matters, we assume that $t(\boldsymbol{R})$ and $S(\boldsymbol{R})$ are negligible beyond the nearest neighbor separation $|\boldsymbol{R}| = a$. Then

$$\begin{aligned} E(\boldsymbol{k}) &= E^{\rm at} - \frac{t(0) + 2\,t(a)\sum_{j=1}^d \cos(k_j a)}{1 + 2\,S(a)\sum_{j=1}^d \cos(k_j a)} \\ &\approx E^{\rm at} - t(0) - 2\big[t(a) - t(0)S(a)\big] \sum_{j=1}^d \cos(k_j a) + \ldots \quad , \end{aligned} \tag{4.63}$$

where we expand in the small quantity $S(a)$.

Remarks

Eqn. 4.58 says that the eigenspectrum of the crystalline Hamiltonian at each crystal momentum $\boldsymbol{k}$ is obtained by simultaneously diagonalizing the matrices $\widehat{H}(\boldsymbol{k})$ and $\widehat{S}(\boldsymbol{k})$. The simultaneous diagonalization of two real symmetric matrices is familiar from the classical mechanics of coupled oscillations. The procedure for complex Hermitian matrices follows along the same lines:

(i) To simultaneously diagonalize $\widehat{H}_{nn'}(\boldsymbol{k})$ and $\widehat{S}_{nn'}(\boldsymbol{k})$, we begin by finding a unitary matrix $U_{na}(\boldsymbol{k})$ such that

$$\sum_{n,n'} U^\dagger_{an}(\boldsymbol{k})\, \widehat{S}_{nn'}(\boldsymbol{k})\, U_{n'a'}(\boldsymbol{k}) = s_a(\boldsymbol{k})\, \delta_{aa'} \quad . \tag{4.64}$$

The eigenvalues $s_a(\boldsymbol{k})$ are all real and are presumed to be positive[10].

(ii) Next construct the Hermitian matrix

$$\widehat{L}_{bb'}(\boldsymbol{k}) = \sum_{n,n'} s_b^{-1/2}(\boldsymbol{k})\, U^\dagger_{bn}(\boldsymbol{k})\, \widehat{H}_{nn'}(\boldsymbol{k})\, U_{n'b'}(\boldsymbol{k})\, s_{b'}^{-1/2}(\boldsymbol{k}) \quad . \tag{4.65}$$

[10] In fact there are many zero eigenvalues, as we shall discuss below. Still this won't prove fatal to our development so long as we operate in some truncated Hilbert space.

This may be diagonalized by a unitary matrix $V_{ab}(\boldsymbol{k})$, *viz.*

$$\sum_{b,b'} V^\dagger_{ab}(\boldsymbol{k})\,\widehat{L}_{bb'}(\boldsymbol{k})\,V_{a'b'}(\boldsymbol{k}) = E_a(\boldsymbol{k})\,\delta_{aa'} \quad . \tag{4.66}$$

The eigenvalues are then the set $\{E_a(\boldsymbol{k})\}$.

(iii) Define the matrix

$$\Lambda_{nb}(\boldsymbol{k}) = \sum_a U_{na}(\boldsymbol{k})\,s_a^{-1/2}(\boldsymbol{k})\,V_{ab}(\boldsymbol{k})\,s_b^{1/2}(\boldsymbol{k}) \quad , \tag{4.67}$$

or, in abbreviated notation, $\Lambda = U s^{-1/2} V s^{1/2}$, suppressing the $\boldsymbol{k}$ label and defining the square matrix $s = \mathsf{diag}(s_1, s_2, \ldots\}$. Note that $\Lambda^\dagger = s^{1/2}\,V^\dagger s^{-1/2}\,U^\dagger$ but $\Lambda^{-1} = s^{-1/2}\,V^\dagger s^{1/2}\,U^\dagger$ and thus $\Lambda^\dagger \neq \Lambda^{-1}$. Nevertheless, Λ simultaneously diagonalizes $\widehat{H}$ and $\widehat{S}$:

$$\begin{aligned} \Lambda^\dagger_{an}(\boldsymbol{k})\,\widehat{S}_{nn'}(\boldsymbol{k})\,\Lambda_{n'a'}(\boldsymbol{k}) &= s_a(\boldsymbol{k})\,\delta_{aa'} \\ \Lambda^\dagger_{an}(\boldsymbol{k})\,\widehat{H}_{nn'}(\boldsymbol{k})\,\Lambda_{n'a'}(\boldsymbol{k}) &= h_a(\boldsymbol{k})\,\delta_{aa'} \quad , \end{aligned} \tag{4.68}$$

where $h_a(\boldsymbol{k}) = s_a(\boldsymbol{k})\,E_a(\boldsymbol{k})$.

Thus, the band energies are then given by

$$E_a(\boldsymbol{k}) = \frac{h_a(\boldsymbol{k})}{s_a(\boldsymbol{k})} \quad . \tag{4.69}$$

This all seems straightforward enough. However, implicit in this procedure is the assumption that the overlap matrix is nonsingular, which is clearly wrong! We know that the atomic eigenstates at any *single* lattice site must form a complete set, therefore we must be able to write

$$\varphi_n(\boldsymbol{r} - \boldsymbol{R}) = \sum_{n'} A_{nn'}(\boldsymbol{R})\,\varphi_{n'}(\boldsymbol{r}) \quad . \tag{4.70}$$

Therefore the set $|\,n\boldsymbol{R}\,)$ is massively degenerate. Fortunately, this problem is not nearly so severe as it might first appear. Recall that the atomic eigenstates consist of *bound states* of negative energy, and *scattering states* of positive energy. If we restrict our attention to a finite set of atomic bound states, the overlap matrix remains nonsingular.

4.3.2 Wannier functions

Suppose a very nice person gives us a complete set of Bloch functions $\psi_{n\boldsymbol{k}}(\boldsymbol{r})$. We can then form the linear combinations

$$W_n(\boldsymbol{r} - \boldsymbol{R}) \equiv \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\,e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})}\,u_{n\boldsymbol{k}}(\boldsymbol{r}) \quad , \tag{4.71}$$

were N is the number of unit cells and $\chi_n(\boldsymbol{k})$ is a smooth function of the wavevector $\boldsymbol{k}$ which satisfies $\chi_n(\boldsymbol{k}+\boldsymbol{G}) = \chi_n(\boldsymbol{k})$. The $\boldsymbol{k}$ sum is over all wavevectors lying within the first Brillouin zone. Writing $W_n(\boldsymbol{r}) = \langle\, \boldsymbol{r} \,|\, n\boldsymbol{R} \,\rangle$, we have

$$|\, n\boldsymbol{R} \,\rangle = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\chi_n(\boldsymbol{k})}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, |\, n\boldsymbol{k} \,\rangle \quad , \tag{4.72}$$

and the overlap matrix is

$$\langle\, n\boldsymbol{R} \,|\, n'\boldsymbol{R}' \,\rangle = \int d^d r\; W_n^*(\boldsymbol{r}-\boldsymbol{R})\, W_{n'}(\boldsymbol{r}-\boldsymbol{R}') = \delta_{nn'}\, \delta_{\boldsymbol{R}\boldsymbol{R}'} \quad . \tag{4.73}$$

The functions $W_n(\boldsymbol{r}-\boldsymbol{R})$ are called *Wannier functions*. They are linear combinations of Bloch states within a single energy band which are localized about a single Bravais lattice site or unit cell. Since the Wannier states are normalized, we have $\int d^d r \left|W_n(\boldsymbol{r}-\boldsymbol{R})\right|^2 = 1$, which means, if the falloff is the same in all symmetry-related directions of the crystal, that the envelope of $W_n(\boldsymbol{r}-\boldsymbol{R})$ must decay faster than $|\boldsymbol{r}-\boldsymbol{R}|^{-d/2}$ in d dimensions. For core ionic orbitals such as the 1s states, the atomic wavefunctions themselves are good approximations to Wannier states. Note that our freedom to choose the phase functions $\chi_n(\boldsymbol{k})$ results in many different possible definitions of the Wannier states. One desideratum we may choose to impose is to constrain the phase functions so as to minimize the expectation of $(\boldsymbol{r}-\boldsymbol{R})^2$ in each band.

Closed form expressions for Wannier functions are hard to come by, but we can obtain results for the case where the cell functions are constant, *i.e.* $u_{n\boldsymbol{k}}(\boldsymbol{r}) = (Nv_0)^{-1/d}$. Consider the cubic lattice case in $d=3$ dimensions, where $v_0 = a^3$. We then have

$$\begin{aligned} W(\boldsymbol{r}-\boldsymbol{R}) &= v_0^{1/2} \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})} \\ &= \left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_x\, e^{ik_x(x-X)}\right]\left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_y\, e^{ik_y(y-Y)}\right]\left[\frac{\sqrt{a}}{2\pi}\int\limits_{-\pi/a}^{\pi/a} dk_z\, e^{ik_z(z-Z)}\right] \\ &= \left[\frac{\sqrt{a}\,\sin\left[\frac{\pi}{a}(x-X)\right]}{\pi(x-X)}\right]\left[\frac{\sqrt{a}\,\sin\left[\frac{\pi}{a}(y-Y)\right]}{\pi(y-Y)}\right]\left[\frac{\sqrt{a}\,\sin\left[\frac{\pi}{a}(z-Z)\right]}{\pi(z-Z)}\right] \quad , \end{aligned} \tag{4.74}$$

which falls off as $|\Delta x\,\Delta y\,\Delta z|^{-1}$ along a general direction in space[11].

The Wannier states are not eigenstates of the crystal Hamiltonian. Indeed, we have

$$\begin{aligned} \langle\, n\boldsymbol{R} \,|\, H \,|\, n'\boldsymbol{R}' \,\rangle &= \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'} e^{-i\chi_n(\boldsymbol{k})}\, e^{i\chi_{n'}(\boldsymbol{k}')}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, e^{-i\boldsymbol{k}'\cdot\boldsymbol{R}'} \langle\, n\boldsymbol{k} \,|\, H \,|\, n'\boldsymbol{k}' \,\rangle \\ &= \delta_{nn'}\, v_0 \int \frac{d^dk}{(2\pi)^d}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')}\, E_n(\boldsymbol{k}) \quad , \end{aligned} \tag{4.75}$$

[11] Note that $W(x,0,0)$ falls off only as $1/|x|$. Still, due to the more rapid decay along a general real space direction, $W(\boldsymbol{r})$ is square integrable.

which is diagonal in the band indices, but not in the unit cell labels.

4.3.3 Tight binding redux

Suppose we have an orthonormal set of orbitals $|\, a\boldsymbol{R}\,\rangle$, where a labels the orbital and $\boldsymbol{R}$ denotes a Bravais lattice site. The label a may refer to different orbitals associated with the atom at $\boldsymbol{R}$, or it may label orbitals on other atoms in the unit cell defined by $\boldsymbol{R}$. We presume that $a \in \{1, \ldots, N_{\rm orb}\}$ with $N_{\rm orb}$ finite[12]. The most general tight binding Hamiltonian we can write is

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, |\, a\boldsymbol{R}\,\rangle\langle\, a'\boldsymbol{R}'\,| \quad , \tag{4.76}$$

where $H_{aa'}(\boldsymbol{R} - \boldsymbol{R}') = H^*_{a'a}(\boldsymbol{R}' - \boldsymbol{R}) = \langle\, a\boldsymbol{R}\,|\, H\,|\, a'\boldsymbol{R}'\,\rangle$ is the Hamiltonian matrix, whose rows and columns are indexed by a composite index combining both the unit cell label $\boldsymbol{R}$ and the orbital label a. When $\boldsymbol{R} = \boldsymbol{R}'$ and $a = a'$, the term $H_{aa}(0) = \varepsilon_a$ is the energy of a single electron in an isolated a orbital. For all other cases, $H_{aa'}(\boldsymbol{R}-\boldsymbol{R}') = -t_{aa'}(\boldsymbol{R}-\boldsymbol{R}')$ is the hopping integral between the a orbital in unit cell $\boldsymbol{R}$ and the a' orbital in unit cell $\boldsymbol{R}'$. Let's write an eigenstate $|\,\psi\,\rangle$ as

$$|\,\psi\,\rangle = \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{4.77}$$

Applying the Hamiltonian to $|\,\psi\,\rangle$, we obtain the coupled equations

$$\sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, \psi_{a'\boldsymbol{R}'}\, |\, a\boldsymbol{R}\,\rangle = E \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{4.78}$$

Since the $|\, a\boldsymbol{R}\,\rangle$ basis is complete, we must have that the coefficients of $|\, a\boldsymbol{R}\,\rangle$ on each side agree. Therefore,

$$\sum_{\boldsymbol{R}'} \sum_{a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, \psi_{a'\boldsymbol{R}'} = E\, \psi_{a\boldsymbol{R}} \quad . \tag{4.79}$$

We now use Bloch's theorem, which says that each eigenstate may be labeled by a wavevector $\boldsymbol{k}$, with $\psi_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}}\, u_a(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$. The $N^{-1/2}$ prefactor is a normalization term. Multiplying each side by by $e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$, we have

$$\sum_{a'} \left(\sum_{\boldsymbol{R}'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \right) u_{a'}(\boldsymbol{k}') = E(\boldsymbol{k})\, u_a(\boldsymbol{k}) \quad , \tag{4.80}$$

which may be written as

$$\sum_{a'} \widehat{H}_{aa'}(\boldsymbol{k})\, u_{a'}(\boldsymbol{k}) = E(\boldsymbol{k})\, u_a(\boldsymbol{k}) \quad , \tag{4.81}$$

[12] If our unit cell contained one s-orbital for each of the r basis sites, then $N_{\rm orb} = r$. But we may have multiple orbitals for each atom/ion within the unit cell.

where

$$\widehat{H}_{aa'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} H_{aa'}(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.82}$$

Thus, for each crystal wavevector $\boldsymbol{k}$, the $u_{a\boldsymbol{k}}$ are the eigenfunctions of the $N_{\rm orb} \times N_{\rm orb}$ Hermitian matrix $\widehat{H}_{aa'}(\boldsymbol{k})$. The energy eigenvalues at wavevector $\boldsymbol{k}$ are given by $\mathsf{spec}\big\{\widehat{H}(\boldsymbol{k})\big\}$, *i.e.* by the set of eigenvalues of the matrix $\widehat{H}(\boldsymbol{k})$. There are $N_{\rm orb}$ such solutions (some of which may be degenerate), which we distinguish with a band index n, and we denote $u_{n a}(\boldsymbol{k})$ and $E_n(\boldsymbol{k})$ as the corresponding eigenvectors and eigenvalues. We sometimes will use the definition $\hat{t}_{aa'}(\boldsymbol{k}) \equiv -\widehat{H}_{aa'}(\boldsymbol{k})$ for the matrix of hopping integrals.

4.3.4 Interlude on Fourier transforms

It is convenient to use second quantized notation and write the Hamiltonian as

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\, c^\dagger_{a\boldsymbol{R}}\, c_{a'\boldsymbol{R}'} \quad , \tag{4.83}$$

where $c^\dagger_{a\boldsymbol{R}}$ creates an electron in orbital a at unit cell $\boldsymbol{R}$. The *second quantized* fermion creation and annihilation operators satisfy the anticommutation relations

$$\big\{c_{a\boldsymbol{R}}\,,\, c^\dagger_{a'\boldsymbol{R}'}\big\} = \delta_{\boldsymbol{R}\boldsymbol{R}'}\,\delta_{aa'} \quad . \tag{4.84}$$

To quantize the wavevectors, we place our system on a d-dimensional torus with N_j unit cells along principal Bravais lattice vector $\boldsymbol{a}_j$ for all $j \in \{1, \ldots, d\}$. The total number of unit cells is then $N = N_1 N_2 \cdots N_d$. Consider now the Fourier transforms,

$$c_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} c_{a\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad c^\dagger_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} c^\dagger_{a\boldsymbol{k}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.85}$$

and their inverses

$$c_{a\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} c_{a\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \qquad , \qquad c^\dagger_{a\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} c^\dagger_{a\boldsymbol{R}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.86}$$

One then has $\big\{c_{a\boldsymbol{k}}\,,\, c_{a'\boldsymbol{k}'}\big\} = \delta_{aa'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'}$, which says that the Fourier space operators satisfy the same anticommutation relations as in real space, *i.e.* the individual $\boldsymbol{k}$ modes are orthonormal. This is equivalent to the result $\langle\, a\boldsymbol{k}\,|\,a'\boldsymbol{k}'\,\rangle = \delta_{aa'}\,\delta_{\boldsymbol{k}\boldsymbol{k}'}$, where $|\,a\boldsymbol{k}\,\rangle = N^{-1/2}\sum_{\boldsymbol{R}} |\,a\boldsymbol{R}\,\rangle\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$. The Hamiltonian may now be expressed as

$$H = \sum_{\boldsymbol{k}} \sum_{a,a'} \widehat{H}_{aa'}(\boldsymbol{k})\, c^\dagger_{a\boldsymbol{k}}\, c_{a'\boldsymbol{k}} \quad , \tag{4.87}$$

where $\widehat{H}_{aa'}(\boldsymbol{k})$ was defined in Eqn. 4.82 above.

You must, at the very deepest level of your soul, internalize Eqn. 4.85. Equivalently, using bra and ket vectors,

$$\langle\, a\boldsymbol{R}\,| = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \langle\, a\boldsymbol{k}\,| \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad |\, a\boldsymbol{R}\,\rangle = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} |\, a\boldsymbol{k}\,\rangle \, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.88}$$

and

$$\langle\, a\boldsymbol{k}\,| = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} \langle\, a\boldsymbol{R}\,| \, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad |\, a\boldsymbol{k}\,\rangle = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} |\, a\boldsymbol{R}\,\rangle \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{4.89}$$

To establish the inverse relations, we evaluate

$$\begin{aligned} |\, a\boldsymbol{R}\,\rangle &= \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} |\, a\boldsymbol{k}\,\rangle \, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ &= \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} \left(\frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}'} |\, a\boldsymbol{R}'\,\rangle \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}'} \right) e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ &= \sum_{\boldsymbol{R}'} \left(\frac{1}{N}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \right) |\, a\boldsymbol{R}'\,\rangle \quad . \end{aligned} \tag{4.90}$$

Similarly, we find

$$|\, a\boldsymbol{k}\,\rangle = \sum_{\boldsymbol{k}'} \left(\frac{1}{N}\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} \right) |\, a\boldsymbol{k}'\,\rangle \quad . \tag{4.91}$$

In order for the inverse relations to be true, then, the quantities in round brackets in the previous two equations must satisfy

$$\frac{1}{N}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')} = \delta_{\boldsymbol{R}\boldsymbol{R}'} \quad , \quad \frac{1}{N}\sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}} = \delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \tag{4.92}$$

Let's see how this works in the $d = 1$ case. Let the lattice constant be a and place our system on a ring of N sites (*i.e.* a one-dimensional torus). The k values are then quantized according to $k_j = 2\pi j/a$, where $j \in \{0, \ldots, N-1\}$. The first equation in Eqn. 4.92 is then

$$\frac{1}{N}\sum_{j=0}^{N-1} e^{-2\pi i j(n-n')/N} = \delta_{nn'} \quad , \tag{4.93}$$

where we have replaced $\boldsymbol{R}$ by na and $\boldsymbol{R}'$ by $n'a$, with $n, n' \in \{1, \ldots, N\}$. Clearly the above equality holds true when $n = n'$. For $n \neq n'$, let $z = e^{-2\pi i(n-n')/N}$. The sum is $1 + z + \ldots + z^{N-1} = (1 - z^N)/(1 - z)$. But $z^N = 1$ and $z \neq 1$, so the identity is again verified.

If we do not restrict n and n' to be among $\{1, \ldots, N\}$ and instead let their values range freely over the integers, then the formula is still correct, provided we write the RHS as $\delta_{n,n' \bmod N}$. Similarly, we must understand $\delta_{\boldsymbol{R}\boldsymbol{R}'}$ in Eqn. 4.92 to be unity whenever $\boldsymbol{R}' = \boldsymbol{R} + l_1 N_1 \boldsymbol{a}_1 + \ldots + l_d N_d \boldsymbol{a}_d$, where each $l_j \in \mathbb{Z}$, and zero otherwise. Similarly, $\delta_{\boldsymbol{k}\boldsymbol{k}'}$ is unity whenever $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{G}$, where $\boldsymbol{G} \in \widehat{\mathcal{L}}$ is any reciprocal lattice vector, and zero otherwise.

4.3.5 Examples of tight binding dispersions

One-dimensional lattice

Consider the case of a one-dimensional lattice. The lattice sites lie at positions $X_n = na$ for $n \in \mathbb{Z}$. The hopping matrix elements are $t(j) = t\,\delta_{j,1} + t\,\delta_{j,-1}$, where j is the separation between sites in units of the lattice constant a. Then $\hat{t}(k) = 2t\cos(ka)$ and the dispersion is $E(k) = -2t\cos(ka)$. Equivalently, and quite explicitly,

$$\begin{aligned} H &= -t\sum_n \Big(|\,n+1\,\rangle\langle\,n\,| + |\,n\,\rangle\langle\,n+1\,| \Big) = -\frac{t}{N}\sum_k \sum_{k'} \sum_n e^{-ik'(n+1)a}\, e^{ikna}\, |\,k\,\rangle\langle\,k'\,| + \text{H.c.} \\ &= -t\sum_k \sum_{k'} \left(\frac{1}{N}\sum_n e^{i(k-k')na} \right) e^{-ik'a}\, |\,k\,\rangle\langle\,k'\,| + \text{H.c.} = -2t\sum_k \cos(ka)\, |\,k\,\rangle\langle\,k\,| \quad , \end{aligned} \tag{4.94}$$

since the term in round brackets is $\delta_{kk'}$, as per Eqn. 4.92.

s-orbitals on cubic lattices

On a Bravais lattice with one species of orbital, there is only one band. Consider the case of s orbitals on a d-dimensional cubic lattice. The hopping matrix elements are

$$t(\boldsymbol{R}) = t\sum_{j=1}^{d} \Big(\delta_{\boldsymbol{R},\boldsymbol{a}_j} + \delta_{\boldsymbol{R},-\boldsymbol{a}_j} \Big) \quad , \tag{4.95}$$

where $\boldsymbol{a}_j = a\,\hat{\mathbf{e}}_j$ is the j^{th} elementary direct lattice vector. Taking the discrete Fourier transform (DFT) as specified in Eqn. 4.82,

$$\hat{t}(\boldsymbol{k}) = 2t\sum_{j=1}^{d} \cos(k_j a) \quad . \tag{4.96}$$

The dispersion is then $E(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k})$. The model exhibits a *particle-hole symmetry*,

$$\widetilde{c}_{\boldsymbol{k}} \equiv c^\dagger_{\boldsymbol{k}+\boldsymbol{Q}} \quad , \tag{4.97}$$

where $\boldsymbol{Q} = \frac{\pi}{a}\big(\hat{\mathbf{e}}_1 + \ldots + \hat{\mathbf{e}}_d\big)$. Note $\hat{t}(\boldsymbol{k}+\boldsymbol{Q}) = -\hat{t}(\boldsymbol{k})$.

s-orbitals on the triangular lattice

The triangular lattice is depicted as the lattice of black dots in the left panel of Fig. 4.9. The elementary direct lattice vectors are

$$\boldsymbol{a}_1 = a\left(\tfrac{1}{2}\hat{\mathbf{e}}_1 - \tfrac{\sqrt{3}}{2}\hat{\mathbf{e}}_2\right) \quad , \qquad \boldsymbol{a}_2 = a\left(\tfrac{1}{2}\hat{\mathbf{e}}_1 + \tfrac{\sqrt{3}}{2}\hat{\mathbf{e}}_2\right) \quad , \tag{4.98}$$

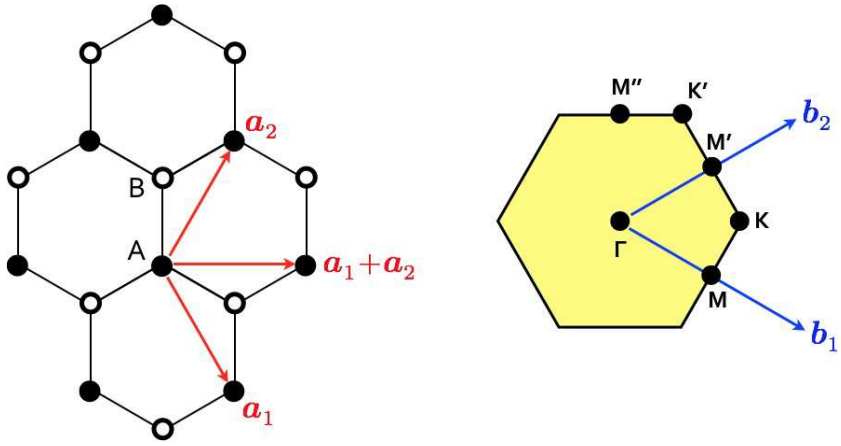


Figure 4.9: Left: The honeycomb lattice is a triangular lattice (black sites) with a two element basis (add white sites). $\boldsymbol{a}_{1,2}$ are elementary direct lattice vectors. Right: First Brillouin zone for the triangular lattice. $\boldsymbol{b}_{1,2}$ are elementary reciprocal lattice vectors. Points of high symmetry Γ, K, K′, M, M′, and M″ are shown.

and the elementary reciprocal lattice vectors are

$$\boldsymbol{b}_1 = \frac{4\pi}{\sqrt{3}\,a}\left(\tfrac{\sqrt{3}}{2}\hat{\mathrm{e}}_1 - \tfrac{1}{2}\hat{\mathrm{e}}_2\right) \quad , \quad \boldsymbol{b}_2 = \frac{4\pi}{\sqrt{3}\,a}\left(\tfrac{\sqrt{3}}{2}\hat{\mathrm{e}}_1 + \tfrac{1}{2}\hat{\mathrm{e}}_2\right) \quad . \tag{4.99}$$

Note that $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\,\delta_{ij}$. The hopping matrix elements are

$$t(\boldsymbol{R}) = t\,\delta_{\boldsymbol{R},\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_2} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_2} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_3} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_3} \quad , \tag{4.100}$$

where $\boldsymbol{a}_3 \equiv \boldsymbol{a}_1 + \boldsymbol{a}_2$. Thus,

$$\begin{aligned}\hat{t}(\boldsymbol{k}) &= 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1) + 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_2) + 2t\,\cos(\boldsymbol{k}\cdot\boldsymbol{a}_3) \\ &= 2t\cos(\theta_1) + 2t\cos(\theta_2) + 2t\cos(\theta_1+\theta_2) \quad .\end{aligned} \tag{4.101}$$

Here we have written

$$\boldsymbol{k} = \frac{\theta_1}{2\pi}\,\boldsymbol{b}_1 + \frac{\theta_2}{2\pi}\,\boldsymbol{b}_2 + \frac{\theta_3}{2\pi}\,\boldsymbol{b}_3 \quad , \tag{4.102}$$

and therefore for a general $\boldsymbol{R} = l_1\,\boldsymbol{a}_1 + l_2\,\boldsymbol{a}_2 + l_3\,\boldsymbol{a}_3$, we have

$$\boldsymbol{k}\cdot(l_1\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + l_3\boldsymbol{a}_3) = l_1\theta_1 + l_2\theta_2 + l_3\theta_3 \quad . \tag{4.103}$$

Again there is only one band, because the triangular lattice is a Bravais lattice. The dispersion relation is $E(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k})$. Unlike the case of the d-dimensional cubic lattice, the triangular lattice energy band does not exhibit particle-hole symmetry. The extrema are at $E_{\rm min} = E(\Gamma) = -6t$, and $E_{\rm max} = E(\mathrm{K}) = +3t$, where $\Gamma = 0$ is the zone center and $\mathrm{K} = \frac{1}{3}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$ is the zone corner, corresponding to $\theta_1 = \theta_2 = \frac{2\pi}{3}$.

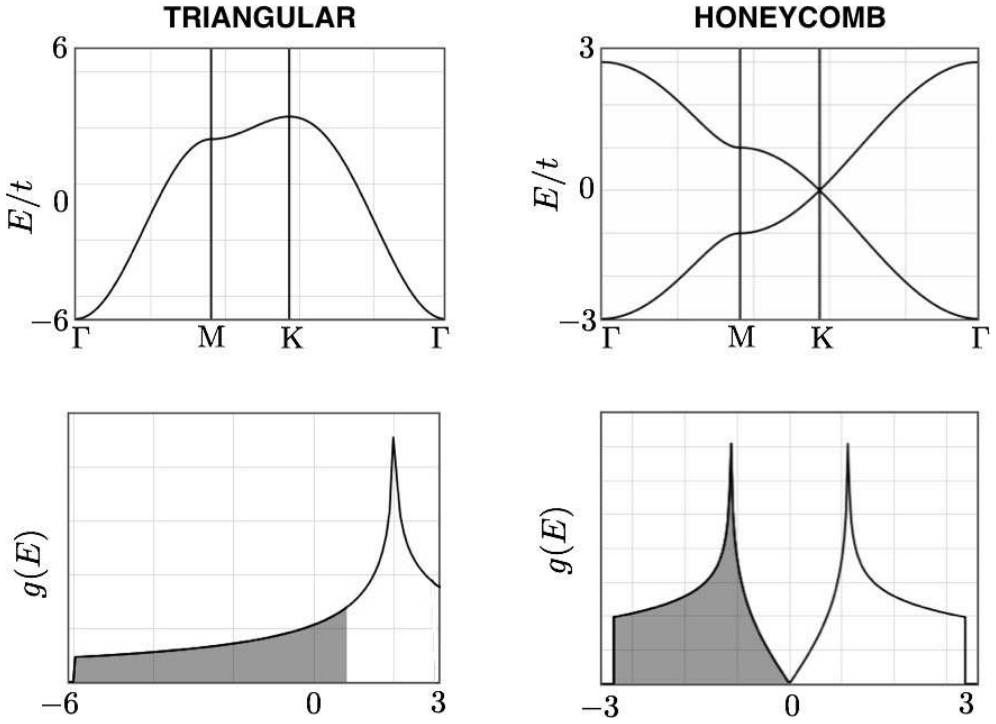


Figure 4.10: Energy bands and density of states for nearest neighbor s-orbital tight binding model on triangular (left) and honeycomb (right) lattices.

Graphene: π-orbitals on the honeycomb lattice

Graphene is a two-dimensional form of carbon arrayed in a honeycomb lattice. The electronic structure of carbon is $1s^2\,2s^2\,2p^2$. The $1s$ electrons are tightly bound and have small overlaps from site to site, hence little dispersion. The $2s$ and $2p_{x,y}$ orbitals engage in sp^2 hybridization. For each carbon atom, three electrons in each atom's sp^2 orbitals are distributed along bonds connecting to its neighbors[13]. Thus each bond gets two electrons (of opposite spin), one from each of its carbon atoms. This is what chemists call a σ-bond. The remaining p_z orbital (the π orbital, to our chemist friends) is then free to hop from site to site. For our purposes it is equivalent to an s-orbital, so long as we don't ask about its properties under reflection in the x-y plane. The underlying Bravais lattice is triangular, with a two element basis (labelled A and B in Fig. 4.9). According to the left panel of Fig. 4.9, the A sublattice site in unit cell $\boldsymbol{R}$ is connected to the B sublattice sites in unit cells $\boldsymbol{R}$, $\boldsymbol{R}+\boldsymbol{a}_1$, and $\boldsymbol{R}-\boldsymbol{a}_2$. Thus, the hopping matrix element between the A sublattice sites in unit sell 0 and the B sublattice site in unit cell $\boldsymbol{R}$ is given by

$$t_{\rm AB}(\boldsymbol{R}) = t\,\delta_{\boldsymbol{R},0} + t\,\delta_{\boldsymbol{R},\boldsymbol{a}_1} + t\,\delta_{\boldsymbol{R},-\boldsymbol{a}_2} \quad , \tag{4.104}$$

[13]In diamond, the carbon atoms are fourfold coordinated, and th orbitals are sp^3 hybridized.

and therefore $\hat{t}_{\rm AB}(\boldsymbol{k}) = t\big(1 + e^{-i\theta_1} + e^{i\theta_2}\big)$. We also have $\hat{t}_{\rm BA}(\boldsymbol{k}) = \hat{t}^*_{\rm AB}(\boldsymbol{k})$ and $\hat{t}_{\rm AA}(\boldsymbol{k}) = \hat{t}_{\rm BB}(\boldsymbol{k}) = 0$. Thus, the Hamiltonian matrix is

$$H(\boldsymbol{k}) = -\begin{pmatrix} 0 & \hat{t}_{\rm AB}(\boldsymbol{k}) \\ \hat{t}^*_{\rm AB}(\boldsymbol{k}) & 0 \end{pmatrix} = -t\begin{pmatrix} 0 & 1 + e^{-i\theta_1} + e^{i\theta_2} \\ 1 + e^{i\theta_1} + e^{-i\theta_2} & 0 \end{pmatrix} \quad , \tag{4.105}$$

and the energy eigenvalues are

$$E_\pm(\boldsymbol{k}) = \pm\big|\hat{t}_{\rm AB}(\boldsymbol{k})\big| = \pm t\sqrt{3 + 2\cos\theta_1 + 2\cos\theta_2 + 2\cos(\theta_1 + \theta_2)} \quad . \tag{4.106}$$

These bands are depicted in the right panels of Fig. 4.10. Note the band touching at K (and K′), which are known as *Dirac points*. In the vicinity of either Dirac point, writing $\boldsymbol{k} = \boldsymbol{K} + \boldsymbol{q}$ or $\boldsymbol{k} = \boldsymbol{K}' + \boldsymbol{q}$, one has $E_\pm(\boldsymbol{k}) = \pm\hbar v_{\rm F}|\boldsymbol{q}|$, where $v_{\rm F} = \frac{\sqrt{3}}{2}\,ta/\hbar$ is the *Fermi velocity*. At the electroneutrality point (*i.e.* one π electron per site), the Fermi levels lies precisely at $\varepsilon_{\rm F} = 0$. The density of states vanishes continuously as one approaches either Dirac point[14].

p-orbitals on the cubic lattice

Finally, consider the case where each site hosts a trio $(p_x\,,\,p_y\,,\,p_z)$ of p-orbitals. Let the separation between two sites be $\boldsymbol{R}$. Then the 3×3 hopping matrix between these sites depends on two tensors, $\delta_{\mu\nu}$ and $\widehat{\boldsymbol{R}}_\mu\widehat{\boldsymbol{R}}_\nu$. When $\hat{\eta}$ lies along one of the principal cubic axes, the situation is as depicted in Fig. 4.11. The hopping matrix is

$$t_{\mu\nu}(\boldsymbol{R}) = t_{\rm w}(R)\,\delta_{\mu\nu} - \big(t_{\rm w}(R) + t_{\rm s}(R)\big)\,\hat{R}_\mu\,\hat{R}_\nu \quad , \tag{4.107}$$

where the weak and strong hoppings $t_{\rm w,s}$ are depicted in Fig. 4.11. We can now write

$$t_{xx}(\boldsymbol{R}) = -t_{\rm s}\left(\delta_{\boldsymbol{R},\boldsymbol{a}_1} + \delta_{\boldsymbol{R},-\boldsymbol{a}_1}\right) + t_{\rm w}\left(\delta_{\boldsymbol{R},\boldsymbol{a}_2} + \delta_{\boldsymbol{R},-\boldsymbol{a}_2} + \delta_{\boldsymbol{R},\boldsymbol{a}_3} + \delta_{\boldsymbol{R},-\boldsymbol{a}_3}\right) \quad , \tag{4.108}$$

and therefore

$$\hat{t}_{\mu\nu}(\boldsymbol{k}) = 2\begin{pmatrix} -t_{\rm s}\cos\theta_1 + t_{\rm w}(\cos\theta_2 + \cos\theta_3) & 0 & 0 \\ 0 & -t_{\rm s}\cos\theta_2 + t_{\rm w}(\cos\theta_1 + \cos\theta_3) & 0 \\ 0 & 0 & -t_{\rm s}\cos\theta_3 + t_{\rm w}(\cos\theta_1 + \cos\theta_2) \end{pmatrix} \tag{4.109}$$

which is diagonal. The three p-band dispersions are given by the diagonal entries.

4.3.6 Bloch's theorem, again

In Eqn. 4.76,

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'}\sum_{a,a'} H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')\,|\,a\boldsymbol{R}\,\rangle\langle\,a'\boldsymbol{R}'\,| \quad ,$$

[14] *I.e.*, the DOS in either the K or K′ valley.

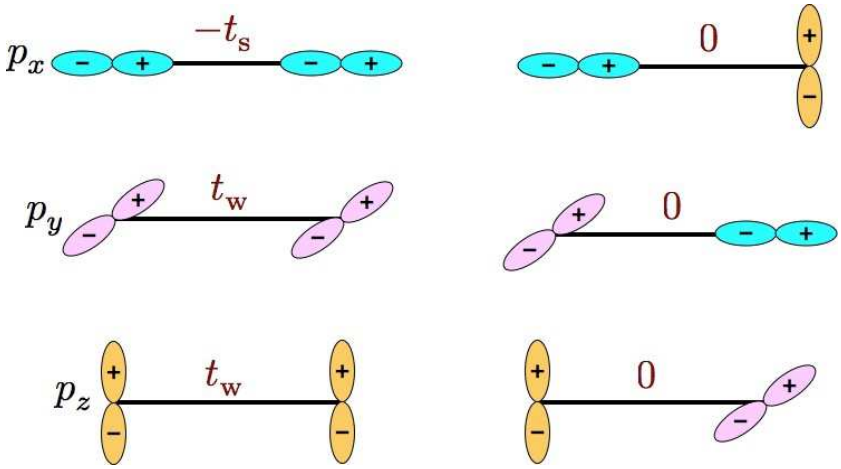


Figure 4.11: Matrix elements for neighboring tight binding orbitals of p symmetry.

$\boldsymbol{R}$ and $\boldsymbol{R}'$ labeled Bravais lattice sites, while a and a' labeled orbitals. We stress that these orbitals don't necessarily have to be located on the same ion. We should think of $\boldsymbol{R}$ and $\boldsymbol{R}'$ labeling *unit cells*, each of which is indeed associated with a Bravais lattice site. For example, in the case of graphene, $|\, a\boldsymbol{R}\,\rangle$ represents an orbital on the a sublattice in unit cell $\boldsymbol{R}$. The eigenvalue equation may be written

$$\widehat{H}_{aa'}(\boldsymbol{k})\, u_{na'}(\boldsymbol{k}) = E_n(\boldsymbol{k})\, u_{na}(\boldsymbol{k}) \quad , \tag{4.110}$$

where n is the band index. The function $u_{na}(\boldsymbol{k})$ is the *internal wavefunction* within a given cell, and corresponds to the cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ in the continuum, with $a \leftrightarrow (\boldsymbol{r} - \boldsymbol{R})$ labeling a position within each unit cell. The full Bloch state may then be written

$$|\, \psi_{n\boldsymbol{k}}\,\rangle = |\, \boldsymbol{k}\,\rangle \otimes |\, u_{n\boldsymbol{k}}\,\rangle \quad , \tag{4.111}$$

so that

$$\begin{aligned} \psi_{n\boldsymbol{k}}(\boldsymbol{R}, a) &= \Big(\langle\, \boldsymbol{R}\,| \otimes \langle\, a\,|\Big)\Big(|\, \boldsymbol{k}\,\rangle \otimes |\, u_{n\boldsymbol{k}}\,\rangle\Big) \\ &= \langle\, \boldsymbol{R}\,|\, \boldsymbol{k}\,\rangle\, \langle\, a\,|\, u_{n\boldsymbol{k}}\,\rangle = \frac{1}{\sqrt{N}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, u_{na}(\boldsymbol{k}) \quad . \end{aligned} \tag{4.112}$$

Here we have chosen a normalization $\sum_a \big|u_{na}(\boldsymbol{k})\big|^2 = 1$ within each unit cell, which entails the overall normalization $\sum_{\boldsymbol{R},a} \big|\psi_{n\boldsymbol{k}}(\boldsymbol{R}, a)\big|^2 = 1$.

4.3.7 Go flux yourself : how to add magnetic fields

To simplify matters, we consider only s-orbitals on two-dimensional lattices. The general tight-binding Hamiltonian is written

$$H = -\sum_{\boldsymbol{r}<\boldsymbol{r}'} \Big(t_{\boldsymbol{r}\boldsymbol{r}'}\, c^\dagger_{\boldsymbol{r}}\, c_{\boldsymbol{r}'} + t^*_{\boldsymbol{r}\boldsymbol{r}'}\, c^\dagger_{\boldsymbol{r}'}\, c_{\boldsymbol{r}} \Big) \quad , \tag{4.113}$$

where the notation $\boldsymbol{r} < \boldsymbol{r}'$ means that each pair $(\boldsymbol{r}, \boldsymbol{r}')$ summed only once. We may write $t_{\boldsymbol{r}\boldsymbol{r}'} = t^*_{\boldsymbol{r}'\boldsymbol{r}} = |t_{\boldsymbol{r}\boldsymbol{r}'}| \exp(iA_{\boldsymbol{r}\boldsymbol{r}'})$, where $A_{\boldsymbol{r}\boldsymbol{r}'}$ is a gauge field living on the links of the lattice. Let p denote a plaquette on the lattice. Then the dimensionless flux ϕ_p (in units of $\hbar c/e$) through plaquette p is

$$\phi_p = \sum_{\langle \boldsymbol{r}\boldsymbol{r}' \rangle \in \partial p} A_{\boldsymbol{r}\boldsymbol{r}'} \quad , \tag{4.114}$$

where the sum is taken in a counterclockwise fashion along the links on the boundary of p. The tight-binding Hamiltonian exhibits a *gauge invariance* under the combined operations

$$\begin{aligned} c_{\boldsymbol{r}} &\to e^{i\alpha_{\boldsymbol{r}}}\, c_{\boldsymbol{r}} \\ A_{\boldsymbol{r}\boldsymbol{r}'} &\to A_{\boldsymbol{r}\boldsymbol{r}'} + \alpha_{\boldsymbol{r}} - \alpha_{\boldsymbol{r}'} \quad . \end{aligned} \tag{4.115}$$

Consider now the case of the square lattice. It is clear that any configuration of the $A_{\boldsymbol{r}\boldsymbol{r}'}$ which is periodic in the structural unit cell, *i.e.* under translations by elementary direct lattice vectors, must correspond to $\phi_p = 0$ for every plaquette p[15]. This is because the phase $A_{\boldsymbol{r}\boldsymbol{r}'}$ is associated with the *directed link* from $\boldsymbol{r}$ to $\boldsymbol{r}'$, and parallel links on opposite sides of the elementary square plaquette will yield equal and opposite values of $A_{\boldsymbol{r}\boldsymbol{r}'}$ because they are traversed in opposite directions. *In order to describe nonzero flux per plaquette, the configuration of the lattice gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ must break lattice translational symmetry*[16]. Consider the case where $\phi = \pi$ in each plaquette. A configuration for the gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ yielding this flux distribution is shown in the left panel of Fig. 4.12. All links have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$, hence $t_{\boldsymbol{r}\boldsymbol{r}'} = t \exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) = t$, except for the links depicted with slashes, for which $A_{\boldsymbol{r}\boldsymbol{r}'} = \pi$ and $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. The *magnetic unit cell* is now a 2×1 block consisting of one cell from each sublattice (blue and red). We call this a magnetic unit cell to distinguish it from the *structural unit cell* of the underlying square lattice. The *structural* Bravais lattice is square, with elementary direct lattice vectors are $\boldsymbol{a}_1 = a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. But the *magnetic* Bravais lattice is rectangular, with elementary RLVs $\boldsymbol{a}_1 = 2a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. From Bloch's theorem, the phase of the wavefunction varies by $\exp(i\boldsymbol{k}\cdot\boldsymbol{a}_1) \equiv \exp(i\theta_1)$ across the unit cell in the x-direction, and by $\exp(i\boldsymbol{k}\cdot\boldsymbol{a}_2) \equiv \exp(i\theta_2)$ in the y-direction. The Hamiltonian is

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 + e^{i\theta_1} \\ 1 + e^{-i\theta_1} & -2\cos\theta_2 \end{pmatrix} \tag{4.116}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\cos^2(\frac{1}{2}\theta_1) + \cos^2\theta_2}$. The band gap collapses at two points: $(\theta_1, \theta_2) = (\pi, \pm\frac{1}{2}\pi)$. Writing $(\theta_1, \theta_2) = (\pi + \delta_1\,,\, \pm\frac{1}{2}\pi + \delta_2)$, we find

$$E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\sin^2(\tfrac{1}{2}\delta_1) + \sin^2\delta_2} = \pm 2ta\,\sqrt{q_1^2 + q_2^2} + \mathcal{O}(|\boldsymbol{q}|^3) \quad , \tag{4.117}$$

[15] More precisely, if $A_{\boldsymbol{r}\boldsymbol{r}'}$ is periodic in the structural unit cell, then each structural unit cell is congruent to a zero flux state. However, it may be that a structural cell is comprised of more than one elementary plaquette, as is the case with the triangular lattice (each structural cell consists of two triangles), or that there are closed loops which don't correspond to a structural unit cell due to further neighbor hoppings. In such cases, there may be closed loops on the lattice whose flux is not congruent to zero. See §4.3.7 for some examples.

[16] By "nonzero" flux, we mean $\phi \bmod 2\pi \neq 0$.

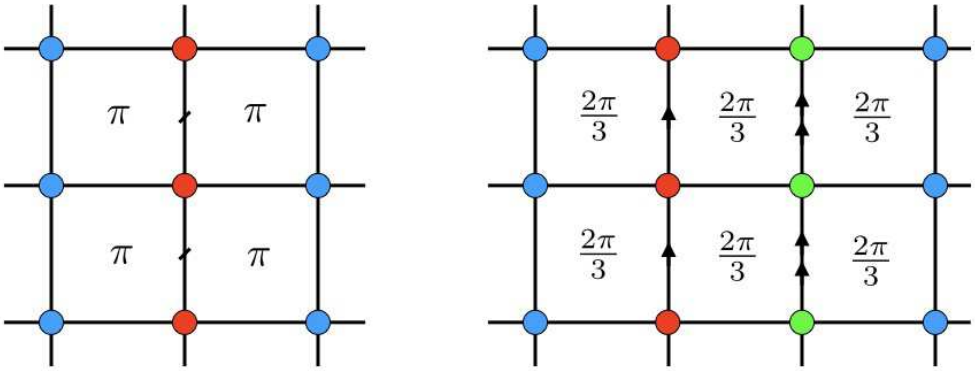


Figure 4.12: Gauges for the square lattice Hofstadter model. Left: $\phi = \pi$ case. $t_{\boldsymbol{r}\boldsymbol{r}'} = t$ on all links except those with slashes, where $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. Right: $\phi = \frac{2}{3}\,\pi$. Each arrow corresponds to a factor of $\exp(2\pi i/3)$.

which is a Dirac cone! Thus, the dispersion for the square lattice π flux model has two Dirac points. Here, $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_{\rm D}$ is the wavevector measured from either Dirac point[17].

The π flux state is time-reversal symmetric, since under time reversal we have $\exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) \to \exp(-iA_{\boldsymbol{r}\boldsymbol{r}'})$, hence $\phi_p \to -\phi_p$. But flux is only defined modulo 2π, hence $\pi \to -\pi \cong \pi$ yields the same flux configuration.

A more interesting state of affairs pertains for the case $\phi = \frac{2}{3}\pi$, for which a valid gauge configuration $A_{\boldsymbol{r}\boldsymbol{r}'}$ is shown in the right panel of Fig. 4.12. Now there are three sites per unit cell: A (blue), B (red), and C (green). The Bloch phase accrued across the magnetic unit cell in the $\pm\hat{\boldsymbol{x}}$ direction is $e^{\pm i\theta_1}$, and in the $\pm\hat{\boldsymbol{y}}$ direction is $e^{\pm i\theta_2}$. Thus

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & e^{i\theta_1} \\ 1 & 2\cos(\theta_2 + \frac{2\pi}{3}) & 1 \\ e^{-i\theta_1} & 1 & 2\cos(\theta_2 + \frac{4\pi}{3}) \end{pmatrix} \quad . \tag{4.118}$$

The general case where the flux per structural unit cell is $\phi = 2\pi p/q$ is known as the Hofstadter model[18]. In this case, the magnetic unit cell is a $q \times 1$ block, and the resulting $q \times q$ Hamiltonian is given by

$$\begin{aligned} H_{jj'}(\boldsymbol{\theta}) = -2t\cos\Big(\theta_2 + \tfrac{2\pi(j-1)p}{q}\Big)\,\delta_{jj'} - t\,\delta_{j',j+1}\big(1-\delta_{j,q}\big) - t\,e^{-i\theta_1}\,\delta_{j,q}\,\delta_{j',1} \\ - t\,\delta_{j',j-1}\big(1-\delta_{j,1}\big) - t\,e^{i\theta_1}\,\delta_{j,1}\,\delta_{j',q} \quad . \end{aligned} \tag{4.119}$$

[17] Note $\theta_1 = 2k_x a = \pi + 2q_x a$ and $\theta_2 = k_y a = \frac{1}{2}\pi + q_y a$, hence $\delta_1 = 2q_x a$ and $\delta_2 = q_y a$.
[18] See D. R. Hofstadter, *Phys. Rev. B* **14**, 2239 (1976).

In other words,

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & 0 & \cdots & 0 & e^{i\theta_1} \\ 1 & 2\cos\left(\theta_2 + \frac{2\pi p}{q}\right) & 1 & & & 0 \\ 0 & 1 & 2\cos\left(\theta_2 + \frac{4\pi p}{q}\right) & 1 & & \vdots \\ \vdots & 0 & 1 & \ddots & & \vdots \\ & & & & & \\ 0 & & & & & 1 \\ e^{-i\theta_1} & 0 & & \cdots & 1 & 2\cos\left(\theta_2 + \frac{2\pi(q-1)p}{q}\right) \end{pmatrix} . \tag{4.120}$$

There are thus q magnetic subbands. Note that

$$H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = XUH(\theta_1, \theta_2)\, U^\dagger X^\dagger \quad , \tag{4.121}$$

where $X_{ij} = \delta_{i,j+1 \bmod q}$ and $U = \mathsf{diag}(1, e^{-i\theta_1}, \cdots, e^{-i\theta_1})$. Thus,

$$\mathsf{spec}\, H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = \mathsf{spec}\, H(\theta_1, \theta_2) \quad , \tag{4.122}$$

as we saw explicitly in the $q = 2$ case above. A plot of the magnetic subbands in (E, ϕ) space, known as *Hofstadter's butterfly*, is shown in Fig. 4.13.

In the limit where the denominator q of the flux $\phi = 2\pi p/q$ is large (for fixed p), the flux per cell is very small. We then expect to recover the continuum Landau level spectrum $E_n = (n+\frac{1}{2})\hbar\omega_{\rm c}$. To express this in terms of the flux ϕ, note that the $B = 0$ dispersion is

$$E(\boldsymbol{k}) = -2t\cos(k_x a) - 2t\cos(k_y a) = -4t + t\boldsymbol{k}^2 a^2 + \ldots \quad , \tag{4.123}$$

which allows us to identify the effective mass m from the coefficient of the $\boldsymbol{k}^2$ term, with the result $m = \hbar^2/2ta^2$. The magnetic field is the flux per unit area, hence $B = \phi\hbar c/ea^2$. Thus,

$$\hbar\omega_{\rm c} = \frac{\hbar eB}{mc} = \frac{\hbar e}{c} \times \frac{\phi\,\hbar c}{ea^2} \times \frac{2ta^2}{\hbar^2} = 2\phi\, t \quad . \tag{4.124}$$

This describes the corners of the Hofstadter butterfly in Fig. 4.13, where continuum Landau levels radiate outward from the energies $\pm 4t$ according to

$$E_n(\phi) = \pm\Big(4t - (2n+1)\,\phi\, t\Big) \qquad \text{and} \qquad E_n(\phi) = \pm\Big(4t - (2n+1)\,(2\pi - \phi)\, t\Big) \quad , \tag{4.125}$$

for $\phi \ll 1$.

Unit cells with zero net flux

As mentioned in a footnote above, it is not quite true that a lattice gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ which is periodic in the underlying Bravais lattice unit cell leads to zero net flux in every plaquette or

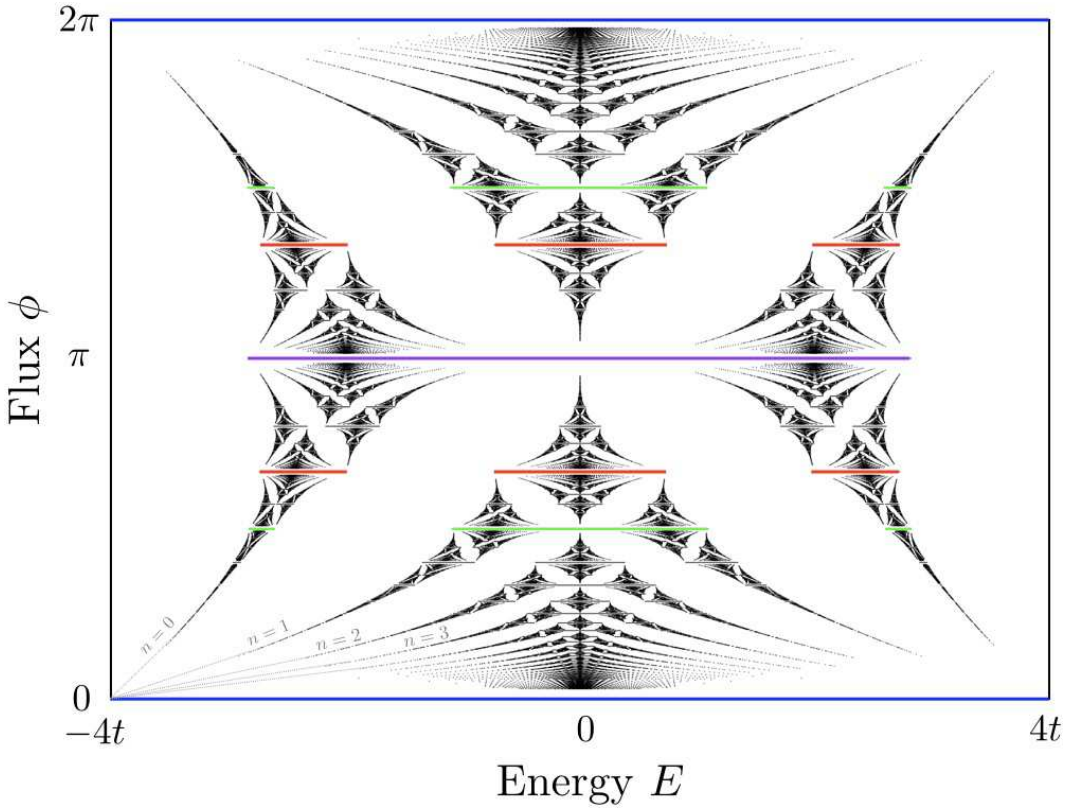


Figure 4.13: Magnetic subbands for the square lattice Hofstadter model for flux per plaquette $\phi \in [0, 2\pi]$. Blue bands at $\phi = 0$ and $\phi = 2\pi$ have the full bandwidth $W = 8t$. At $\phi = \pi$ (purple), there are two subbands with $E_- \in [-2\sqrt{2}t, 0]$ and $E_+ \in [0, 2\sqrt{2}t]$ which touch at $E = 0$. Similarly, at $\phi = \pm\frac{1}{2}\pi$ (green), there are four subbands, with the central two bands touching at $E = 0$. At $\phi = \pm\frac{2}{3}\pi$ (red), there are three subbands. Continuum Landau levels are shown radiating from the lower left corner.

closed loop of links on the lattice. Two counterexamples are shown in Fig. 4.14. The first example is that of the triangular lattice, where each structural unit cell is a rhombus consisting of two elementary triangular plaquettes. Consider now the situation where each horizontal link carries a U(1) phase α, *i.e.* $A_{\boldsymbol{r}\boldsymbol{r}'} = t\, e^{i\alpha}$, while the remaining links all have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$. Computing the U(1) flux by taking the directed sum counterclockwise over each triangle, we see that all the up triangles carry flux $\phi_\triangle = \alpha$, while all the down triangles carry flux $\phi_\triangledown = -\alpha \cong 2\pi - \alpha$. If, as before, we take the elementary direct lattice vectors to be $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ and write $\boldsymbol{k} = \sum_{j=1}^{2} \theta_j\, \boldsymbol{b}_j/2\pi$, with $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\,\delta_{ij}$, then the tight binding Hamiltonian for this triangular lattice model is given by $H = \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}\, a^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}}$, where

$$\begin{aligned} E_{\boldsymbol{k}} &= -2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1 + \boldsymbol{k}\cdot\boldsymbol{a}_2 + \alpha) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_2) \\ &= -2t\cos(\theta_1 + \theta_2 + \alpha) - 2t\cos\theta_1 - 2t\cos\theta_2 \quad . \end{aligned} \tag{4.126}$$

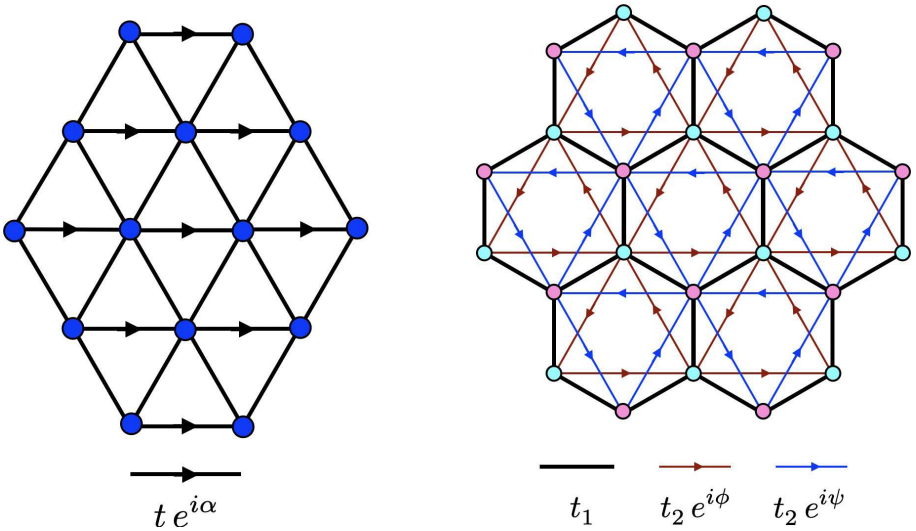


Figure 4.14: Two models with zero net flux per unit cell which still break time reversal symmetry. Left: The unit cell of the triangular lattice consists of two triangles.

A more interesting state of affairs is depicted in the right panel of Fig. 4.14, which is graphene augmented by nearest neighbor same-sublattice hopping terms, which is the celebrated Haldane honeycomb lattice model[19]. Inscribed in each hexagonal cell are one up-triangle of A sites, depicted by blue dots in the figure, and one down-triangle of B sites, depicted as pink dots in the figure. Again we take $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ for the underlying A Bravais lattice, with the basis vectors $\mathbf{0}$ and $\boldsymbol{\delta}_1 = a\hat{\boldsymbol{y}}$. The nearest neighbor hoppings between A and B sites all are taken to have (real) amplitude t_1, while the inscribed A and B same-sublattice hoppings are taken to be $t_2\,e^{i\phi}$ and $t_2\,e^{i\psi}$, respectively, and taken in the counterclockwise direction around the inscribed $\triangle$ and $\triangledown$ paths. An on-site energy term $\pm m$, called the *Semenoff mass*, is added to the hopping terms. Taking the product of the real t_1 amplitudes around any hexagonal unit cell along the thick black lines, we see that the flux per unit cell is congruent to zero. However, *within* each unit cell there is nonzero flux, distributed nonuniformly such that the total flux is zero. The corresponding tight binding Hamiltonian is given by

$$\begin{aligned} H = m\sum_i \big(a_i^\dagger\,a_i - b_i^\dagger\,b_i\big) - t_1\sum_{\langle ij\rangle}\big(a_i^\dagger\,b_j + b_j^\dagger\,a_i\big) \\ - t_2\sum_{\triangle}\big(e^{i\phi}\,a_i^\dagger\,a_j + e^{-i\phi}\,a_j^\dagger\,a_i\big) - t_2\sum_{\triangledown}\big(e^{i\psi}\,b_i^\dagger\,b_j + e^{-i\psi}\,b_j^\dagger\,b_i\big) \quad , \end{aligned} \tag{4.127}$$

where the same-sublattice hopping along the triangles $\triangle$ and $\triangledown$ inscribed within each hexagon

[19] F. D. M. Haldane, *Phys. Rev. Lett.* **61**, 2015 (1988).

are to be traversed in a counterclockwise direction. Fourier transforming to $\boldsymbol{k}$-space, one has

$$\begin{aligned} H = \sum_{\boldsymbol{k}} \Big\{ &- t_1 \left(1 + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) a^\dagger_{\boldsymbol{k}}\, b_{\boldsymbol{k}} - t_1 \left(1 + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) b^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}} \\ &+ \Big[m - 2t_2\,\mathsf{Re}\left(e^{i\phi}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\phi}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\phi}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big]\, a^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}} \\ &+ \Big[-m - 2t_2\,\mathsf{Re}\left(e^{i\psi}\, e^{-i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\psi}\, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\psi}\, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big]\, b^\dagger_{\boldsymbol{k}}\, b_{\boldsymbol{k}} \Big\} \quad , \end{aligned} \tag{4.128}$$

and thus we may write

$$H = \sum_{\boldsymbol{k}} \begin{pmatrix} a^\dagger_{\boldsymbol{k}} & b^\dagger_{\boldsymbol{k}} \end{pmatrix} \begin{pmatrix} H_{\mathsf{AA}}(\boldsymbol{k}) & H_{\mathsf{AB}}(\boldsymbol{k}) \\ H_{\mathsf{BA}}(\boldsymbol{k}) & H_{\mathsf{BB}}(\boldsymbol{k}) \end{pmatrix} \begin{pmatrix} a_{\boldsymbol{k}} \\ b_{\boldsymbol{k}} \end{pmatrix} \quad , \tag{4.129}$$

where

$$\begin{aligned} H_{\mathsf{AA}}(\boldsymbol{k}) &= m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi \\ H_{\mathsf{BB}}(\boldsymbol{k}) &= -m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\psi + 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\psi \\ H_{\mathsf{AB}}(\boldsymbol{k}) &= H^*_{\mathsf{BA}}(\boldsymbol{k}) = -t_1\big(1 + e^{i\theta_1} + e^{-i\theta_2}\big) \quad . \end{aligned} \tag{4.130}$$

In the Haldane model one has $\psi = \phi$, in which case we may write $H(\boldsymbol{k})$ in terms of Pauli matrices, as

$$\begin{aligned} H(\boldsymbol{k}) = &-2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - t_1\big(1 + \cos\theta_1 + \cos\theta_2\big)\,\sigma^x \\ &+ t_1\big(\sin\theta_1 - \sin\theta_2\big)\,\sigma^y + \Big(m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi\Big)\,\sigma^z \quad . \end{aligned} \tag{4.131}$$

What makes the Haldane model so interesting is that its band structure is *topological* over a range of the dimensionless parameters m/t_2 and ϕ.[20] More on this below!

4.3.8 General flux configuration on the square lattice

More generally, consider a magnetic unit cell formed by an $M \times N$ block of structural unit cells, as depicted in Fig. 4.15. Each structural cell p is labeled by the indices (m, n), where the Bravais lattice site its lower left corner is $ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$. To assign the lattice gauge fields, do the following. For $\boldsymbol{r} = ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = ma\hat{\boldsymbol{x}} + (n+1)a\hat{\boldsymbol{y}}$ with $n < N$, let $A_{\boldsymbol{r}\boldsymbol{r}'} = \sum_{i=1}^{m-1} \phi_{i,n}$. For the $n = N$, we include the Bloch phase θ_2, so that $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_2 + \sum_{i=1}^{m-1} \phi_{i,n}$, also noting that $(m, N+1) \cong (m, 1)$. This sets $A_{\boldsymbol{r}\boldsymbol{r}'}$ for all vertical (y-directed) links. The only horizontal links for which $A_{\boldsymbol{r}\boldsymbol{r}'}$ are nonzero are those with $\boldsymbol{r} = Ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = a\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$; note $(M+1, n) \cong (1, n)$. Then $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_1 - \sum_{i=1}^{M}\sum_{j=1}^{n} \phi_{i,j}$. One can check that this prescription yields the desired flux configuration, as well as the two Bloch phases.

[20] Without loss of generality, we may set $t_1 \equiv 1$.

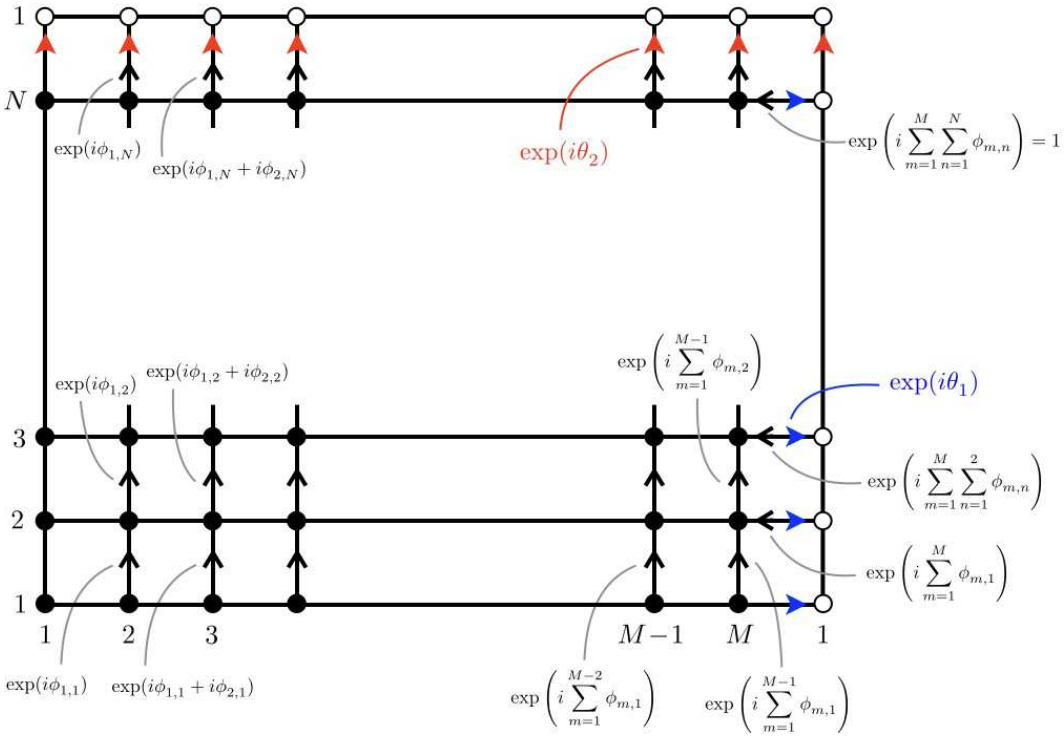


Figure 4.15: Lattice gauge field configuration for a general $M \times N$ rectangular lattice with flux $\phi_{m,n}$ in unit cell (m, n) and Bloch phases (θ_1, θ_2).

4.4 Topological Band Structures

4.4.1 SSH model

Consider the long chain polymer polyacetylene, with chemical formula $(\mathrm{CH})_x$, where x can be fairly large ($x \sim 10^4$ for example). In the *trans* form of $(\mathrm{CH})_x$, each C atom has two C neighbors and one H neighbor (see Fig. 4.16). The electronic structure of the carbon atom is $1\mathrm{s}^2\,2\mathrm{s}^2\,2\mathrm{p}^2$. The two $1\mathrm{s}$ electrons are tightly bound and are out of the picture. Three of the remaining $n = 2$ electrons form sp^2 planar hybrid bonding orbitals. This leaves one electron per carbon, which is denoted by π and has the symmetry of p_z, where backbone of the molecule is taken to lie in the (x, y) plane. If we were to model $(\mathrm{CH})_x$ by a one-dimensional tight binding model for these π-electrons, we'd write

$$H = -t \sum_{n=1}^{N} \sum_{\sigma=\pm} \left(c^\dagger_{n,\sigma}\, c_{n+1,\sigma} + c^\dagger_{n+1,\sigma}\, c_{n,\sigma} \right) = -2t \sum_{k,\sigma} \cos(ka)\, c^\dagger_{k,\sigma}\, c_{k,\sigma} \quad , \tag{4.132}$$

where σ is the spin polarization and where $c_{k,\sigma} = N^{-1/2} \sum_n e^{-ikna}\, c_{n,\sigma}$. We assume periodic boundary conditions $c_{n+N,\sigma} = c_{n,\sigma}$, which entails the mode quantization $k = 2\pi j/Na$ with

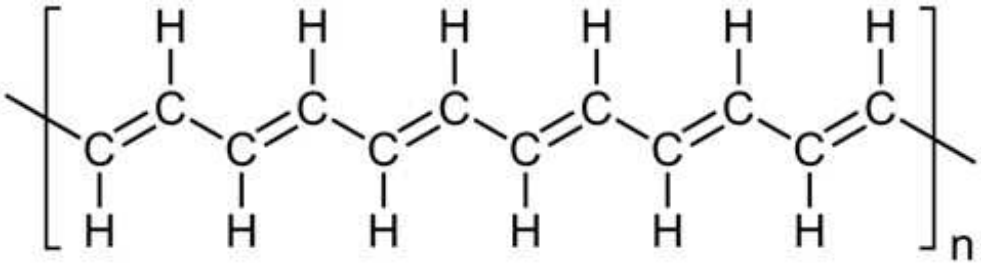


Figure 4.16: *trans*-polyacetylene, CH_x, consists of a backbone of C atoms with alternating single and double bonds.

$j \in \{1, \ldots, N\}$. In the thermodynamic limit $N \to \infty$ we may restrict $|k| \leqslant \frac{\pi}{a}$ to the first Brillouin zone.

The ground state is obtained by filling each of the $\frac{1}{2}N$ lowest energy states $\varepsilon_k = -2t\cos(ka)$ with a pair of $\uparrow$ and $\downarrow$ states, thus accounting for the N π-electrons present. Thus, with $\theta \equiv ka$, we have

$$E_0 = -4Nt \int_{-\pi/2}^{\pi/2} \frac{d\theta}{2\pi} \cos\theta = -\frac{4Nt}{\pi} \quad . \tag{4.133}$$

But there is a way for the system to lower its energy through *spontaneous dimerization*. Let's see how this works. We allow the nuclear centers of the carbon atoms to move a bit, and we write

$$H = -t \sum_{n=1}^{N} \Big(1 - \alpha(u_{n+1} - u_n)\Big)\Big(c^\dagger_{n,\sigma}\, c_{n+1,\sigma} + c^\dagger_{n+1,\sigma}\, c_{n,\sigma}\Big) + \sum_{n=1}^{N} \left(\frac{p_n^2}{2M} + \tfrac{1}{2}K(u_{n+1} - u_n)^2\right) \quad , \tag{4.134}$$

where u_n is the displacement of the $n^{\rm th}$ carbon atom and p_n is its momentum in the direction along the $(\mathrm{CH})_x$ backbone, and M is the atomic mass of carbon. The parameter α accounts for the exponential falloff of the π-electron wavefunctions from each nuclear center. We expect then that the hopping integral between atoms n and $n+1$ will be given by $t_{n,n+1} = t\, e^{-\alpha(u_{n+1} - u_n)}$, so $t_{n,n+1}$ increases if the distance $x_{n+1} - x_n = a + u_{n+1} - u_n$ decreases, *i.e.* if $u_{n+1} - u_n < 0$, and decreases if $u_{n+1} - u_n > 0$. The parameter α has dimensions of inverse length, and we presume that the lattice distortions are all weak, which licenses us to expand the exponential and write $t_{n,n+1} \approx t\,\big(1 - \alpha(u_{n+1} - u_n)\big)$. This is known as the Su-Schrieffer-Heeger (SSH) model[21].

For spontaneous dimerization, which breaks the lattice translation symmetry from $\mathbb{Z}_N$ to $\mathbb{Z}_{N/2}$, we write

$$u_n = (-1)^n\, \zeta + \delta u_n \quad . \tag{4.135}$$

[21] W. P. Su, J. R. Schrieffer, and A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698 (1979).

The phonon part of the Hamiltonian now becomes

$$H_{\rm ph} = \overbrace{\sum_{n=1}^{N}\left(\frac{p_n^2}{2M} + \tfrac{1}{2}K\,(\delta u_{n+1} - \delta u_n)^2\right)}^{H^0_{\rm ph}} + 2NK\zeta^2 + 4K\zeta\sum_n (-1)^n\,\delta u_n \quad . \tag{4.136}$$

We assume N to be even. We take as a trial state for the phonons the ground state of $H^0_{\rm ph}$, which is a product of harmonic oscillator states for each of the phonon modes with

$$H^0_{\rm ph} = \sum_k \hbar\omega_k\left(A_k^\dagger\,A_k + \tfrac{1}{2}\right) \quad , \tag{4.137}$$

where $\omega_k = 2(K/M)^{1/2}\,|\sin(\frac{1}{2}ka)|$ is the phonon dispersion and

$$A_k = \left(\frac{1}{2\hbar M\omega_k}\right)^{1/2}\hat{p}_k + \left(\frac{M\omega_k}{2\hbar}\right)^{1/2}\delta\hat{u}_k \tag{4.138}$$

is the (first quantized) phonon annihilation operator, with $\begin{pmatrix}\hat{p}_k\\ \delta\hat{u}_k\end{pmatrix} = N^{-1/2}\sum_n e^{-ikna}\begin{pmatrix}p_n\\ \delta u_n\end{pmatrix}$. At $T=0$, we have $E_0^{\rm ph} = \frac{1}{2}\sum_k\hbar\omega_k = 4N\hbar\sqrt{K/M}$. Note also that in the phonon ground state $|\,\Psi_0^{\rm ph}\,\rangle$ we have $\langle\,\Psi_0^{\rm ph}\,|\,\delta u_n\,|\,\Psi_0^{\rm ph}\,\rangle = 0$ for all n. If we take the expectation value of H from Eqn. 4.134 in the state $|\,\Psi_0^{\rm ph}\,\rangle$, we obtain the effective electronic Hamiltonian

$$\begin{aligned} H_{\rm eff} &= E_0^{\rm ph} + 2NK\zeta^2 - \sum_{n=1}^{N_{\rm c}}\sum_\sigma\Big(t_1\,a^\dagger_{n,\sigma}\,b_{n,\sigma} + t_2\,b^\dagger_{n,\sigma}\,a_{n+1,\sigma} + {\rm H.c.}\Big) \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \sum_{k,\sigma}\begin{pmatrix}a^\dagger_{k,\sigma} & b^\dagger_{k,\sigma}\end{pmatrix}\begin{pmatrix}0 & t_1 + t_2\,e^{-ikb}\\ t_1 + t_2\,e^{ikb} & 0\end{pmatrix}\begin{pmatrix}a_{k,\sigma}\\ b_{k,\sigma}\end{pmatrix} \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 + \sum_{k,\sigma}\left|t_1 + t_2\,e^{ikb}\right|\left(\gamma^\dagger_{+,k,\sigma}\,\gamma_{+,k,\sigma} - \gamma^\dagger_{-,k,\sigma}\,\gamma_{-,k,\sigma}\right) \quad , \end{aligned} \tag{4.139}$$

where $N_{\rm c} = \frac{1}{2}N$, $b = 2a$ is the unit cell size, $a_{n,\sigma} \equiv c_{2n-1,\sigma}$, $b_{n,\sigma} \equiv c_{2n,\sigma}$, and

$$t_{1,2} = \big(1 \mp 2\alpha\zeta\big)\,t \quad . \tag{4.140}$$

We see that there are two states for each wavevector k and spin polarization σ, labeled by indices $\pm$, with energies $\varepsilon_{\pm,k,\sigma} = \pm\big|t_1 + t_2\,e^{ikb}\big|$. At $T=0$, we fill the bottom $(-)$ band with electrons of both spin polarizations, and the total energy is then (substituting $\theta \equiv kb \in [-\pi,\pi]$)

$$\begin{aligned} E_0^{\rm tot} &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \frac{N}{2\pi}\int_{-\pi}^{\pi}d\theta\,\sqrt{t_1^2 + t_2^2 + 2t_1t_2\cos\theta} \\ &= 4N\hbar\sqrt{\frac{K}{M}} + 2NK\zeta^2 - \frac{4Nt}{\pi}\,\mathsf{E}(\kappa) \quad , \end{aligned} \tag{4.141}$$

where

$$\kappa = \frac{2\sqrt{t_1 t_2}}{|t_1 + t_2|} = \sqrt{1 - 4\alpha^2\zeta^2} \quad . \tag{4.142}$$

and where $\mathsf{E}(\kappa)$ is the complete elliptic integral of the second kind,

$$\mathsf{E}(\kappa) = \int_0^{\pi/2} d\vartheta \, \sqrt{1 - \kappa^2 \sin^2\vartheta} \quad . \tag{4.143}$$

We are interested in the case where $\alpha|\zeta| \ll 1$, *i.e.* small dimerization amplitude. Thus $\kappa \approx 1$, and in this limit we have[22]

$$\mathsf{E}(\kappa) = 1 + \tfrac{1}{2}\kappa'^2\Big(\ln(4/\kappa') - \tfrac{1}{2}\Big) + \ldots \quad , \tag{4.144}$$

where $\kappa' = \sqrt{1-\kappa^2} = 2\alpha\zeta$. We now have that the energy per site is given by

$$\frac{E_0^{\rm tot}}{N}(\zeta) = 4\hbar\sqrt{\frac{K}{M}} + 2K\zeta^2 - \frac{4t}{\pi} - \frac{8t}{\pi}\,\alpha^2\zeta^2 \ln\left(\frac{2}{\sqrt{e}\,\alpha\,\zeta}\right) + \ldots \quad . \tag{4.145}$$

Differentiating with respect to the dimerization parameter ζ, we obtain a minimum at

$$\zeta^* = \frac{2}{\sqrt{e}\,\alpha}\,\exp\left(-\frac{\pi K}{4\alpha^2 t}\right) \quad , \tag{4.146}$$

in the limit where $\alpha^2 t \ll K$. Thus, the system prefers to break the discrete translational symmetry of the one-dimensional lattice and to be spontaneously dimerized. The tendency for one-dimensional electronic systems coupled to phonons to spontaneously dimerize is known as the *Peierls instability*.

Topological aspects of the SSH model

We now suppress spin and investigate the following Hamiltonian,

$$\begin{aligned} H_0 &= -\sum_{n=1}^{N_{\rm c}} \Big(t_1\, a_n^\dagger\, b_n + t_2\, b_n^\dagger\, a_{n+1} + \text{H.c.} \Big) \\ &= -\sum_k \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} \begin{pmatrix} 0 & t_1 + t_2\, e^{-ikb} \\ t_1 + t_2\, e^{ikb} & 0 \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \quad , \end{aligned} \tag{4.147}$$

where the second and third lines are written in terms of second-quantized fermionic creation $(a^\dagger, b^\dagger)$ and annihilation (a, b) operators. This corresponds to a model with $N_{\rm c}$ unit cells where

[22] See Gradshteyn and Ryzhik, formula 8.114.3.

there are alternating strong (t_1) and weak (t_2) bonds, with periodic boundary conditions. The number of sites is $N = 2N_{\rm c}$. An arbitrary single particle eigenstate may be written

$$|\,\psi\,\rangle = \sum_{n=1}^{N_{\rm c}} \Big(A_n\, a_n^\dagger + B_n\, b_n^\dagger \Big)\,|\,0\,\rangle \quad . \tag{4.148}$$

In terms of the c-number coefficients $\{A_n, B_n\}$, the Schödinger equation may be written as the system

$$\begin{aligned} E A_n &= -t_2 B_{n-1} - t_1 B_n \\ E B_n &= -t_1 A_n - t_2 A_{n+1} \end{aligned} \quad , \tag{4.149}$$

with periodic boundary conditions $A_{n+N_{\rm c}} = A_n$ and $B_{n+N_{\rm c}} = B_n$. We can write this system in matrix form as

$$\begin{pmatrix} 0 & t_1 \\ t_2 & E \end{pmatrix} \begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = - \begin{pmatrix} E & t_2 \\ t_1 & 0 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} E^2 - t_1^2 & E\,t_2 \\ -E\,t_2 & -t_2^2 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad . \tag{4.150}$$

With translational invariance, we can demand

$$\begin{pmatrix} A_{n+1} \\ B_n \end{pmatrix} = z \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} \quad , \tag{4.151}$$

where $z = \exp(ikb)$. With this substitution in the above equation, we obtain the system

$$\begin{pmatrix} E^2 - t_1^2 - z\,t_1\,t_2 & E\,t_2 \\ -E\,t_2 & -t_2^2 - z\,t_1\,t_2 \end{pmatrix} \begin{pmatrix} A_n \\ B_{n-1} \end{pmatrix} = 0 \quad , \tag{4.152}$$

and setting the determinant to zero recovers the dispersion $E = \pm|t_1 + z\,t_2|$.

Now suppose we remove the link between the b orbital in unit cell $n = N_{\rm c}$ and the a orbital in unit cell $n = 1$. We now have

$$\begin{pmatrix} A_2 \\ B_1 \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} E^2 - t_1^2 & 0 \\ -E\,t_2 & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ B_{N_{\rm c}} \end{pmatrix} \quad , \quad \begin{pmatrix} A_{N_{\rm c}} \\ B_{N_{\rm c}-1} \end{pmatrix} = \frac{1}{t_1 t_2} \begin{pmatrix} 0 & -E\,t_2 \\ 0 & E^2 - t_1^2 \end{pmatrix} \begin{pmatrix} A_1 \\ B_{N_{\rm c}} \end{pmatrix} \quad , \tag{4.153}$$

as well as Eqn. 4.150 for $n \in \{2, \ldots, N_{\rm c} - 1\}$. We now show that there exist two eigenstates in the thermodynamic limit which are degenerate with $E = 0$ and which are exponentially localized in the vicinity of one of the ends of the chain provided $|r| < 1$, where $r = t_1/t_2$. With $E = 0$ we have from the first of the above equations that $A_2 = -rA_1$ and $B_1 = 0$. Iterating Eqn. 4.150, we then find $A_n = (-r)^{n-1} A_1$ and $B_n = 0$. In the limit $N_{\rm c} \to \infty$, so long as $|r| < 1$, we obtain a normalized wavefunction with $A_1 = \sqrt{1 - r^2}\, e^{i\alpha}$, where α is an arbitrary phase. A second zero mode is elicited by starting with the second of the above equations, which yields $B_{N_{\rm c}-n} = (-r)^n B_{N_{\rm c}}$ and $A_{N_{\rm c}-n} = 0$, which is normalized by setting $B_{N_{\rm c}} = \sqrt{1 - r^2}\, e^{i\beta}$. For $N_{\rm c}$ finite, these modes will mix and undergo level repulsion, resulting in two levels of opposite energy exponentially close to $E = 0$. Thus, for $|r| < 1$ there are two $E = 0$ edge states. For $|r| > 1$, there are no normalizable mid-gap (*i.e.* $E = 0$) modes. Of course, the midgap states *will* appear for $|r| > 1$ if we instead cut the bond between the a and b orbitals within the $n = 1$ cell.

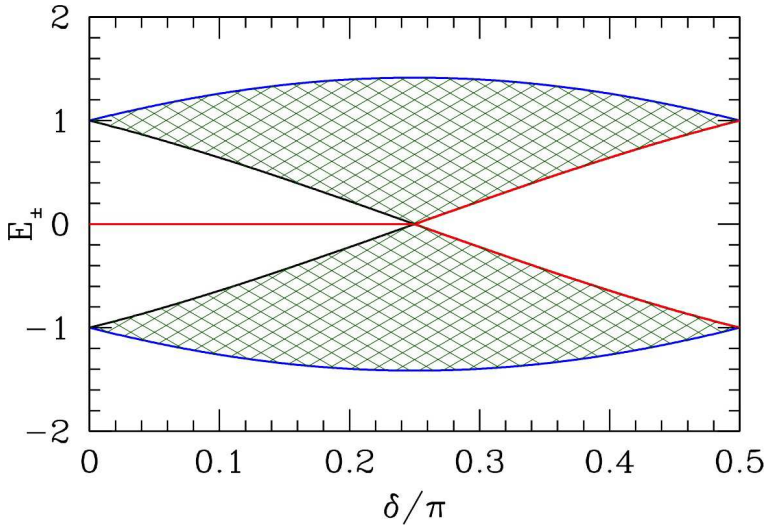


Figure 4.17: Spectrum of the SSH Hamiltonian on a finite chain with $N_{\rm c} \to \infty$ with $t_1 = t \sin\delta$ and $t_2 = t\cos\delta$. The two $E = 0$ edge states exist provided $|t_1| < |t_2|$, *i.e.* $|\delta| < \frac{\pi}{4}$.

4.4.2 Polarization and geometric phase

The individual cell functions for the Bloch states of the SSH model are given by the spinors

$$\vec{u}_\pm(k) = \begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix} \quad , \tag{4.154}$$

which are eigenstates of the Hamiltonian $H(k)$, *viz.*

$$\overbrace{-\begin{pmatrix} 0 & t(k) \\ t^*(k) & 0 \end{pmatrix}}^{H(k)} \overbrace{\begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix}}^{\vec{u}_\pm(k)} = E_\pm(k) \overbrace{\begin{pmatrix} u_{\mathsf{A}\pm}(k) \\ u_{\mathsf{B}\pm}(k) \end{pmatrix}}^{\vec{u}_\pm(k)} \quad , \tag{4.155}$$

where $t(k) \equiv t_1 + t_2\, e^{-ikb}$ and $E_\pm(k) = \pm|t(k)|$. We define the *polarization* $P_\pm$ of each band as

$$P_\pm = i \int\limits_{-\pi/b}^{\pi/b} \frac{dk}{2\pi} \, \langle\, \vec{u}_\pm(k) \,|\, \partial_k \,|\, \vec{u}_\pm(k) \,\rangle = \int\limits_{-\pi/b}^{\pi/b} \frac{dk}{2\pi} \, A_\pm(k) \quad , \tag{4.156}$$

where

$$A_\pm(k) = \left\langle\, \vec{u}_\pm(k) \,\middle|\, i\,\frac{\partial}{\partial k} \,\middle|\, \vec{u}_\pm(k) \,\right\rangle \tag{4.157}$$

is the *Berry connection*, which plays a role similar to an electromagnetic vector potential. However, the polarization is defined only up to an integer multiple, because if we make an allowed gauge transformation $|\,\vec{u}_\pm(k)\,\rangle \to e^{-i\varphi(k)}\,|\,\vec{u}_\pm(k)\,\rangle$, where $e^{-i\varphi(k)}$ is *single-valued* on the Brillouin

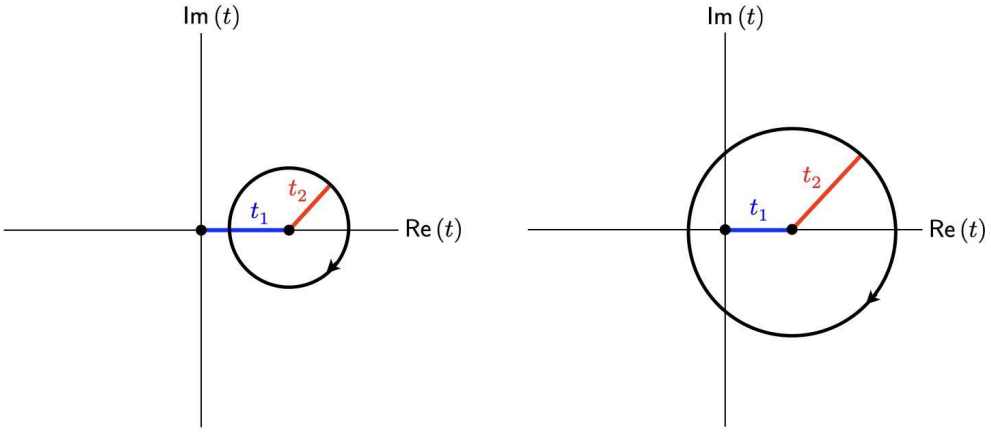


Figure 4.18: Winding of $t(k) = t_1 + t_2\, e^{-ikb}$ in the SSH model. Left: $|t_1| > |t_2|$, $W = 0$. Right: $|t_1| < |t_2|$, $W = -1$.

zone, then $P_\pm \to P_\pm + \big[\varphi(\pi/b) - \varphi(-\pi/b)\big]/2\pi = P_\pm + n$ where $n \in \mathbb{Z}$. Thus, it is the *phase* $\exp(2\pi i P_\pm)$ which is well-defined and gauge-invariant. We now compute this phase and show that it characterizes the two sectors $|r| > 1$ and $|r| < 1$, where $r \equiv t_1/t_2$.

Solving the Schrödinger equation for the cell functions, we obtain the solutions

$$u_{\mathsf{A}\pm}(k) = \frac{1}{\sqrt{2}} \quad , \qquad u_{\mathsf{B}\pm}(k) = \mp\frac{1}{\sqrt{2}}\,\frac{t^*(k)}{|t(k)|} = \mp\frac{1}{\sqrt{2}}\,e^{-i\Theta(k)} \quad , \tag{4.158}$$

where $\Theta(k) \equiv \arg t(k)$. Thus, $A_\pm(k) = \frac{1}{2}\frac{\partial\Theta}{\partial k}$ and

$$2\pi P_\pm = \frac{1}{2}\oint d\Theta = \pi W \quad , \tag{4.159}$$

where W is the integer *winding number* of the angle $\Theta(k)$ around the Brillouin zone $kb \in [-\pi, \pi]$. Examining the function $t(k) = t_1 + t_2\, e^{-ikb}$, we see that $\Theta(k)$ has winding $W = 0$ if $|t_1| > |t_2|$ and $W = -1$ if $|t_1| < |t_2|$. Thus, $\exp(2\pi i P_\pm) = +1$ for both $\pm$ bands in the nontopological sector where $|t_1| > |t_2|$, and $\exp(2\pi i P_\pm) = -1$ for both $\pm$ bands in the topological sector where $|t_1| < |t_2|$. Thus, the polarization phase neatly distinguishes the two sectors, which, as we saw in the previous section, also differ in their eigenspectra on open chains, where the topological sector exhibits two $E = 0$ edge states which are not present in the nontopological sector. The different winding sectors are depicted in Fig. 4.18.

Explicit breaking of translational symmetry

Suppose the A and B site ions are not both carbon, but are chemically different species. Then there will in general be a difference in the on-site energies for the individual π orbitals, and we

should write

$$H(k) = -\begin{pmatrix} m & t(k) \\ t^*(k) & -m \end{pmatrix} = -E_k \begin{pmatrix} \cos\theta_k & \sin\theta_k\, e^{i\phi_k} \\ \sin\theta_k\, e^{-i\phi_k} & -\cos\theta_k \end{pmatrix} \tag{4.160}$$

where $E_k = \sqrt{|t(k)|^2 + m^2}$ and

$$\cos\theta_k = \frac{m}{E_k} \quad , \quad \sin\theta_k\, e^{i\phi_k} = \frac{t(k)}{E_k} \quad . \tag{4.161}$$

The eigenvectors of $H(k)$ are

$$|\,\vec{u}_-(k)\,\rangle = \begin{pmatrix} \cos\frac{1}{2}\theta_k \\ \sin\frac{1}{2}\theta_k\, e^{i\phi_k} \end{pmatrix} \quad , \quad |\,\vec{u}_-(k)\,\rangle = \begin{pmatrix} \sin\frac{1}{2}\theta_k\, e^{i\phi_k} \\ -\cos\frac{1}{2}\theta_k \end{pmatrix} \quad , \tag{4.162}$$

with corresponding eigenvalues $\mp E_k$. We then have

$$\begin{aligned} A_\pm(k) &= \left\langle\, \vec{u}_\pm(k) \,\middle|\, i\,\frac{\partial}{\partial k} \,\middle|\, \vec{u}_\pm(k) \,\right\rangle = \pm\tfrac{1}{2}(1-\cos\theta_k)\,\frac{\partial\phi_k}{\partial k} \\ &= \frac{1}{2}\left(1 - \frac{m}{E_k}\right)\frac{\partial\,\mathsf{arg}\,t_k}{\partial k} \end{aligned} \tag{4.163}$$

Adopting again the definition $\Theta(k) = \phi_k = \mathsf{arg}\,t(k)$, note the following limits:

$$A_\pm(k\,;\,m=0) = \frac{1}{2}\,\frac{\partial\Theta}{\partial k} \quad , \quad A_\pm(k\,;\,m=-\infty) = \frac{\partial\Theta}{\partial k} \quad , \quad A_\pm(k\,;\,m=\infty) = 0 \quad . \tag{4.164}$$

For general m, we have that $\exp(2\pi i P_\pm)$ are no longer quantized at ± 1 and can take any value on the unit circle. The reason is that the Hamiltonian $H(k)$ in Eqn. 4.160 is intrinsically complex, which is a consequence of the breaking of inversion symmetry $\mathcal{I}$. You might think that the $m = 0$ case also breaks $\mathcal{I}$ as it is also complex, but in that case $H(k) = -\mathsf{Re}\,t(k)\,X + \mathsf{Im}\,t(k)\,Y$ is unitarily equivalent to a real Hamiltonian by rotating by $\frac{1}{2}\pi$ about the $\hat{\boldsymbol{x}}$ axis in internal space, *i.e.* by writing $H'(k) = U^\dagger H(k)\,U$ with $U = 2^{-1/2}(1+iX)$, whence $H'(k) = -\mathsf{Re}\,t(k)\,X + \mathsf{Im}\,t(k)\,Z$. When $m \to \pm\infty$, we obtain a trivial topology, because $\exp(2\pi i P_\pm) = +1$ in both limits[23]. Note also that for $m = 0$ the center of each link is a valid center of inversion symmetry, which is not the case when $m \neq 0$.

What happens to the edge states? Clearly they must split in energy because the sites on either end are members of different sublattices. Since the edge states we obtained previously live entirely on the A and B sublattices, respectively, their energies when $m \neq 0$ are, simply, $E_{\rm edge} = \pm m$. The bulk dispersion takes the form $E_{\rm bulk} = \pm E_k = \pm\sqrt{|t(k)|^2 + m^2}$. The edge states once again join the bulk spectrum when $|t_1| = |t_2|$, where $E_{k=\pi/b} = \pm m$.

[23] When $m = -\infty$ and $A_\pm(k) = \partial\Theta/\partial k$, we have $P_\pm = W$ rather than $P_\pm = \frac{1}{2}W$ as in the $m = 0$ case we studied previously, hence $\exp(2\pi i P_\pm) = +1$.

4.4.3 Domain wall states in the Dirac equation

Dirac Hamiltonians and Clifford algebras

A famous exchange (though perhaps somewhat apocryphal)[24]:

> Bohr to Dirac (∼1928): "What are you working on, Mr. Dirac?"
>
> Dirac to Bohr: "I am trying to take the square root of something."

What was Dirac trying to take the square root of? The Klein-Gordon equation, apparently:

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 + \frac{m^2c^2}{\hbar^2}\right)\psi = 0 \quad . \tag{4.165}$$

Dirac didn't like the fact that the KG equation is second order in space and time. The famous Dirac Hamiltonian is given by

$$H = c\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta mc^2 \quad , \tag{4.166}$$

where $\{\alpha_1, \alpha_2, \alpha_3, \beta\}$ are mutually anticommuting Hermitian *matrices* that are elements of a *Clifford algebra*, which is an associative algebra consisting of the identity (1) and d elements Γ^μ with $\mu \in \{1, \ldots, d\}$, each with $\mathrm{Tr}\,\Gamma^\mu = 0$, and satisfying the relations

$$\left\{\Gamma^\mu, \Gamma^\nu\right\} = 2\delta^{\mu\nu} \quad , \tag{4.167}$$

where $\{\bullet, \bullet\}$ is the anticommutator[25]. Note that $(\Gamma^\mu)^2 = 1$ for each element of the algebra. This can be accomplished with matrices of rank $2^{[d/2]}$, where $[d/2]$ is the greatest integer less than or equal to $\frac{1}{2}d$, in the following manner. Starting with $d = 2$, we may take $\Gamma^1 = X$ and $\Gamma^2 = Y$, where X and Y are the Pauli matrices σ^x and σ^y. For $d = 3$, we add $\Gamma^3 = -i\Gamma^1\Gamma^2 = Z$. For $d = 4$, take

$$\Gamma^1 = X \otimes \mathbb{I} = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \\ +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \end{pmatrix} \quad , \quad \Gamma^2 = Y \otimes \mathbb{I} = \begin{pmatrix} 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -i \\ +i & 0 & 0 & 0 \\ 0 & +i & 0 & 0 \end{pmatrix}$$
$$\Gamma^3 = Z \otimes X = \begin{pmatrix} 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad , \quad \Gamma^4 = Z \otimes Y = \begin{pmatrix} 0 & -i & 0 & 0 \\ +i & 0 & 0 & 0 \\ 0 & 0 & 0 & +i \\ 0 & 0 & -i & 0 \end{pmatrix} \quad , \tag{4.168}$$

[24] See P. A. M. Dirac and the Discovery of Quantum Mechanics by Kurt Gottfried (2010).

[25] More generally, we can take $d = t+s$ and $\{\Gamma^\mu, \Gamma^\nu\} = 2\eta^{\mu\nu}$, where $\eta^{\mu\nu} = 0$ if $\mu \neq \nu$, $\eta^{\mu\nu} = +1$ if $\mu = \nu \in \{1, \ldots, t\}$, and $\eta^{\mu\nu} = -1$ if $\mu = \nu \in \{t+1, \ldots, d\}$, which define the Clifford algebra $\mathcal{C}_{t,s}$. We restrict our attention to the Clifford algebras $\mathcal{C}_{d,0}$.

where $\mathbb{I}$ is the 2×2 unit matrix. For $d = 5$, we add

$$\Gamma^5 = -\Gamma^1\Gamma^2\Gamma^3\Gamma^4 = Z \otimes Z = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} \quad . \tag{4.169}$$

In general, for $d = 2k$ or $d = 2k + 1$, we may write

$$\begin{aligned} \Gamma^1 &= \overbrace{X \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I}}^{k \text{ terms}} & \Gamma^{2k-1} &= Z \otimes Z \otimes \cdots \otimes Z \otimes X \\ \Gamma^2 &= Y \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I} & \Gamma^{2k} &= Z \otimes Z \otimes \cdots \otimes Z \otimes Y \\ \Gamma^3 &= Z \otimes X \otimes \cdots \otimes \mathbb{I} \otimes \mathbb{I} & \Gamma^{2k+1} &= Z \otimes Z \otimes \cdots \otimes Z \otimes Z \, . \end{aligned} \tag{4.170}$$

The last of these pertains when $d = 2k + 1$ and is equivalent to

$$\Gamma^{2k+1} = (-i)^k \, \Gamma^1 \, \Gamma^2 \cdots \Gamma^{2k} \, . \tag{4.171}$$

Now you know something about Clifford algebras!

A very nice thing about Clifford algebras is that if we write

$$H = \boldsymbol{a} \cdot \boldsymbol{\Gamma} = a_\mu \, \Gamma^\mu \quad , \tag{4.172}$$

then

$$H^2 = a_\mu \, a_\nu \, \Gamma^\mu \, \Gamma^\nu = \tfrac{1}{2} \, a_\mu \, a_\nu \left\{ \Gamma^\mu, \Gamma^\nu \right\} = \boldsymbol{a}^2 \quad . \tag{4.173}$$

Since both H and H^2 are of rank 2^k, where $d = 2k$ or $d = 2k+1$, we have that the spectrum of H must consist of a 2^{k-1}-fold degenerate set of eigenvalues $E_- = -|\boldsymbol{a}|$, and a 2^{k-1}-fold degenerate set of eigenvalues $E_+ = +|\boldsymbol{a}|$.

Note that the dimension of the space of rank-n Hermitian matrices is n^2 (n independent real elements along the diagonal and $\frac{1}{2}n(n-1)$ independent complex elements above the diagonal). Thus for $n = 2^k$ we require 2^{2k} parameters to describe the most general Hamiltonian, whereas $H = \boldsymbol{a} \cdot \boldsymbol{\Gamma}$ is specified only $2k$ or $2k + 1$ parameters (or $2k + 2$ if we include real multiples of the identity). You may have noticed that the matrix $Z \otimes \mathbb{I} \otimes \cdots \otimes \mathbb{I}$, for example, is missing among the list Eqn. 4.170 for all $k > 1$. We can use the matrices in the Clifford algebra to build a full set of 2^{2k} independent Hermitian matrices, which will no longer be mutually anticommuting, by taking various products of matrices in the Clifford algebra. For example, with $k = 2$ and $d = 2k+1 = 5$ we have only five matrices in the CA but 15 independent traceless 4×4 Hermitian matrices in all. The missing 10 matrices are given by $\Gamma^{\mu\nu} = -i\Gamma^\mu\Gamma^\nu$ for $\mu < \nu$.

But we digress! Let us get back to the Dirac equation! For the Dirac Hamiltonian $H = \boldsymbol{a}(\boldsymbol{p}) \cdot \boldsymbol{\Gamma}$ with $\boldsymbol{a}(\boldsymbol{p}) = (p_x, p_y, p_z, m, 0)$, we have two massive Dirac cones with

$$E_\pm(\boldsymbol{p}) = \pm\sqrt{c^2\boldsymbol{p}^2 + m^2c^4} \quad . \tag{4.174}$$

Each energy level is doubly degenerate, corresponding to two electron spin polarizations. The gamma matrices are taken from the $k = 2$ Clifford algebra and are of rank four. In one space dimension, $H(p) = cpX + mc^2Z$ where X and Z are Pauli matrices (*i.e.* $k = 1$ Clifford algebra).

Domain wall bound states in $d = 1$

The bulk dispersion and ground state energy of the massive Dirac Hamiltonian are even functions of m. What happens if we continuously interpolate between negative and positive values of m? Consider the $d = 1$ Hamiltonian

$$H = -i\hbar c\, X\, \frac{\partial}{\partial x} + m(x)\, c^2\, Z = \begin{pmatrix} m(x)\, c^2 & -i\hbar c\, \partial_x \\ -i\hbar c\, \partial_x & -m(x)\, c^2 \end{pmatrix} \quad . \tag{4.175}$$

We have in mind a $m(x)$ smoothly interpolating between $m(-\infty) = -m$ and $m(+\infty) = +m$, but we can easily generalize to the case where $m(-\infty) = m_1$ and $m(+\infty) = m_2$. Over most of $x \in \mathbb{R}$, the mass $m(x)$ is roughly constant with either $m(x) = m_1$ or $m(x) = m_2$. In these regions, the bulk dispersion $E_{\pm,k} = \pm\sqrt{(\hbar ck)^2 + m_{1,2}^2\, c^4}$ holds. We now show that when the product $m_1 m_2 < 0$ that there is a precise zero energy eigenstate localized along the interface between the two regions[26]. Setting $E = 0$, the Schrödinger equation $H\vec{\psi}(x) = E\vec{\psi}(x)$ can be written as

$$\frac{\partial\vec{\psi}(x)}{\partial x} = -\frac{c}{\hbar}\, m(x)\, Y\, \vec{\psi}(x) \quad . \tag{4.176}$$

This requires that the spinor component of $\vec{\psi}(x)$ must be an eigenstate of Y, so we write

$$\vec{\psi}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i\eta \end{pmatrix} f(x) \quad , \tag{4.177}$$

where $\eta = \pm 1$, hence $Y\vec{\psi}(x) = \eta\,\vec{\psi}(x)$. Thus,

$$\frac{\partial \ln f(x)}{\partial x} = -\eta\, \frac{c}{\hbar}\, m(x) \qquad \Rightarrow \qquad f(x) = A\, \exp\left\{ -\,\eta\, \frac{c}{\hbar} \int_0^x dx'\, m(x') \right\} \quad , \tag{4.178}$$

where A is a normalization constant. In order to have a normalizable solution, though, we must have that $m(-\infty)\, m(+\infty) < 0$, in which case we choose

$$\eta = \mathsf{sgn}\big[m(+\infty) - m(-\infty)\big] \quad . \tag{4.179}$$

Helical edge states in $d = 2$

Let's now proceed to $d = 2$ and write

$$H = -i\hbar c\, \Gamma^1\, \frac{\partial}{\partial x} - i\hbar c\, \Gamma^2\, \frac{\partial}{\partial y} + m(x)\, c^2\, \Gamma^4 \quad . \tag{4.180}$$

[26] Two efficiently brief and readable texts on the general subject of topological bands in condensed matter, see S.-Q. Shen, *Topological Insulators* (Springer, 2012) or J. K. Asbóth, L. Oroszlány, and A. Pályi, *A Short Course on Topological Insulators* (Springer, 2016). More extensive treatments may be found in the now standard texts by Bernevig and Hughes (Princeton, 2013) and by Vanderbilt (Cambridge, 2018).

The mass $m(x)$ is presumed to be dependent only on the x coordinate, again interpolating between positive and negative values at $x = \pm\infty$. The wavefunction $\vec{\psi}(x,y)$ is now a four component spinor. We assume that $\vec{\psi}(x,y)$ corresponds to a plane wave in the y-direction, *i.e.* parallel to the domain wall. Note that $\Gamma^2 = Y \otimes \mathbb{I}$ has the following normalized eigenvectors:

$$\vec{\xi}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\i\\0\end{pmatrix} \quad , \quad \vec{\xi}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\i\end{pmatrix} \quad , \quad \vec{\xi}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-i\\0\end{pmatrix} \quad , \quad \vec{\xi}_4 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\-i\end{pmatrix} \quad , \tag{4.181}$$

where $\Gamma^2\,\vec{\xi}_{1,2} = +\vec{\xi}_{1,2}$ and $\Gamma^2\,\vec{\xi}_{3,4} = -\vec{\xi}_{3,4}$. We will also need the following, which you can easily check with $\Gamma^1 = X \otimes \mathbb{I}$ and $\Gamma^4 = Z \otimes Y$:

$$\begin{aligned}
&\Gamma^1\,\vec{\xi}_1 = +i\,\vec{\xi}_3 \quad , \quad \Gamma^1\,\vec{\xi}_2 = +i\,\vec{\xi}_4 \quad , \quad \Gamma^1\,\vec{\xi}_3 = -i\,\vec{\xi}_1 \quad , \quad \Gamma^1\,\vec{\xi}_4 = -i\,\vec{\xi}_2 \\
&\Gamma^4\,\vec{\xi}_1 = +i\,\vec{\xi}_4 \quad , \quad \Gamma^4\,\vec{\xi}_2 = -i\,\vec{\xi}_3 \quad , \quad \Gamma^4\,\vec{\xi}_3 = +i\,\vec{\xi}_2 \quad , \quad \Gamma^4\,\vec{\xi}_4 = -i\,\vec{\xi}_1 \quad .
\end{aligned} \tag{4.182}$$

The domain wall bound states are of two types, which we call I and II:

$$\begin{aligned}
\vec{\psi}_{\rm I}(x,y) &= A\,f(x)\,e^{ik_y y}\left(\alpha\,\vec{\xi}_1 + \beta\,\vec{\xi}_2\right) \\
\vec{\psi}_{\rm II}(x,y) &= B\,g(x)\,e^{ik_y y}\left(\gamma\,\vec{\xi}_3 + \delta\,\vec{\xi}_4\right) \quad .
\end{aligned} \tag{4.183}$$

For type I solutions, we write $E = \hbar c k_y$, in which case insertion into the Schrödinger equation $H\vec{\psi} = E\vec{\psi}$ yields the equation

$$\hbar c\left(\alpha\,\vec{\xi}_3 + \beta\,\vec{\xi}_4\right)\partial_x f + m(x)\,c^2\left(-i\beta\,\vec{\xi}_3 + i\alpha\,\vec{\xi}_4\right) f = 0 \quad . \tag{4.184}$$

The solvability condition here requires that the ratios of the $\vec{\xi}_4$ and $\vec{\xi}_3$ coefficients inside each of the round brackets must be the same, *i.e.* $\beta/\alpha = (-i\alpha)/(i\beta) = -\alpha/\beta$. Thus $\beta^2 = -\alpha^2$ and we may take

$$\begin{aligned}
(\alpha,\beta) = (1,+i) \;:\; \partial_x \ln f &= -\frac{c}{\hbar}\,m(x) \\
(\alpha,\beta) = (1,-i) \;:\; \partial_x \ln f &= +\frac{c}{\hbar}\,m(x) \quad .
\end{aligned} \tag{4.185}$$

For type II solutions, we write $E = -\hbar c k_y$, whence

$$-\hbar c\left(\gamma\,\vec{\xi}_1 + \delta\,\vec{\xi}_2\right)\partial_x g + m(x)\,c^2\left(-i\delta\,\vec{\xi}_1 + i\gamma\,\vec{\xi}_2\right) g = 0 \quad , \tag{4.186}$$

and we have the following possibilities:

$$\begin{aligned}
(\gamma,\delta) = (1,+i) \;:\; \partial_x \ln g &= +\frac{c}{\hbar}\,m(x) \\
(\gamma,\delta) = (1,-i) \;:\; \partial_x \ln g &= -\frac{c}{\hbar}\,m(x) \quad .
\end{aligned} \tag{4.187}$$

We therefore have the following solutions:

- Type I : $E = +\hbar c k_y$, $\Gamma^2\,\vec{\psi} = +\vec{\psi}$:

$$\begin{aligned}\vec{\psi}_{\rm Ia}(x,y) &= A\,e^{ik_y y}\,\exp\left\{-\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_1 + i\vec{\xi}_2\big)\\ \vec{\psi}_{\rm Ib}(x,y) &= A\,e^{ik_y y}\,\exp\left\{+\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_1 - i\vec{\xi}_2\big)\end{aligned} \tag{4.188}$$

  and

- Type II: $E = -\hbar c k_y$, $\Gamma^2\,\vec{\psi} = -\vec{\psi}$:

$$\begin{aligned}\vec{\psi}_{\rm IIa}(x,y) &= B\,e^{ik_y y}\,\exp\left\{+\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_3 + i\vec{\xi}_4\big)\\ \vec{\psi}_{\rm IIb}(x,y) &= B\,e^{ik_y y}\,\exp\left\{-\frac{c}{\hbar}\int\limits_0^x dx'\,m(x')\right\}\big(\vec{\xi}_3 - i\vec{\xi}_4\big)\quad .\end{aligned} \tag{4.189}$$

Thus, if $m(-\infty) < 0 < m(+\infty)$, the normalizable solutions are $\vec{\psi}_{\rm Ia}$ and $\vec{\psi}_{\rm IIb}$, while for the case $m(-\infty) > 0 > m(+\infty)$, the normalizable solutions are $\vec{\psi}_{\rm Ib}$ and $\vec{\psi}_{\rm IIa}$. Note that the time-dependences are as follows:

$$\begin{aligned}\text{type I}\ :\ \vec{\psi}_{\rm I}(x,y) \propto f(x)\,e^{ik_y(y-ct)} \quad &\Rightarrow \quad \text{up-mover, } v_y = +c\\ \text{type II}\ :\ \vec{\psi}_{\rm II}(x,y) \propto g(x)\,e^{ik_y(y+ct)} \quad &\Rightarrow \quad \text{down-mover, } v_y = -c\quad .\end{aligned} \tag{4.190}$$

Note that the product of $\mathsf{sgn}(v_y)$ and the eigenvalue η of Γ^2 is $\chi = \eta\ \mathsf{sgn}(v_y) = +1$ in both cases. We call χ the *chirality*.

It should now come as no surprise that for the $d = 3$ Dirac Hamiltonian

$$H = -i\hbar c\,\Gamma^1\frac{\partial}{\partial x} - i\hbar c\,\Gamma^2\frac{\partial}{\partial y} - i\hbar c\,\Gamma^3\frac{\partial}{\partial z} + m(x)\,c^2\,\Gamma^4 \tag{4.191}$$

that we obtain two chiral domain-wall surface states with $E = \pm\hbar c\sqrt{k_y^2 + k_z^2}$.

4.4.4 The adiabatic theorem and Berry's phase

Consider a Hamiltonian $H(\boldsymbol{\lambda})$ dependent on a set of parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_K\}$, and let $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ satisfy the time-independent Schrödinger equation,

$$H(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = E_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \quad . \tag{4.192}$$

Now let $\boldsymbol{\lambda}(t)$ be continuously time-dependent, and consider the time-*dependent* Schrödinger equation

$$i\hbar\frac{d}{dt}\,|\,\Psi(t)\,\rangle = H\big(\boldsymbol{\lambda}(t)\big)\,|\,\Psi(t)\,\rangle \quad . \tag{4.193}$$

The adiabatic theorem states that if $\boldsymbol{\lambda}(t)$ evolves extremely slowly, then each solution $|\,\Psi_n(t)\,\rangle$ is proportional to $|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$, with

$$\big|\,\Psi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle = \exp\big(i\gamma_n(t)\big)\,\exp\!\left(-\,\frac{i}{\hbar}\int^t\! dt'\;E_n\big(\boldsymbol{\lambda}(t')\big)\right)\big|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle \quad , \tag{4.194}$$

with corrections which vanish in the limit $|\dot{\boldsymbol{\lambda}}|/|\boldsymbol{\lambda}| \to 0$. Taking the time derivative and then the overlap with the bra vector $\langle\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,|$, one obtains the result

$$\frac{d\gamma_n(t)}{dt} = i\,\big\langle\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big|\,\frac{d}{dt}\,\big|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\big\rangle = \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})\cdot\frac{d\boldsymbol{\lambda}}{dt} \equiv \mathcal{A}_n(t) \quad , \tag{4.195}$$

where

$$\mathcal{A}_n^\mu(\boldsymbol{\lambda}) = i\,\big\langle\,\varphi_n(\boldsymbol{\lambda})\,\big|\,\frac{\partial}{\partial\lambda_\mu}\,\big|\,\varphi_n(\boldsymbol{\lambda})\,\big\rangle \tag{4.196}$$

is the *Berry connection*. We have already met with such a quantity in Eqn. 4.157 above. Note that $\mathcal{A}_n^\mu(\boldsymbol{\lambda})$ is real. In particular, if $\boldsymbol{\lambda}(t)$ traverses a closed loop $\mathcal{C}$ with infinitesimal speed, then the wavefunction $|\,\Psi_n(t)\,\rangle$ will accrue a *geometric phase* $\gamma_n(\mathcal{C})$, given by

$$\gamma_n(\mathcal{C}) = \oint\limits_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{4.197}$$

also called *Berry's phase*[27].

In the adiabatic limit, the dynamical phase $\hbar^{-1}\int^t dt'\;E_n\big(\boldsymbol{\lambda}(t')\big)$ becomes very large whenever $\langle E_n\rangle \neq 0$, because the path $\boldsymbol{\lambda}(t)$ is traversed very slowly. We may remove this dynamical phase by defining the Hamiltonian

$$\widetilde{H}_n(\boldsymbol{\lambda}) \equiv H(\boldsymbol{\lambda}) - E_n(\boldsymbol{\lambda}) \quad . \tag{4.198}$$

We define $|\,\widetilde{\Psi}_n(t)\,\rangle$ as the solution to the Schrödinger equation

$$i\hbar\,\frac{d}{dt}\,|\,\widetilde{\Psi}_n(t)\,\rangle = \widetilde{H}_n\big(\boldsymbol{\lambda}(t)\big)\,|\,\widetilde{\Psi}_n(t)\,\rangle \tag{4.199}$$

in the adiabatic limit. The adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ are the same as before, but now satisfy the zero energy condition $\widetilde{H}_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = 0$. Clearly $|\,\widetilde{\Psi}_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle = \exp\big(i\gamma_n(t)\big)\,|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$

[27] See M. V. Berry, *Proc. Roy. Soc. A* **392**, 45 (1984).

and the dynamical phase has been removed. However, note that the geometrical phase γ_n does not depend on the elapsed time, but only on the path traversed, *viz.*

$$\gamma_n = \gamma_n(\boldsymbol{\lambda}) = \int_{\boldsymbol{\lambda}_0}^{\boldsymbol{\lambda}} d\boldsymbol{\lambda}' \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}') \quad , \tag{4.200}$$

where $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(0)$, and where the integral is taken along the path in traversed by $\boldsymbol{\lambda}$.

4.4.5 Connection, curvature, and Chern numbers

The mathematical structure underlying this discussion is that of the *Hermitian line bundle*, the ingredients of which are (i) a *base space* $\mathcal{M}$ which is a topological manifold; this is the parameter space for $\boldsymbol{\lambda}$, and (ii) to each point $\boldsymbol{\lambda} \in \mathcal{M}$ is associated a *fiber* which is the adiabatic wavefunction $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathcal{H}$, which is a complex one-dimensional subspace of some Hilbert space $\mathcal{H}$. As $\boldsymbol{\lambda}$ moves around the base space $\mathcal{M}$, the fiber twists around. The adiabatic theorem furnishes us with a way of defining *parallel transport* of $|\,\widetilde{\Psi}_n(\boldsymbol{\lambda})\,\rangle$ along the curve $\mathcal{C}$[28]. The object $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ is the *connection* and the geometric phase $\gamma(\mathcal{C})$ is the *holonomy* of the connection[29]. As a holonomy, $\gamma(\mathcal{C})$ depends only on the curve $\mathcal{C}$ and not on where along the curve one starts.

The *curvature tensor* for the bundle is given by

$$\begin{aligned} \Omega_n^{\mu\nu}(\boldsymbol{\lambda}) &= \frac{\partial \mathcal{A}_n^\nu}{\partial \lambda_\mu} - \frac{\partial \mathcal{A}_n^\mu}{\partial \lambda_\nu} \\ &= i \left\langle \frac{\partial \varphi_n}{\partial \lambda_\mu} \,\middle|\, \frac{\partial \varphi_n}{\partial \lambda_\nu} \right\rangle - i \left\langle \frac{\partial \varphi_n}{\partial \lambda_\nu} \,\middle|\, \frac{\partial \varphi_n}{\partial \lambda_\mu} \right\rangle \quad . \end{aligned} \tag{4.201}$$

Using completeness of the $|\,\varphi_n\,\rangle$ basis, we may write the curvature tensor as

$$\Omega_n^{\mu\nu}(\boldsymbol{\lambda}) = i \sum_l{}' \frac{\left\langle \varphi_n \,\middle|\, \frac{\partial H}{\partial \lambda_\mu} \,\middle|\, \varphi_l \right\rangle \left\langle \varphi_l \,\middle|\, \frac{\partial H}{\partial \lambda_\nu} \,\middle|\, \varphi_n \right\rangle - (\mu \leftrightarrow \nu)}{(E_n - E_l)^2} \quad , \tag{4.202}$$

where the prime on the sum indicates that the term $l = n$ is to be excluded. We see that in this formulation the curvature tensor is actually independent of any phase convention for the adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$. So long as the adiabatic eigenstate $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ remains nondegenerate, the denominator in Eqn. 4.202 remains nonzero, hence the curvature tensor $\Omega(\boldsymbol{\lambda})$ is nonsingular. The same cannot be said about the connection $\boldsymbol{\mathcal{A}}(\boldsymbol{\lambda})$, however, because it is

[28] Note that $|\,\Psi_n(t)\,\rangle$, which depends explicitly on elapsed time and not solely on the position $\boldsymbol{\lambda}$ along its trajectory, can not be said to be parallel transported along any curve.

[29] See B. Simon, *Phys. Rev. Lett.* **51**, 2167 (1983).

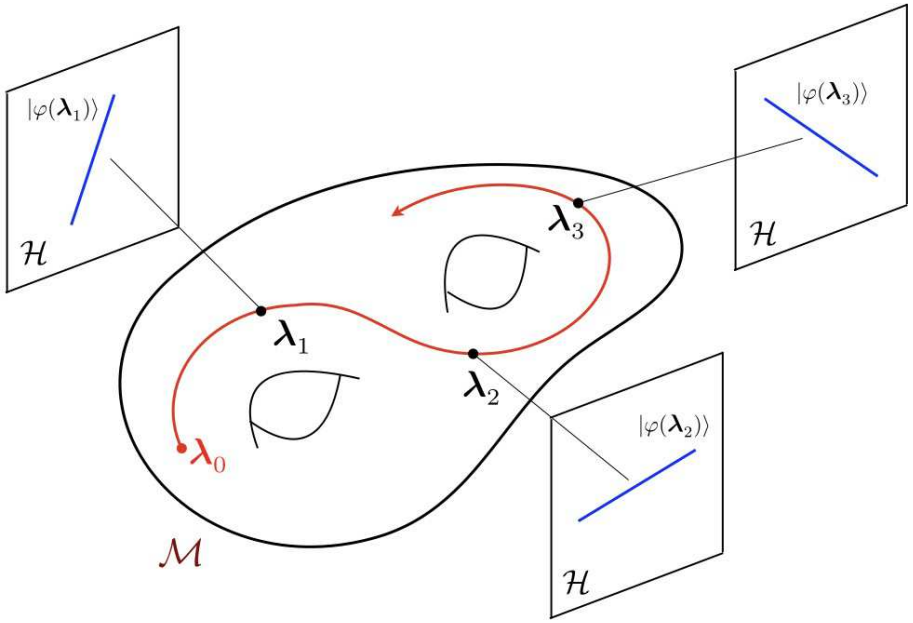


Figure 4.19: A Hermitian line bundle consists of a base space $\mathcal{M}$ and a fiber $|\,\varphi(\boldsymbol{\lambda})\,\rangle$ which twists as the point $\boldsymbol{\lambda}$ moves around the base space.

gauge-covariant. This means that if we multiply the adiabatic wavefunctions by phases, with $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \to \exp(if_n(\boldsymbol{\lambda}))\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$, the connection changes accordingly, with

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \to \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) - \frac{\partial f_n(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \quad . \tag{4.203}$$

How can we fix a gauge in order to give unambiguous meaning to $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$? One way might be to demand that the adiabatic wavefunction amplitude be real and positive at some particular point in space $\boldsymbol{r}_0$, *i.e.* $\langle\, \boldsymbol{r}_0\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathbb{R}_+$ for all $\boldsymbol{\lambda} \in \mathcal{M}$. For lattice-based models, where the adiabatic wavefunction is a vector of amplitudes for each orbital and each site within the appropriate unit cell, we could similarly demand that one of these amplitudes be real and positive. *This prescription fails if there exists a value of* $\boldsymbol{\lambda}$ *for which this wavefunction amplitude vanishes.*

As we are about to discover, the integral of the curvature over a two-dimensional base space is a topological invariant, meaning that it remains fixed (and indeed quantized) under continuous deformations of the Hamiltonian $H(\boldsymbol{\lambda})$. Using Stokes' theorem, we can turn an area integral of the curvature into line integrals of the connection. However, having chosen a particular gauge for the adiabatic wavefunctions, it may be that the connection is singular at certain points. Therefore the line integrals cannot be completely collapsed, and we obtain the result

$$\int\limits_{\mathcal{M}} d^2\lambda\; \Omega_n^{12}(\boldsymbol{\lambda}) = -\sum_i \oint\limits_{\mathcal{C}_i} d\boldsymbol{\lambda} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{4.204}$$

where the loop $\mathcal{C}_i$ encloses the $i^{\rm th}$ singularity $\boldsymbol{\lambda}_i$ of the connection in a counterclockwise man-

ner[30]. This is the generalization to Hermitian line bundles of the index formula in Eqn. 4.310 for the Gauss-Bonnet theorem. Quantization follows by writing $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = e^{iq_i\zeta(\boldsymbol{\lambda}-\boldsymbol{\lambda}_i)}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ in the vicinity of $\boldsymbol{\lambda}=\boldsymbol{\lambda}_i$, where q_i is an integer and

$$\zeta(\boldsymbol{\lambda}-\boldsymbol{\lambda}_i) = \tan^{-1}\left(\frac{\lambda_2-\lambda_{i,2}}{\lambda_1-\lambda_{i,1}}\right) \quad . \tag{4.205}$$

The integers q_i are chosen to 'unwind' the singularity at each $\boldsymbol{\lambda}_i$, so as to make the gauge transformed connection $\widetilde{\boldsymbol{\mathcal{A}}}{}^{\mu}_n(\boldsymbol{\lambda}) \equiv i\,\langle\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,|\,\boldsymbol{\nabla}_{\boldsymbol{\lambda}}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ nonsingular[31]. We then obtain

$$C_n \equiv \frac{1}{2\pi}\int\limits_{\mathcal{M}} d^2\lambda\;\Omega^{12}_n(\boldsymbol{\lambda}) = \sum_i q_i \quad . \tag{4.206}$$

Thus $C_n \in \mathbb{Z}$ is the *Chern number* of the Hermitian line bundle corresponding to the adiabatic wavefunction $|\,\varphi_n\,\rangle$.

The simplest nontrivial example is that of a spin-$\frac{1}{2}$ object in a magnetic field $\boldsymbol{B}(t)$, with

$$H(t) = g\mu_{\rm B}\boldsymbol{B}\cdot\boldsymbol{\sigma} = g\mu_{\rm B}B\begin{pmatrix}\cos\theta & \sin\theta\exp(-i\phi)\\ \sin\theta\exp(i\phi) & \cos\theta\end{pmatrix} \quad , \tag{4.207}$$

where $\boldsymbol{B} = B\,\hat{\boldsymbol{n}}$ is the adiabatic parameter which varies extremely slowly in time. The adiabatic wavefunctions are

$$|\,\varphi_+(\theta,\phi)\,\rangle = \begin{pmatrix}u\\ v\end{pmatrix} \qquad , \qquad |\,\varphi_-(\theta,\phi)\,\rangle = \begin{pmatrix}-\bar{v}\\ \bar{u}\end{pmatrix} \quad , \tag{4.208}$$

where $\hat{\boldsymbol{n}} = (\sin\theta\cos\phi\,,\,\sin\theta\sin\phi\,,\,\cos\theta)$, $u = \cos(\frac{1}{2}\theta)$, and $v = \sin(\frac{1}{2}\theta)\exp(i\phi)$. The energy eigenvalues are $E_\pm = \pm g\mu_{\rm B}B$. We now compute the connections,

$$\begin{aligned}\mathcal{A}_+ &= i\,\langle\,\varphi_+\,|\,\frac{d}{dt}\,|\,\varphi_+\,\rangle = i\big(\bar{u}\dot{u}+\bar{v}\dot{v}\big) = -\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = -\tfrac{1}{2}\,\dot{\omega}\\ \mathcal{A}_- &= i\,\langle\,\varphi_-\,|\,\frac{d}{dt}\,|\,\varphi_-\,\rangle = i\big(u\dot{\bar{u}}+v\dot{\bar{v}}\big) = +\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = +\tfrac{1}{2}\,\dot{\omega} \quad ,\end{aligned} \tag{4.209}$$

where $d\omega = \dot{\omega}\,dt$ is the differential solid angle subtended by the path $\hat{\boldsymbol{n}}(t)$. Thus, $\gamma_\pm(\mathcal{C}) = \mp\frac{1}{2}\,\omega_{\mathcal{C}}$ is $\mp$ half the solid angle subtended by the path $\hat{\boldsymbol{n}}_{\mathcal{C}}(t)$ on the Bloch sphere. We may now read off the components $\mathcal{A}^\theta_\pm = 0$ and $\mathcal{A}^\phi_\pm = \mp\frac{1}{2}(1-\cos\theta)$ and invoke Eqn. 4.201 to compute the curvature,

$$\Omega^{\theta\phi}_\pm(\theta,\phi) = \mp\tfrac{1}{2}\sin\theta \quad . \tag{4.210}$$

[30] We must assume that the base space $\mathcal{M}$ is orientable.

[31] Note that we have employed a singular gauge transformation, which is necessary to do the desired unwinding. Also note that the integers q_i should also carry a band index n, which we have suppressed here for notational simplicity.

Note then that the integral of the curvature over the entire sphere is given by

$$\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \; \Omega_\pm^{\theta\phi}(\theta,\phi) = 2\pi\, C_\pm \quad , \tag{4.211}$$

where $C_\pm = \mp 1$ is the Chern number. Equivalently, note that both connections are singular at $\theta = \pi$, where the azimuthal angle is ill-defined. This singularity can be gentled through an appropriate singular gauge transformation $|\,\varphi_\pm\,\rangle = e^{\pm i\phi}\,|\,\widetilde{\varphi}_\pm\,\rangle = e^{\mp i\zeta}\,|\,\widetilde{\varphi}_\pm\,\rangle$, where ζ is defined to be the angle which increases as one winds counterclockwise around the *south pole*, hence $\zeta = -\phi$. This corresponds to $q_\pm = \mp 1$ in our earlier notation, hence again $C_\pm = \mp 1$.

As we saw above, this is a general result: when the base space $\mathcal{M}$ is two dimensional: the integral of the curvature over $\mathcal{M}$ is 2π times an integer. This result calls to memory the famous Gauss-Bonnet theorem (see §4.7 below for more), which says that the integral of the Gaussian curvature K over a two-dimensional manifold $\mathcal{M}$ is

$$\int_{\mathcal{M}} dS\; K = 4\pi(1-g) \quad , \tag{4.212}$$

where g is the *genus* (number of holes) in the manifold $\mathcal{M}$. In the Gauss-Bonnet case, the bundle construction is known as the *tangent bundle* of $\mathcal{M}$, and the corresponding connection is determined by the Riemannian metric one places on $\mathcal{M}$. However, *independent of the metric*, the integral of K is determined solely by the global topology of $\mathcal{M}$, *i.e.* by its genus. Thus, in three-dimensional space, a sphere inherits a metric from its embedding in $\mathbb{R}^3$. If you distort the sphere by denting it, locally its curvature K will change, being the product of the principal radii of curvature at any given point. But the integral of K over the surface will remain fixed at 4π. Just as the genus g of a Riemann surface is unaffected by simple deformations but can change if one does violence to it, such as puncturing and resewing it[32], so is the Chern number invariant under deformations of the underlying Hamiltonian, provided one does not induce a level crossing of the adiabatic eigenstate $|\,\varphi_n\,\rangle$ with a neighboring one. Also, note that if the connection $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ can be defined globally on $\mathcal{M}$, *i.e.* with no singularities, then $C_n = 0$.

4.4.6 Two-band models

For the two band ($S = \frac{1}{2}$) system with Hamiltonian $H = g\mu_{\rm B} B\,\hat{\boldsymbol{n}}(\boldsymbol{\lambda})\cdot\boldsymbol{\sigma}$, one can verify that we may also write the Chern numbers as[33]

$$C_\pm = \pm\frac{1}{4\pi}\int_{\mathcal{M}} d^2\lambda\; \hat{\boldsymbol{n}}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_1}\times\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_2} \quad . \tag{4.213}$$

[32] M. Gilbert's two commandments of topology: (I) Thou shalt not cut. (II) Thou shalt not glue.

[33] The dependence of the magnitude $B = |\boldsymbol{B}|$ on $\boldsymbol{\lambda}$ is irrelevant to the calculation of the Chern numbers. The equivalence $\hat{\boldsymbol{n}} = z^\dagger\boldsymbol{\sigma}z$ for $z = \begin{pmatrix} u \\ v \end{pmatrix}$ is known as the *first Hopf map* from CP^1 to $\mathbb{S}^2$.

In this formulation, the Chern number has the interpretation of a *Pontrjagin number*, which is a topological index classifying real vector bundles (more in §4.7.2 below). Thus, for a tight binding model on a bipartite lattice, the most general Hamiltonian may be written

$$H(\boldsymbol{k}) = d_0(\boldsymbol{k}) + \boldsymbol{d}(\boldsymbol{k}) \cdot \boldsymbol{\sigma} \quad , \tag{4.214}$$

where $\boldsymbol{k}$ is the wavevector and where each $d^\mu(\boldsymbol{k})$ is periodic under translations of $\boldsymbol{k}$ by any reciprocal lattice vector $\boldsymbol{G}$. In this case $\lambda_{1,2} = k_{x,y}$ are the components of $\boldsymbol{k}$, and $\mathcal{M} = \mathbb{T}^2$ is the Brillouin zone torus. Note that the sum of the Chern numbers for each of the $+$ and $-$ bands is zero. As we shall see below with the TKNN problem, for a larger spin generalization, *i.e.* when the magnetic unit cell contains more than two basis elements, the sum $\sum_a C_a$ of the Chern numbers over all bands also vanishes. Consider the two band model with

$$H(\boldsymbol{\theta}) = \begin{pmatrix} m - 2t\cos\theta_1 - 2t\cos\theta_2 & \Delta\,(\sin\theta_1 - i\sin\theta_2) \\ \Delta\,(\sin\theta_1 + i\sin\theta_2) & -m + 2t\cos\theta_1 + 2t\cos\theta_2 \end{pmatrix} \quad . \tag{4.215}$$

As before, $\theta_\mu = \boldsymbol{k}\cdot\boldsymbol{a}_\mu$. Note $H(\boldsymbol{\theta}) = \boldsymbol{d}(\boldsymbol{\theta})\cdot\boldsymbol{\sigma}$ with

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{\theta}) &= \big(\Delta\sin\theta_1\,,\,\Delta\sin\theta_2\,,\,m - 2t\cos\theta_1 - 2t\cos\theta_2\big) \\ &\equiv |\boldsymbol{d}|\,\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big) \quad . \end{aligned} \tag{4.216}$$

Note the adiabatic parameters here are θ_1 and θ_2, upon which ϑ and χ are parametrically dependent. Does $\boldsymbol{d}(\boldsymbol{\theta})$ wind around the Brillouin zone torus, yielding a nonzero Chern number?

First, you might be wondering, where does this model come from? Actually, it is the Hamiltonian for a $p_x + ip_y$ superconductor, but we can back out of $H(\boldsymbol{\theta})$ a square lattice insulator model involving two orbitals a and b which live on top of each other at each site, and are not spatially separated[34]. The parameter m reflects the difference in the local energies of the two orbitals. The nearest neighbor hopping integrals between like orbitals are $t_{aa} = t$ and $t_{bb} = -t$, but $t_{ab}(\pm\boldsymbol{a}_1) = \pm\frac{i}{2}\,\Delta$ and $t_{ab}(\pm\boldsymbol{a}_2) = \pm\frac{1}{2}\Delta$, with $t_{ba}(-\boldsymbol{a}_{1,2}) = t^*_{ab}(+\boldsymbol{a}_{1,2})$ due to hermiticity.

The energy eigenvalues are

$$E_\pm(\boldsymbol{\theta}) = \pm\sqrt{\Delta^2\sin^2\theta_1 + \Delta^2\sin^2\theta_2 + (m - 2t\cos\theta_1 - 2t\cos\theta_2)^2} \quad . \tag{4.217}$$

The Wigner-von Neumann theorem says that degeneracy for complex Hamiltonians like ours has codimension three, meaning one must fine tune three parameters in order to get a degeneracy. The reason is that for $H = \boldsymbol{d}\cdot\sigma$ describing two nearby levels, the gap is $2|\boldsymbol{d}|$, thus in order for the gap to vanish we must require three conditions: $d_x = d_y = d_z = 0$. For the real case where $d_y = 0$ is fixed, we only require two conditions, *i.e.* $d_x = d_z = 0$. For our model, the gap collapse requires

$$\begin{aligned} \Delta\,\sin\theta_1 &= 0 \\ \Delta\,\sin\theta_2 &= 0 \\ m - 2t\cos\theta_1 - 2t\cos\theta_2 &= 0 \quad . \end{aligned} \tag{4.218}$$

[34] In this model they are both s-orbitals, which is unphysical.

Thus, degeneracies occur at $(\theta_1, \theta_2) = (0, 0)$ when $m = 4t$, at (π, π) when $m = -4t$, and at $(0, \pi)$ and $(\pi, 0)$ when $m = 0$. It is clear that for $|m| > 4t$ both Chern numbers must be zero. This is because for $m > 4t$ we have $d_z(\theta_1, \theta_2) > 0$ for all values of the Bloch phases, while for $m < -4t$ we have $d_z(\theta_1, \theta_2) < 0$. Thus in neither case can the $\boldsymbol{d}$ vector wind around the Bloch sphere, and the Pontrjagin/Chern indices accordingly vanish for both bands.

Now consider the case $m \in [0, 4t]$. Recall that the eigenfunctions are given by

$$|\,\varphi_+\,\rangle = \begin{pmatrix} \cos(\frac{1}{2}\vartheta) \\ \sin(\frac{1}{2}\vartheta)\, e^{i\chi} \end{pmatrix} \quad , \quad |\,\varphi_-\,\rangle = \begin{pmatrix} -\sin(\frac{1}{2}\vartheta)\, e^{-i\chi} \\ \cos(\frac{1}{2}\vartheta) \end{pmatrix} \quad , \tag{4.219}$$

with eigenvalues $\pm|\boldsymbol{d}|$. The singularity in both $|\,\varphi_\pm(\theta_1, \theta_2)\,\rangle$ occurs at $\vartheta = \pi$. Recall that $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, which entails $\boldsymbol{d} = (0\,,\,0\,,\,-|\boldsymbol{d}|)$, *i.e.* $d_x = d_y = 0$ and $d_z < 0$. This only occurs for $(\theta_1, \theta_2) = (0, 0)$. All we need to do to compute the Chern numbers is to identify the singularity in $\zeta(\theta_1, \theta_2)$ about this point, *i.e.* does $\zeta = -\chi$ wind clockwise or counterclockwise, in which case $C_+ = -1$ or $C_+ = +1$, respectively. Treating $\theta_{1,2}$ as very small, one easily obtains $\zeta = -\tan^{-1}(\theta_2/\theta_1)$, which is to say clockwise winding, hence $C_\pm = \mp 1$. *Exercise* : Find $C_\pm$ for $m \in [-4t, 0]$.

Haldane honeycomb model

In §4.3.7, we met Haldane's famous honeycomb lattice model, $H(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) + \boldsymbol{d}(\boldsymbol{\theta}) \cdot \boldsymbol{\sigma}$, with

$$\begin{aligned} d_0(\boldsymbol{\theta}) &= -2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1 + \theta_2)\big]\cos\phi \\ d_x(\boldsymbol{\theta}) &= -t_1\big(1 + \cos\theta_1 + \cos\theta_2\big) \\ d_y(\boldsymbol{\theta}) &= t_1\big(\sin\theta_1 - \sin\theta_2\big) \\ d_z(\boldsymbol{\theta}) &= m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1 + \theta_2)\big]\sin\phi \quad . \end{aligned} \tag{4.220}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) \pm |\boldsymbol{d}(\boldsymbol{\theta})|$. Now is quite easy to demonstrate that $\big|\sin\theta_1 + \sin\theta_2 - \sin(\theta_1 + \theta_2)\big| \leqslant \frac{3}{2}\sqrt{3}$, and therefore that the $\boldsymbol{d}(\boldsymbol{\theta})$ cannot wind if $|m| > 3\sqrt{3}\,t_2\,|\sin\phi|$ and $C_\pm = 0$. As above, we set $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, and the singularity in both wavefunctions occurs at $\vartheta = \pi$, which requires $d_x(\boldsymbol{\theta}) = d_y(\boldsymbol{\theta}) = 0$ and $d_z(\boldsymbol{\theta}) < 0$. This in turn requires $\theta_1 = \theta_2 = \frac{2}{3}\pi s$ where $s = \mathsf{sgn}(\sin\phi)$. We now write $\theta_j = \frac{2}{3}\,\pi s + \delta_j$ and find

$$\tan\chi = \frac{d_y(\boldsymbol{\theta})}{d_x(\boldsymbol{\theta})} = \left(\frac{\delta_2 - \delta_1}{\delta_1 + \delta_2}\right)\mathsf{sgn}(\sin\phi) = s\,\tan(\alpha - \tfrac{\pi}{4}) \quad , \tag{4.221}$$

where $\boldsymbol{\delta} \equiv |\boldsymbol{\delta}|\,(\cos\alpha, \sin\alpha)$. Thus, $\zeta = -\chi$ winds in the same sense as α if $s < 0$ and in the opposite sense if $s > 0$. Thus we conclude $C_\pm = \mp\mathsf{sgn}(\sin\phi)$. The topological phase diagram for the Haldane honeycomb lattice model is shown in Fig. 4.20. The phase space is a cylinder in the dimensionless parameters $\phi \in [-\pi, \pi]$ and $m/t_2 \in \mathbb{R}$. Regions are labeled by the Chern numbers $C_\pm$ of the two energy bands.

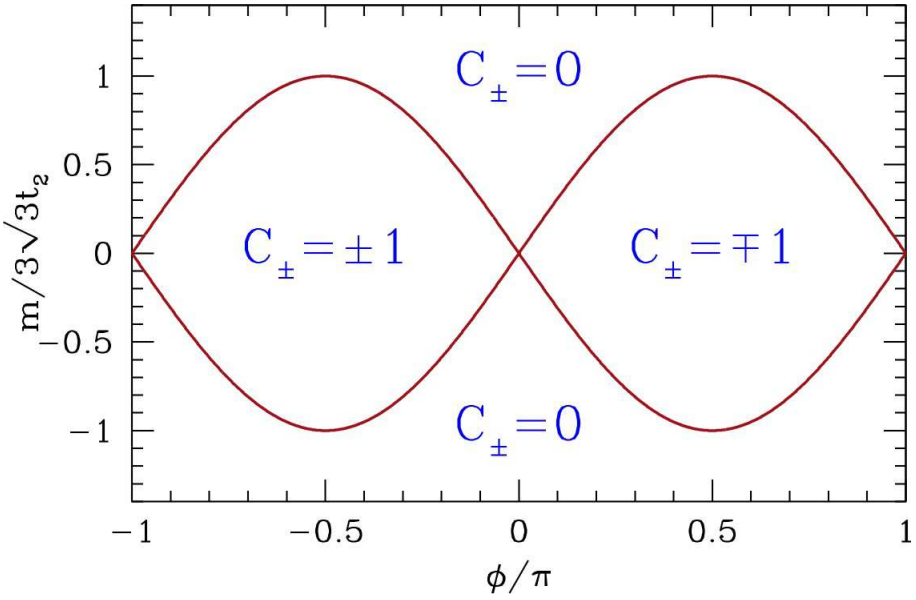


Figure 4.20: Topological phase diagram for the Haldane honeycomb lattice model, in which regions in the $(\sin\phi, m/t_2)$ cylinder are labeled by the Chern numbers $C_\pm$ of the energy bands.

Note on broken symmetries

We saw above how the band structure of spinless π-orbitals on graphene results in two Dirac points at wavevectors $\boldsymbol{K}$ and $\boldsymbol{K}' = -\boldsymbol{K}$. If we adopt a pseudospin convention for the valleys, with Pauli matrices $\boldsymbol{\tau}$, and where $\tau^z = \pm 1$ corresponds to the $\pm\boldsymbol{K}$ valley, it turns out that we may in one stroke write the long wavelength graphene Hamiltonian as

$$H_0 = \tfrac{\sqrt{3}}{2}\, ta\big(q_x\, \sigma^x\, \tau^z + q_y\, \sigma^y\big) \quad , \tag{4.222}$$

where σ^y again operates in the A/B sublattice space. This Hamiltonian is symmetric under the operations of parity ($\mathcal{P}$) and time-reversal ($\mathcal{T}$). Under $\mathcal{P}$, we switch valleys, switch sublattices, and send $q_x \to -q_x$. Under $\mathcal{T}$, we switch valleys and send $\boldsymbol{q} \to -\boldsymbol{q}$. It is also important to remember that $\mathcal{T}$ is antiunitary, and includes the complex conjugation operator $\check{K}$. The matrix parts of these operators, *i.e.* other than their actions on the components of $\boldsymbol{q}$, are given by

$$\mathcal{P} = \sigma^y\, \tau^z \qquad , \qquad \mathcal{T} = i\tau^y \check{K} \quad . \tag{4.223}$$

Note that $\mathcal{T}^2 = -1$ and $\mathcal{T}^{-1} = -\mathcal{T} = \check{K}\tau^y\,(-i)$. Of course $\mathcal{P}^2 = 1$ and thus $\mathcal{P}^{-1} = \mathcal{P}$. One can now check explicitly that $\mathcal{P}H_0\mathcal{P}^{-1} = \mathcal{T}H_0\mathcal{T}^{-1} = H_0$.

There are three ways to introduce a gap into the model, *i.e.* to gap out the dispersion at the $\boldsymbol{K}$ and $\boldsymbol{K}'$ points at the two inequivalent Brillouin zone corners:

1. The first way is by introducing a Semenoff mass term, which is of the form $V_{\rm S} = \Delta_{\rm S}\,\sigma^z$. This turns graphene into boron nitride (BN), distinguishing the local π-orbital energies

on the B and N sites. One can check that

$$\mathcal{P}\,\sigma^z\,\mathcal{P}^{-1} = -\sigma^z \qquad , \qquad \mathcal{T}\,\sigma^z\,\mathcal{T}^{-1} = +\sigma^z \quad , \tag{4.224}$$

and therefore the Semenoff mass breaks parity and preserves time-reversal.

2. The second way comes from the Haldane honeycomb lattice model at $m = 0$, where $V_{\rm H} = \Delta_{\rm H}\,\sigma^z\tau^z$.

$$\mathcal{P}\,\sigma^z\tau^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z \qquad , \qquad \mathcal{T}\,\sigma^z\tau^z\,\mathcal{T}^{-1} = -\sigma^z\tau^z \quad . \tag{4.225}$$

This term, the Haldane mass, preserves parity but breaks time-reversal. It leads to a topological band structure in which the bands are classified by nonzero Chern numbers.

3. The third way involves introducing the physical electron spin, and arises from spin-orbit effects. It essentially is described by two copies of the Haldane model, in which the up and down spin electrons couple oppositely to magnetic flux. This was first discussed by Kane and Mele[35], and is described by the perturbation $V_{\rm KM} = \Delta_{\rm KM}\,\sigma^z\tau^z s^z$. where $\boldsymbol{s}$ is the electron spin operator . The Kane-Mele mass term preserves $\mathcal{P}$ and $\mathcal{T}$ symmetries:

$$\mathcal{P}\,\sigma^z\tau^z s^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z s^z \qquad , \qquad \mathcal{T}\,\sigma^z\tau^z s^z\,\mathcal{T}^{-1} = +\sigma^z\tau^z s^z \quad . \tag{4.226}$$

Therefore, following the tried and true rule in physics that "everything which is not forbidden is compulsory", there *must* be a KM mass term in real graphene. The catch is that it is extremely small because graphene is a low-Z atom, and first principles calculations[36] conclude that the spin-orbit gap is on the order of $10\,$mK – too small to be observed due to finite temperature and disorder effects. However, there are many materials (Bi bilayers, HgTe/CdTe heterostructures, various three-dimensional materials such as α-Sn, $\rm Bi_xSb_{1-x}$ and others) where the effect is predicted to be sizable and where it is indeed observed. This is the essence of topological insulator behavior.

4.4.7 Quantum Metric and Geometric Curvature Tensors

Wilczek and Zee[37] considered adiabatic evolution of an N-fold degenerate block of quantum states with a vector of parameters $\boldsymbol{\lambda}$, spanned by basis states $|\,\varphi_j(\boldsymbol{\lambda})\,\rangle$ with $j \in \{1, \ldots, N\}$. With

$$|\,\Psi(\boldsymbol{\lambda})\,\rangle = \sum_{j=1}^{N} C_j(\boldsymbol{\lambda})\,|\,\varphi_j(\boldsymbol{\lambda})\,\rangle \quad , \tag{4.227}$$

[35] C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
[36] See Y. Yao *et al.*, *Phys. Rev. B* **75**, 041401(R) (2007).
[37] See *Phys. Rev. Lett.* **52**, 2111 (1984)

the distance ds between $|\,\Psi(\boldsymbol{\lambda})\,\rangle$ and $|\,\Psi(\boldsymbol{\lambda}+d\boldsymbol{\lambda})\,\rangle$ is defined by

$$ds^2 \equiv \langle\, d\Psi(\boldsymbol{\lambda}) \,|\, d\Psi(\boldsymbol{\lambda})\,\rangle = \mathcal{Q}_{\mu\nu}(\boldsymbol{\lambda})\, d\lambda^\mu\, d\lambda^\nu \quad , \tag{4.228}$$

where

$$\mathcal{Q}_{\mu\nu}(\boldsymbol{\lambda}) = \langle\, \partial_\mu \Psi(\boldsymbol{\lambda}) \,|\, \partial_\nu \Psi(\boldsymbol{\lambda})\,\rangle \quad . \tag{4.229}$$

The question now arises as to how to take the derivative of $|\,\Psi(\boldsymbol{\lambda})\,\rangle$ with respect to λ^μ. The reason this is unclear is that at each value of $\boldsymbol{\lambda}$ we can rotate the basis by an arbitrary $\boldsymbol{\lambda}$-dependent $\mathrm{U}(N)$ transformation. The solution is to define the changes of the wavefunction via *parallel transport*. The ordinary derivative of $|\,\Psi(\boldsymbol{\lambda})\,\rangle$ is given by

$$\partial_\mu\,|\,\Psi(\boldsymbol{\lambda})\,\rangle = \sum_j \Big(\partial_\mu C_j(\boldsymbol{\lambda})\,|\,\varphi_j(\boldsymbol{\lambda})\,\rangle + C_j(\boldsymbol{\lambda})\,|\,\partial_\mu\varphi_j(\boldsymbol{\lambda})\,\rangle\Big) \tag{4.230}$$

The *covariant* derivative is defined by projecting the ordinary derivative onto the space of adiabatic wavefunctions at each value of $\boldsymbol{\lambda}$, *viz.*

$$\begin{aligned} |\,D_\mu\Psi(\boldsymbol{\lambda})\,\rangle &\equiv \sum_n \Big(\partial_\mu C_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle + C_j(\boldsymbol{\lambda})\,\mathrm{P}(\boldsymbol{\lambda})\,|\,\partial_\mu\varphi_j(\boldsymbol{\lambda})\,\rangle\Big) \\ &= \sum_{j,n}\Big(\delta_{jn}\,\partial_\mu C_n(\boldsymbol{\lambda}) - i\,\mathcal{A}^\mu_{jn}(\boldsymbol{\lambda})\,C_n(\boldsymbol{\lambda})\Big)|\,\varphi_j(\boldsymbol{\lambda})\,\rangle \quad , \end{aligned} \tag{4.231}$$

where

$$\mathrm{P}(\boldsymbol{\lambda}) = \sum_n |\,\varphi_n(\boldsymbol{\lambda})\,\rangle\,\langle\,\varphi_n(\boldsymbol{\lambda})\,| \quad , \tag{4.232}$$

is the projector onto the degenerate subspace and $\mathcal{A}^\mu_{jn}(\boldsymbol{\lambda}) = i\,\langle\,\varphi_j(\boldsymbol{\lambda})\,|\,\partial_\mu\varphi_n(\boldsymbol{\lambda})\,\rangle$ is the Berry connection. The condition for parallel transport is $D_\mu\,|\,\Psi(\boldsymbol{\lambda})\,\rangle = 0$, which entails

$$\partial_\mu C_j(\boldsymbol{\lambda}) = i\,\mathcal{A}^\mu_{jn}(\boldsymbol{\lambda})\,C_n(\boldsymbol{\lambda}) \qquad \text{(parallel transport)} \quad . \tag{4.233}$$

Note that normalization is preserved, *i.e.* $\partial_\mu \sum_j \big|C_j(\boldsymbol{\lambda})\big|^2 = 0$, which follows from the condition $(\mathcal{A}^\mu_{jn})^* = \mathcal{A}^\mu_{nj}$, *i.e.* for each μ, $\mathcal{A}_\mu$ is an $N\times N$ Hermitian matrix in orbital space. We then have

$$|\,\partial_\mu\Psi(\boldsymbol{\lambda})\,\rangle = |\,D_\mu\Psi(\boldsymbol{\lambda})\,\rangle + \big[1-\mathrm{P}(\boldsymbol{\lambda})\big]\,\sum_n C_n(\boldsymbol{\lambda})\,|\,\partial_\mu\varphi_n(\boldsymbol{\lambda})\,\rangle \quad . \tag{4.234}$$

Thus, under the conditions of parallel transport, we have

$$|\,\partial_\mu\Psi(\boldsymbol{\lambda})\,\rangle = \big[1-\mathrm{P}(\boldsymbol{\lambda})\big]\,\sum_n C_n(\boldsymbol{\lambda})\,|\,\partial_\mu\varphi_n(\boldsymbol{\lambda})\,\rangle \qquad \text{(parallel transport)} \quad . \tag{4.235}$$

We may now write

$$\begin{aligned} ds^2 &= C^*_m\,Q^{mn}_{\mu\nu}\,C_n\,d\lambda^\mu\,d\lambda^\nu \\ Q^{mn}_{\mu\nu} &= \langle\,\partial_\mu\varphi_m\,|\,\big(1-\mathrm{P}\big)\,|\,\partial_\nu\varphi_n\,\rangle \quad . \end{aligned} \tag{4.236}$$

For each $(\mu\nu)$, we separate $Q^{mn}_{\mu\nu} \equiv g^{mn}_{\mu\nu} - \frac{i}{2} F^{mn}_{\mu\nu}$ into hermitian and anti-hermitian components. Thus,

$$\begin{aligned} Q^{mn}_{\mu\nu} &= \langle\, \partial_\mu \varphi_m \,|\, \partial_\nu \varphi_n \,\rangle - \mathcal{A}^\mu_{mj}\, \mathcal{A}^\nu_{jn} \\ g^{mn}_{\mu\nu} &= \tfrac{1}{2} \langle\, \partial_\mu \varphi_m \,|\, \partial_\nu \varphi_n \,\rangle + \tfrac{1}{2} \langle\, \partial_\mu \varphi_n \,|\, \partial_\nu \varphi_m \,\rangle - \tfrac{1}{2}\, \mathcal{A}^\mu_{mj}\, \mathcal{A}^\nu_{jn} - \tfrac{1}{2}\, \mathcal{A}^\mu_{nj}\, \mathcal{A}^\nu_{jm} \\ F^{mn}_{\mu\nu} &= i \langle\, \partial_\mu \varphi_m \,|\, \partial_\nu \varphi_n \,\rangle - i \langle\, \partial_\mu \varphi_n \,|\, \partial_\nu \varphi_m \,\rangle - i\, \mathcal{A}^\mu_{mj}\, \mathcal{A}^\nu_{jn} + i\, \mathcal{A}^\mu_{nj}\, \mathcal{A}^\nu_{jm} \quad , \end{aligned} \tag{4.237}$$

with both $g_{\mu\nu}$ and $F_{\mu\nu}$ hermitian in orbital space.

4.4.8 The TKNN formula

Recall the Hamiltonian of Eqn. 4.120 for the isotropic square lattice Hofstadter model with flux $\phi = 2\pi p/q$ per unit cell. A more general version, incorporating anisotropy which breaks 90° rotational symmetry, is given by[38]

$$H(\theta_1, \theta_2) = - \begin{pmatrix} 2t_2 \cos\theta_2 & t_1 & 0 & \cdots & 0 & t_1\, e^{-i\theta_1} \\ t_1 & 2t_2 \cos\left(\theta_2 + \frac{2\pi p}{q}\right) & t_1 & & & 0 \\ 0 & t_1 & 2t_2 \cos\left(\theta_2 + \frac{4\pi p}{q}\right) & t_1 & & \vdots \\ \vdots & 0 & t_1 & \ddots & & \vdots \\ & & & & & \\ 0 & & & & & t_1 \\ t_1\, e^{i\theta_1} & 0 & & \cdots & t_1 & 2t_2 \cos\left(\theta_2 + \frac{2\pi(q-1)p}{q}\right) \end{pmatrix} . \tag{4.238}$$

This is a $q \times q$ matrix, and the q eigenvectors $|\, \varphi_n(\boldsymbol{\theta}) \,\rangle$ are labeled by a band index $n \in \{1, \ldots, q\}$, with component amplitudes $\varphi_{a,n}(\boldsymbol{\theta})$ satisfying

$$H_{aa'}(\boldsymbol{\theta})\, \varphi_{a',n}(\boldsymbol{\theta}) = E_n(\boldsymbol{\theta})\, \varphi_{a,n}(\boldsymbol{\theta}) \quad . \tag{4.239}$$

From Wigner-von Neumann, we expect generically that neighboring bands will not cross as a function of the two parameters (θ_1, θ_2), because degeneracy has codimension three. Thus, associated with each band n is a Chern number C_n. By color coding each spectral gap according to the Chern number of all bands below it, J. Avron produced a beautiful and illustrative image of Hofstadter's butterfly, shown in Fig. 4.21 for the isotropic square lattice and in Fig. 4.22 for the isotropic honeycomb lattice.

It turns out that the Chern number is not just an abstract topological index. It is in fact the dimensionless Hall conductivity σ_{xy} itself, provided the Fermi level lies in a gap between magnetic subbands. This was first discovered by Thouless, Kohmoto, Nightingale, and den Nijs,

[38] We drop the hat on $\hat{H}(\boldsymbol{\theta})$ but fondly recall that $\hat{H}_{aa'}(\boldsymbol{\theta})$ is the lattice Fourier transform of $H_{aa'}(\boldsymbol{R} - \boldsymbol{R}')$.

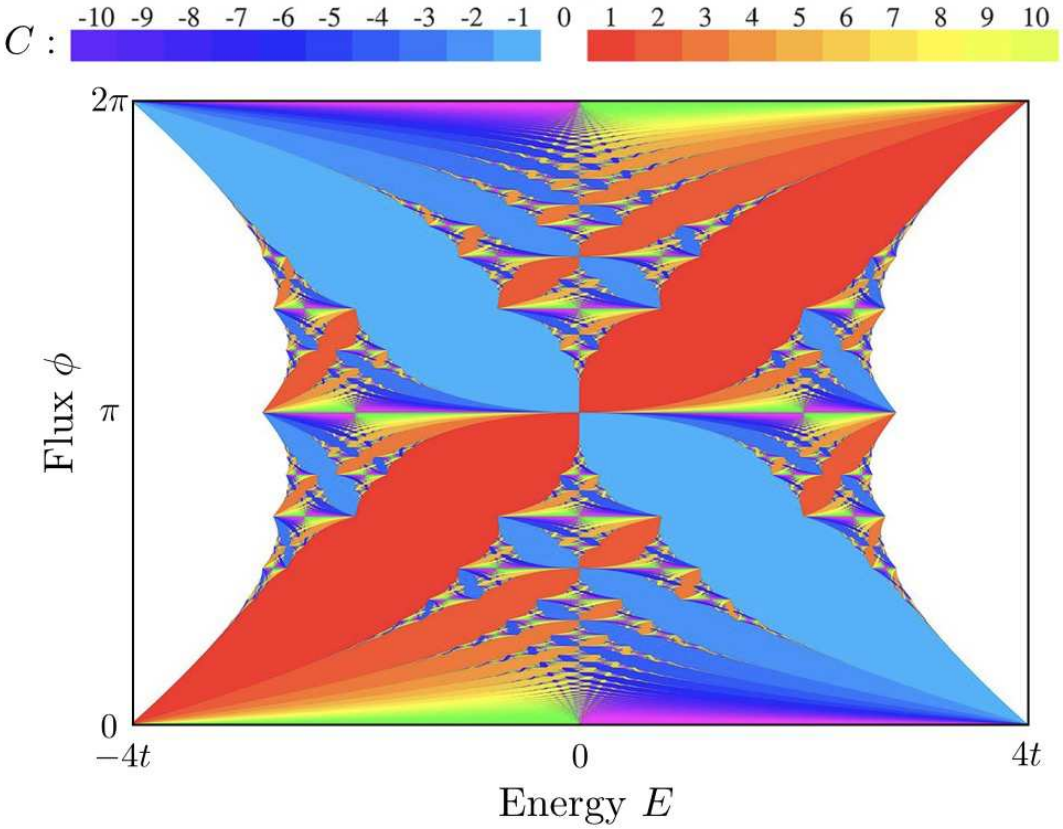


Figure 4.21: Avron's depiction of the Hofstadter butterfly for the isotropic square lattice system. The band gap regions are color coded by Chern number, C, which is the sum of the Chern numbers of all bands below a given gap. White regions correspond to $C = 0$. See J. E. Avron, *Colored Hofstadter butterflies*, in *Multiscale Methods in Quantum Mechanics*, P. Blanchard and G. Dell'Antonio, eds. (Birkhäuser, 2004).

in a seminal paper known by its authors' initials, TKNN[39]. In fact, we've developed the theory here in reverse chronological order. First came TKNN, who found that the contribution $\sigma_{xy}^{(n)}$ to the total Hall conductivity from a band lying entirely below the Fermi level is given by $\sigma_{xy}^{(n)} = \frac{e^2}{h}\, C_n$, where

$$C_n = \frac{i}{2\pi} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \left(\left\langle \frac{\partial \varphi_n}{\partial \theta_1} \,\middle|\, \frac{\partial \varphi_n}{\partial \theta_2} \right\rangle - \left\langle \frac{\partial \varphi_n}{\partial \theta_2} \,\middle|\, \frac{\partial \varphi_n}{\partial \theta_1} \right\rangle \right) \tag{4.240}$$

is an integral over the Brillouin zone. They proved that this expression is an integer, because

[39] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).

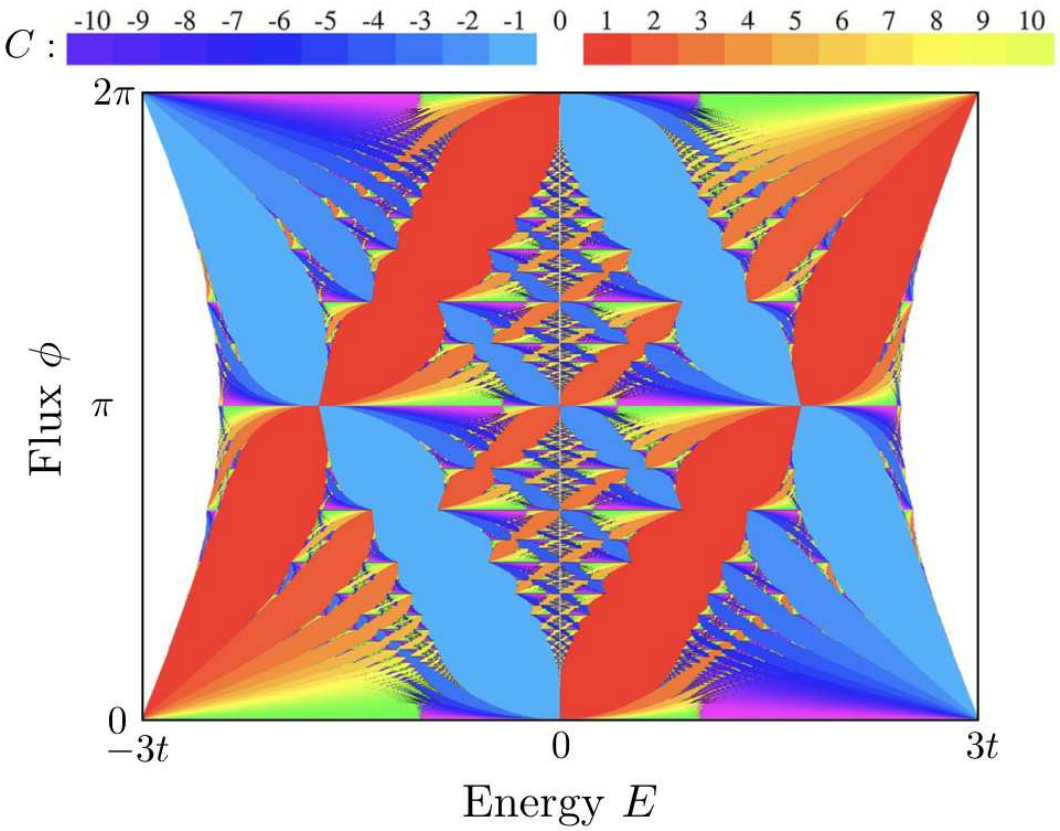


Figure 4.22: Colored Hofstadter butterfly for the honeycomb lattice system, from A. Agazzi, J.-P. Eckman, and G. M. Graf, *J. Stat. Phys.* **156**, 417 (2014).

invoking Stokes' theorem,

$$C_n = \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_2 \left\langle\, \varphi_n \,\middle|\, \frac{\partial \varphi_n}{\partial \theta_2} \,\right\rangle\Big|_{\theta_1=0}^{\theta_1=2\pi} - \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_1 \left\langle\, \varphi_n \,\middle|\, \frac{\partial \varphi_n}{\partial \theta_1} \,\right\rangle\Big|_{\theta_2=0}^{\theta_2=2\pi} \quad . \tag{4.241}$$

But since $H(\theta_1, \theta_2)$ is doubly periodic with period 2π in each of its arguments, we must have

$$\begin{aligned} |\,\varphi_n(\theta_1, 2\pi)\,\rangle &= e^{if_n(\theta_1)}\,|\,\varphi_n(\theta_1, 0)\,\rangle \\ |\,\varphi_n(2\pi, \theta_2)\,\rangle &= e^{ig_n(\theta_2)}\,|\,\varphi_n(0, \theta_2)\,\rangle \quad . \end{aligned} \tag{4.242}$$

Thus, one finds

$$C_n = \frac{1}{2\pi}\Big(f_n(2\pi) - f_n(0) + g_n(0) - g_n(2\pi)\Big) \quad . \tag{4.243}$$

But we also have

$$\begin{aligned} |\,\varphi_n(0,0)\,\rangle &= e^{-if_n(0)}\,|\,\varphi_n(0,2\pi)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,|\,\varphi_n(2\pi,2\pi)\,\rangle \\ &= e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,|\,\varphi_n(2\pi,0)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,e^{ig_n(0)}\,|\,\varphi_n(0,0)\,\rangle \quad , \end{aligned} \tag{4.244}$$

and therefore $\exp(2\pi i\, C_n) = 1$ and $C_n \in \mathbb{Z}$. But just as Berry didn't know he had found a holonomy, TKNN didn't know they had found a Chern number. That mathematical feature was first elucidated by Avron, Seiler, and Simon[40], in a paper which is widely appreciated but which, understandably, is not known by its authors' initials.

To see why Hall conductivity is related to Berry curvature, consider an electric field $\boldsymbol{E} = E_y\,\hat{\boldsymbol{y}}$, and the single electron Hamiltonian $H(E_y) = H(0) - eE_y y$, where $H(0) = \frac{\boldsymbol{\pi}^2}{2m} + V(\boldsymbol{r})$ has eigenstates $|\,\alpha\,\rangle$ and eigenvalues ε_α. First order perturbation theory in the electric field term says

$$|\,\alpha'\,\rangle = |\,\alpha\,\rangle - eE_y \sum_\beta{}' \frac{|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{4.245}$$

where the prime on the sum means the term with $\beta = \alpha$ is excluded. Let's now compute the expectation of the velocity operator v_x in the perturbed state $|\,\alpha'\,\rangle$. We have, to lowest order,

$$\langle\,\alpha'\,|\,v_x\,|\,\alpha'\,\rangle = -eE_y \sum_\beta{}' \frac{\langle\,\alpha\,|\,v_x\,|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle + \langle\,\alpha\,|\,y\,|\,\beta\,\rangle\langle\,\beta\,|\,v_x\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \tag{4.246}$$

We now invoke the Feynman-Hellman theorem, which says

$$\langle\,\alpha\,|\,y\,|\,\beta\,\rangle = \frac{\hbar}{i}\,\frac{\langle\,\alpha\,|\,v_y\,|\,\beta\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{4.247}$$

multiply by the electron charge $-e$, divide by the area of the system Ω, and sum using the Fermi distribution over the levels $|\,\alpha\,\rangle$, to obtain the current density j_x :

$$j_x = E_y \cdot \frac{e^2}{h} \cdot \frac{2\pi i\hbar^2}{\Omega} \sum_\alpha \sum_\beta f_\alpha\,(1 - f_\beta)\,\epsilon_{ij}\,\frac{\langle\,\alpha\,|\,v_i\,|\,\beta\,\rangle\langle\,\beta\,|\,v_j\,|\,\alpha\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} \quad , \tag{4.248}$$

where f_α is the Fermi function at temperature T, chemical potential μ, and energy ε_α. The above expression for $\sigma_{xy} = j_x/E_y$ is known as the *Kubo formula* for the Hall conductivity. At $T = 0$, the Fermi distribution becomes the step function $f_\alpha = \Theta(E_{\mathsf{F}} - E_\alpha)$.

Suppose our system lies on a torus defined by the spatial periods $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$. Define the gauge transformed Hamiltonian

$$\widetilde{H}(\boldsymbol{\theta}) \equiv e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,H\,e^{+i\boldsymbol{q}\cdot\boldsymbol{r}} \quad , \tag{4.249}$$

where

$$\boldsymbol{q} = \theta_2\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_1}{\Omega} - \theta_1\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_2}{\Omega} \quad , \tag{4.250}$$

with $\Omega = \hat{\boldsymbol{z}}\cdot\boldsymbol{L}_1\times\boldsymbol{L}_2 = 2\pi\ell^2 p$ with $p \in \mathbb{Z}$, *i.e.* the total magnetic flux through the system is an integer multiple of the Dirac quantum. Then

$$\frac{\partial\widetilde{H}}{\partial\theta_i} = \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,\hbar\boldsymbol{v}\,e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \equiv \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot\hbar\widetilde{\boldsymbol{v}} \quad , \tag{4.251}$$

[40] J. E. Avron, R. Seiler, and B. Simon, *Phys. Rev. Lett.* **51**, 51 (1983).

because $[H, \boldsymbol{r}] = (\hbar/i)\,\boldsymbol{v}$. Thus, defining $|\,\widetilde{\alpha}\,\rangle \equiv \exp(-i\boldsymbol{q}\cdot\boldsymbol{r})\,|\,\alpha\,\rangle$, and recalling the definition of the wavevector $\boldsymbol{q} = \epsilon_{ab}\,\theta_a\,\boldsymbol{L}_b \times \hat{\boldsymbol{z}}/\Omega$, we find

$$\frac{\partial \widetilde{H}}{\partial \theta_a} = \hbar\,\epsilon_{ab}\,\epsilon_{ij}\,\frac{\widetilde{v}_i\,L_b^j}{\Omega} \quad . \tag{4.252}$$

We then find

$$\sigma_{xy} = \frac{j_x}{E_y} = \sum_{\alpha\ \mathrm{occ}} \sigma_{xy}^{(\alpha)} \quad , \tag{4.253}$$

where the sum is over occupied states below the Fermi level, and where

$$\sigma_{xy}^{(\alpha)} = \frac{e^2}{h}\cdot 2\pi i \sum_\beta{}' \epsilon_{ij}\,\frac{\langle\,\widetilde{\alpha}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_i}\,|\,\widetilde{\beta}\,\rangle\langle\,\widetilde{\beta}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_j}\,|\,\widetilde{\alpha}\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} \quad , \tag{4.254}$$

which is precisely of the form of Eqn. 4.202. Thus, if we now *uniformly average over the boundary phases* θ_1 *and* θ_2, we obtain

$$\big\langle \sigma_{xy}^{(\alpha)} \big\rangle = \frac{e^2}{h}\cdot\frac{i}{2\pi}\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \sum_\beta{}' \epsilon_{ij}\,\frac{\langle\,\widetilde{\alpha}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_i}\,|\,\widetilde{\beta}\,\rangle\langle\,\widetilde{\beta}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_j}\,|\,\widetilde{\alpha}\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} = \frac{e^2}{h}\,C_\alpha \quad , \tag{4.255}$$

i.e. each filled band α contributes $\frac{e^2}{h}\,C_\alpha$ to the total Hall conductivity whenever the Fermi level at $T = 0$ lies in a gap between energy bands. For a crystalline (periodic) system, averaging over $\theta_{1,2}$ is tantamount to integrating over the Brillouin zone.

4.5 Semiclassical Dynamics of Bloch Electrons

Consider a time-evolving quantum mechanical state $|\psi(t)\rangle$. The time dependence of the expectation value $\mathcal{O}(t) = \langle\psi(t)|\mathcal{O}|\psi(t)\rangle$ satisfies

$$\frac{d\mathcal{O}}{dt} = \frac{i}{\hbar}\,\langle\,\psi(t)\,|\,[H,\mathcal{O}]\,|\,\psi(t)\,\rangle \quad . \tag{4.256}$$

Thus for $H = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r})$ we have $\frac{d}{dt}\,\langle\boldsymbol{r}\rangle = \langle\frac{\boldsymbol{p}}{m}\rangle$ and $\frac{d}{dt}\,\langle\boldsymbol{p}\rangle = -\langle\boldsymbol{\nabla}V\rangle$, a result known as Ehrenfest's theorem. There are a couple of problems in applying this to electrons in crystals, though. One is that the momentum $\boldsymbol{p}$ in a Bloch state is defined only modulo $\hbar\boldsymbol{G}$, where $\boldsymbol{G}$ is any reciprocal lattice vector. Another is that the potential $\Delta V(\boldsymbol{r}) = e\boldsymbol{E}\cdot\boldsymbol{r}$ breaks the lattice periodicity present in the crystal.

4.5.1 Adiabatic evolution

Here we assume $d = 3$ dimensions. Recall $\boldsymbol{E} = -\boldsymbol{\nabla}\phi - c^{-1}\partial_t \boldsymbol{A}$, so rather that taking $\boldsymbol{A} = 0$ and $\phi = -\boldsymbol{E}\cdot\boldsymbol{r}$ in the case of a uniform electric field, we can instead take $\boldsymbol{A}(t) = -c\boldsymbol{E}t$ and $\phi = 0$ and write

$$H(t) = \frac{\big(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}(t)\big)^2}{2m} + V(\boldsymbol{r}) \quad , \tag{4.257}$$

with $\partial_t \boldsymbol{A} = -c\boldsymbol{E}$, and $\boldsymbol{\nabla}\times\boldsymbol{A} = \boldsymbol{B}$ if there is a magnetic field as well. We assume that the electric field is very weak, which means that we can treat the time dependent Hamiltonian $H(t)$ in the adiabatic limit[41].

We begin by reiterating some key results from §4.4.4. Consider a setting in which a Hamiltonian $H(\boldsymbol{\lambda})$ depends on a set of parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_K\}$. The adiabatic eigenstates of $H(\boldsymbol{\lambda})$ are denoted as $|n(\boldsymbol{\lambda})\rangle$, where $H(\boldsymbol{\lambda})\,|n(\boldsymbol{\lambda})\rangle = E_n(\boldsymbol{\lambda})\,|n(\boldsymbol{\lambda})\rangle$. Now suppose that $\lambda(t)$ varies with time. Since the set $\big\{|n(\boldsymbol{\lambda})\rangle\big\}$ is complete, we may expand the wavefunction $|\psi(t)\rangle$ in the adiabatic basis, *viz.*

$$|\,\psi(t)\,\rangle = \sum_n a_n(t)\,|\,n\big(\boldsymbol{\lambda}(t)\big)\,\rangle \quad . \tag{4.258}$$

Now we impose the condition $i\hbar\,\partial_t\,|\psi(t)\rangle = H\big(\boldsymbol{\lambda}(t)\big)|\psi(t)\rangle$. We first define the phases $\phi_n(t)$ and $\gamma_n(t)$, where

$$\phi_n(t) = -\frac{1}{\hbar}\int^t dt'\; E_n\big(\boldsymbol{\lambda}(t')\big) \tag{4.259}$$

and where $\gamma_n(t)$ satisfies

$$\frac{d\gamma_n}{dt} = i\,\langle\,n\big(\boldsymbol{\lambda}(t)\big)\,|\,\frac{d}{dt}\,|\,n\big(\boldsymbol{\lambda}(t)\big)\,\rangle \quad . \tag{4.260}$$

Then, writing $a_n(t) \equiv e^{i\phi_n(t)}\,e^{i\gamma_n(t)}\,\alpha_n(t)$, we find

$$\frac{d\alpha_n}{dt} = -{\sum_l}'\,e^{i(\gamma_l-\gamma_n)}\,e^{i(\phi_l-\phi_n)}\,\alpha_l \quad , \tag{4.261}$$

where the prime on the sum indicates that the term $l = n$ is to be excluded. Now consider initial conditions where $a_l(0) = \delta_{ln}$. Since the evolution is adiabatic, the phases $\phi_l(t)$ are the fastest evolving quantities, with $\partial_t\,\phi_l = -E_l/\hbar = \mathcal{O}(1)$, as opposed to γ_l and α_l, which vary on the slow time scale associated with the evolution of $\boldsymbol{\lambda}(t)$. This allows us to approximately integrate the above equations to obtain

$$\alpha_n(t) \approx \alpha_n(0) = 1 \qquad , \qquad \alpha_l(t) \approx -i\hbar\,\frac{\langle\,l\,|\,\partial_t\,|\,n\,\rangle}{E_l - E_n} \quad . \tag{4.262}$$

Thus,

$$|\,\psi(t)\,\rangle \approx e^{i\phi_n(t)}\,e^{i\gamma_n(t)}\left\{|\,n(t)\,\rangle - i\hbar\,{\sum_l}'\,\frac{|\,l(t)\,\rangle\langle\,l(t)\,|\,d_t\,|\,n(t)\,\rangle}{E_l(t) - E_n(t)}\right\} \quad , \tag{4.263}$$

[41] Technically, we should require there be a finite energy gap in order to justify adiabatic evolution.

where each $|\, l(t)\,\rangle = |\, l\big(\boldsymbol{\lambda}(t)\big)\,\rangle$, and where $d_t = d/dt$ is the total time derivative. Note that we can write

$$\frac{d\,|n\rangle}{dt} = \frac{\partial\,|n\rangle}{\partial\lambda_\mu}\cdot\frac{d\lambda_\mu}{dt} \quad , \tag{4.264}$$

with an implied sum on μ, and with $i\,\langle n|\,d_t\,|n\rangle = \mathcal{A}_n^\mu\,\dot\lambda_\mu$, where

$$\mathcal{A}_n^\mu(\boldsymbol{\lambda}) \equiv i\,\Big\langle\, n(\boldsymbol{\lambda})\,\Big|\,\frac{\partial}{\partial\lambda_\mu}\,\Big|\,n(\boldsymbol{\lambda})\,\Big\rangle \tag{4.265}$$

is the *geometric connection* (or *Berry connection*) for the state $|\,n(\boldsymbol{\lambda})\,\rangle$. Note that the Berry connection is gauge-dependent, in that redefining $|\,\widetilde{n}(\boldsymbol{\lambda})\,\rangle \equiv e^{if_n(\boldsymbol{\lambda})}\,|\,n(\boldsymbol{\lambda})\,\rangle$ results in

$$\widetilde{\mathcal{A}}_n^\mu(\boldsymbol{\lambda}) = i\,\Big\langle\,\widetilde{n}(\boldsymbol{\lambda})\,\Big|\,\frac{\partial}{\partial\lambda_\mu}\,\Big|\,\widetilde{n}(\boldsymbol{\lambda})\,\Big\rangle = \mathcal{A}_n^\mu(\boldsymbol{\lambda}) - \frac{\partial f_n}{\partial\lambda_\mu} \quad . \tag{4.266}$$

If we require that the adiabatic wavefunctions be single-valued as a function of $\boldsymbol{\lambda}$, then *the integral of the Berry connection around a closed path is a gauge-invariant quantity*,

$$\gamma_n(\mathcal{C}) \equiv \oint\limits_{\mathcal{C}} d\lambda_\mu\,\mathcal{A}_n^\mu(\boldsymbol{\lambda}) \quad , \tag{4.267}$$

since $f_n(\boldsymbol{\lambda})$ can wind only by $2\pi k$, with $k \in \mathbb{Z}$, around a closed loop $\mathcal{C}$ in parameter space. If $\mathcal{C}$ is contractable to a point, then $k = 0$.

Now consider the cell function $|u_{n\boldsymbol{k}}\rangle$ as a function of the Bloch wavevector $\boldsymbol{k}$ for each band index n. The velocity operator is $\boldsymbol{v}(\boldsymbol{k}) = \hbar^{-1}\,\partial H(\boldsymbol{k})/\partial\boldsymbol{k}$, where $H(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}He^{i\boldsymbol{k}\cdot\boldsymbol{r}}$. We then have[42]

$$\frac{d\boldsymbol{r}}{dt} = \langle\,\psi(t)\,|\,\boldsymbol{v}(\boldsymbol{k})\,|\,\psi(t)\,\rangle = \frac{1}{\hbar}\frac{\partial E_n(\boldsymbol{k})}{\partial\boldsymbol{k}} - \frac{d\boldsymbol{k}}{dt}\times\boldsymbol{\Omega}_n(\boldsymbol{k}) \tag{4.268}$$

where

$$\mathcal{A}_n^\mu(\boldsymbol{k}) = i\,\Big\langle\, u_{n\boldsymbol{k}}\,\Big|\,\frac{\partial}{\partial k^\mu}\,\Big|\,u_{n\boldsymbol{k}}\,\Big\rangle \qquad , \qquad \Omega_n^\mu(\boldsymbol{k}) = \epsilon_{\mu\nu\lambda}\,\frac{\partial\mathcal{A}_n^\lambda(\boldsymbol{k})}{\partial k^\nu} \quad . \tag{4.269}$$

In vector notation, $\boldsymbol{\mathcal{A}}_n(\boldsymbol{k}) = i\,\langle\, u_{n\boldsymbol{k}}\,|\,\boldsymbol{\nabla}_{\boldsymbol{k}}\,|\,u_{n\boldsymbol{k}}\,\rangle$ and $\boldsymbol{\Omega}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}}\times\boldsymbol{\mathcal{A}}_n(\boldsymbol{k})$, where $\boldsymbol{\nabla}_{\boldsymbol{k}} = \frac{\partial}{\partial\boldsymbol{k}}$. Eqn. 4.268 is the first of our semiclassical equations of motion for an electron wavepacket in a crystal. The quantity $\boldsymbol{\Omega}_n(\boldsymbol{k})$, which has dimensions of area, is called the *Berry curvature* of the band $|u_{n\boldsymbol{k}}\rangle$. The second term in Eqn. 4.268 is incorrectly omitted in many standard solid state physics texts! When the orbital moment of the Bloch electrons is included, we must substitute[43]

$$E_n(\boldsymbol{k}) \to E_n(\boldsymbol{k}) - \boldsymbol{M}_n(\boldsymbol{k})\cdot\boldsymbol{B}(\boldsymbol{r},t) \quad , \tag{4.270}$$

[42] See §4.8.

[43] See G. Sundaram and Q. Niu, *Phys. Rev. B* **59**, 14915 (1999); also D. Xiao, M. Chang, and Q. Niu, *Rev. Mod. Phys.* **82**, 1959 (2010).

where

$$M_n^\mu(\boldsymbol{k}) = e\,\epsilon_{\mu\nu\lambda}\,\mathsf{Im}\left\langle \frac{\partial u_{n\boldsymbol{k}}}{\partial k^\nu}\,\middle|\,\Big(E_n(\boldsymbol{k}) - H(\boldsymbol{k})\Big)\,\middle|\,\frac{\partial u_{n\boldsymbol{k}}}{\partial k^\lambda}\right\rangle \quad , \tag{4.271}$$

and $H(\boldsymbol{k}) = \frac{(\boldsymbol{p}+\hbar\boldsymbol{k})^2}{2m} + V(\boldsymbol{r})$ as before.

The second equation of semiclassical motion is for $d\boldsymbol{k}/dt$. This is the familiar equation derived from Newton's second law[44],

$$\frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar}\,\boldsymbol{E} - \frac{e}{\hbar c}\,\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B} - \frac{e}{2\hbar mc}\,\boldsymbol{\nabla}(\boldsymbol{\sigma}\cdot\boldsymbol{B}) \quad , \tag{4.272}$$

where we include the contribution from the Zeeman Hamiltonian $H_{\rm Z} = (e\hbar/2mc)\,\boldsymbol{\sigma}\cdot\boldsymbol{B}$. If we choose $\hat{\boldsymbol{z}}$ as the spin quantization axis, so $H_{\rm Z} = (e\hbar/2mc)\,\sigma B^z$, then we can combine the spin orbit force with that of the electric field and write

$$\frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar}\,\boldsymbol{E}_\sigma - \frac{e}{\hbar c}\,\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B} \quad , \tag{4.273}$$

where

$$\boldsymbol{E}_\sigma = \boldsymbol{E} + \frac{\sigma}{2mc}\boldsymbol{\nabla}B^z \quad . \tag{4.274}$$

4.5.2 Violation of Liouville's theorem and its resolution

Our equations of motion for a wavepacket are thus

$$\begin{aligned} \dot{x}^\alpha + \epsilon_{\alpha\beta\gamma}\,\dot{k}^\beta\,\Omega_n^\gamma &= v_n^\alpha \\ \dot{k}^\alpha + \frac{e}{\hbar c}\,\epsilon_{\alpha\beta\gamma}\,\dot{x}^\beta B^\gamma &= -\frac{e}{\hbar}\,E^\alpha \quad , \end{aligned} \tag{4.275}$$

where $\boldsymbol{v}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}}\,E_n(\boldsymbol{k})/\hbar$. These equations may be recast as

$$\begin{pmatrix} \delta_{\alpha\beta} & \epsilon_{\alpha\beta\gamma}\,\Omega_n^\gamma \\ \frac{e}{\hbar c}\,\epsilon_{\alpha\beta\gamma}\,B^\gamma & \delta_{\alpha\beta} \end{pmatrix} \begin{pmatrix} \dot{x}^\alpha \\ \dot{k}^\alpha \end{pmatrix} = \begin{pmatrix} v_n^\alpha \\ -\frac{e}{\hbar}\,E^\alpha \end{pmatrix} \quad . \tag{4.276}$$

Inverting, we find

$$\begin{aligned} \dot{x}^\alpha &= \big(1 + \tfrac{e}{\hbar c}\boldsymbol{B}\cdot\boldsymbol{\Omega}_n\big)^{-1}\Big\{v_n^\alpha + \tfrac{e}{\hbar c}\,(\boldsymbol{v}_n\cdot\boldsymbol{\Omega}_n)\,B^\alpha + \tfrac{e}{\hbar}\,\epsilon_{\alpha\beta\gamma}\,E^\beta\,\Omega_n^\gamma\Big\} \\ \dot{k}^\alpha &= -\tfrac{e}{\hbar}\,\big(1 + \tfrac{e}{\hbar c}\boldsymbol{B}\cdot\boldsymbol{\Omega}_n\big)^{-1}\Big\{E^\alpha + \tfrac{e}{\hbar c}\,(\boldsymbol{E}\cdot\boldsymbol{B})\,\Omega_n^\alpha + \tfrac{1}{c}\,\epsilon_{\alpha\beta\gamma}\,v_n^\beta\,B^\gamma\Big\} \quad . \end{aligned} \tag{4.277}$$

[44] Some subtleties in the derivation are discussed in A. Manohar, *Phys. Rev. B* **34**, 1287 (1986).

It is straightforward to derive the result

$$\frac{\partial \dot{x}^\alpha}{\partial x^\alpha} + \frac{\partial \dot{k}^\alpha}{\partial k^\alpha} = -\frac{\partial \ln D_n}{\partial x^\alpha}\,\frac{dx^\alpha}{dt} - \frac{\partial \ln D_n}{\partial k^\alpha}\,\frac{dk^\alpha}{dt} - \frac{\partial \ln D_n}{\partial t} = -\frac{d \ln D_n}{dt} \quad , \tag{4.278}$$

where

$$D_n(\boldsymbol{r}, \boldsymbol{k}, t) = 1 + \frac{e}{\hbar c}\,\boldsymbol{B}(\boldsymbol{r}, t) \cdot \boldsymbol{\Omega}_n(\boldsymbol{k}) \tag{4.279}$$

is dimensionless. As discussed by Xiao, Shi, and Niu[45], this implies a violation of Liouville's theorem, as phase space volumes will then expand according to

$$\frac{d \ln \Delta V}{dt} = \boldsymbol{\nabla}_{\boldsymbol{r}} \cdot \dot{\boldsymbol{r}} + \boldsymbol{\nabla}_{\boldsymbol{k}} \cdot \dot{\boldsymbol{k}} = -d \ln D_n(\boldsymbol{r}, \boldsymbol{k}, t)/dt \quad , \tag{4.280}$$

where $\Delta V = \Delta \boldsymbol{r}\, \Delta \boldsymbol{k}$ is a phase space volume element. Thus, $\Delta V(t) = \Delta V(0)/D_n(\boldsymbol{r}, \boldsymbol{k}, t)$, and this inconvenience can be eliminated by redefining the phase space metric as

$$d\mu = \frac{d^3r\, d^3k}{(2\pi)^3} \quad \longrightarrow \quad d\tilde{\mu} \equiv D_n(\boldsymbol{r}, \boldsymbol{k}, t)\,\frac{d^3r\, d^3k}{(2\pi)^3} \quad . \tag{4.281}$$

This means that the expectation of any local observable $\mathcal{O}$ is given by

$$\begin{aligned} \langle \mathcal{O} \rangle(\boldsymbol{r}', t) &= \sum_n \int\limits_{\hat{\Omega}} d\tilde{\mu}\, f_n(\boldsymbol{r}, \boldsymbol{k}, t)\, \langle\, u_{n\boldsymbol{k}} \,|\, \mathcal{O} \,|\, u_{n\boldsymbol{k}} \,\rangle\, \delta(\boldsymbol{r} - \boldsymbol{r}') \\ &= \sum_n \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, D_n(\boldsymbol{r},' \boldsymbol{k}, t)\, f_n(\boldsymbol{r}', \boldsymbol{k}, t)\, \langle\, u_{n\boldsymbol{k}} \,|\, \mathcal{O} \,|\, u_{n\boldsymbol{k}} \,\rangle \quad , \end{aligned} \tag{4.282}$$

where $f_n(\boldsymbol{r}, \boldsymbol{k}, t)$ is the mean occupation number of the state $|\, n\boldsymbol{k} \,\rangle$ in the region spatially centered at $\boldsymbol{r}$, and where we have absorbed the spin polarization label σ into the band label n.[46] In equilibrium, $f_n(\boldsymbol{k})$ is the Fermi function $f\big(E_n(\boldsymbol{k}) - \mu\big)$. The electrical current density carried by a given band n is then

$$\begin{aligned} j_n^\alpha(\boldsymbol{r}, t) &= \int\limits_{\hat{\Omega}} \frac{d^dk}{(2\pi)^d}\, D_n(\boldsymbol{r}, \boldsymbol{k}, t)\, f_n(\boldsymbol{k})\, (-e\dot{x}) \\ &= -e \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \left\{ v_n^\alpha + \frac{e}{\hbar c}\,(\boldsymbol{v}_n \cdot \boldsymbol{\Omega}_n)\, B^\alpha + \frac{e}{\hbar}\,\epsilon_{\alpha\beta\gamma}\, E^\beta\, \Omega_n^\gamma \right\} f_n(\boldsymbol{k}) \quad . \end{aligned} \tag{4.283}$$

Note the cancellation of the $D_n(\boldsymbol{k})$ factors in $d\mu$ and $\dot{x}$. Consider the case of a filled band, with $\boldsymbol{B} = 0$. The total current density is then

$$\boldsymbol{j}_n = -\frac{e^2}{\hbar}\,\boldsymbol{E} \times \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \boldsymbol{\Omega}_n(\boldsymbol{k}) \quad . \tag{4.284}$$

Thus, *when the geometric curvature $\boldsymbol{\Omega}_n(\boldsymbol{k})$ is nonzero, a filled band may carry current.*

[45] See D. Xiao, J. Shi, and Q. Niu, *Phys. Rev. Lett.*. **95**, 137204 (2005).

[46] This allows for intrinsic spin structure in the cell wavefunctions, which is the case when spin-orbit terms are present.

4.5.3 Bloch oscillations

Let's consider the simplest context for our semiclassical equations of motion: $d = 1$ dimension, which means $\boldsymbol{B} = 0$. We'll take a nearest neighbor s orbital hopping Hamiltonian, whose sole tight binding band has the dispersion[47] $E(k) = -2\beta\cos(ka)$. The semiclassical equations of motion are

$$\dot{x} = \frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{2\beta a}{\hbar}\,\sin(ka) \qquad , \qquad \dot{k} = -\frac{e}{\hbar}\,E \quad . \tag{4.285}$$

The second of these equations is easily integrated for constant E:

$$k(t) = k(0) - \frac{e}{\hbar}Et \quad , \tag{4.286}$$

in which case

$$\dot{x} = \frac{2\beta a}{\hbar}\,\sin\Big(k(0)\,a - \frac{eaEt}{\hbar}\Big) \qquad \Rightarrow \qquad x(t) = x(0) + \frac{2\beta}{eE}\left[\cos\Big(k(0)\,a - \frac{eaEt}{\hbar}\Big) - \cos\big(k(0)\,a\big)\right] \quad . \tag{4.287}$$

Note that $x(t)$ oscillates in time! This is quite unlike the free electron case, where we have $m\ddot{x} = -eE$, yielding ballistic motion $x(t) = x(0) + \dot{x}(0)\,t - \frac{eE}{2m}\,t^2$, *i.e.* uniform acceleration $(-eE/m)$. This remarkable behavior is called a *Bloch oscillation*.

The period of the Bloch oscillations is $\tau_{\text{B}} = \hbar/eaE$. Let's estimate τ_{B}, taking $E = 1\,\text{V/cm}$ and $a = 3\,\text{Å}$. We find $\tau_{\text{B}} = 1.4\times10^{-7}\,\text{s}$, which is much larger than typical scattering times due to phonons or lattice impurities. For example, the thermal de Broglie lifetime is $\hbar/k_{\text{B}}T = 2.5\times10^{-14}\,\text{s}$ at $T = 300\,\text{K}$. Thus, the wavepacket never makes it across the Brillouin zone - not even close. In the next chapter, we will see how to model charge transport in metals.

4.6 *Ab initio* Calculations of Electronic Structure

4.6.1 Orthogonalized plane waves

The plane wave expansion of Bloch states in Eqn. 4.10,

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})\,e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}} \tag{4.288}$$

is formally correct, but in practice difficult to implement. The main reason is that one must keep a large number of coefficients $C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})$ in order to get satisfactory results, because the interesting valence or conduction band Bloch functions must be orthogonal to the core Bloch states derived from the atomic 1s, 2s, *etc.* levels. If the core electrons are localized within a

[47] We write the hopping integral as β so as to avoid any confusion with the time variable, t.

volume $v_{\rm c}$, Heisenberg tells us that the spread in wavevector needed to describe such states is given by $\Delta k_x\,\Delta k_y\,\Delta k_z \gtrsim v_{\rm c}^{-1}$. In $d = 3$, the number of plane waves we need to describe a Bloch state of crystal momentum $\hbar\boldsymbol{k}$ is then

$$N_{\rm pw} \approx \frac{\Delta k_x\,\Delta k_y\,\Delta k_z}{\hat{v}_0} = \frac{1}{8\pi^3}\cdot\frac{v_0}{v_{\rm c}} \quad , \tag{4.289}$$

where $\hat{v}_0 = \mathsf{vol}(\hat{\Omega})$ is the volume of the first Brillouin zone (with dimensions $[\hat{v}_0] = L^{-d}$). If the core volume $v_{\rm c}$ is much smaller than the Wigner-Seitz cell volume v_0, this means we must retain a large number of coefficients in the expansion of Eqn. 4.288.

Suppose, however, an eccentric theorist gives you a good approximation to these core Bloch states. Indeed, according to Eqn. 4.58, if we define $C_n(\boldsymbol{k}) \equiv S^{1/2}_{nn'}(\boldsymbol{k})\,D_{n'}(\boldsymbol{k})$, then the coefficients $D_n(\boldsymbol{k})$ satisfy the eigenvalue equation

$$S^{-1/2}(\boldsymbol{k})\,H(\boldsymbol{k})\,S^{-1/2}(\boldsymbol{k})\boldsymbol{D}_n(\boldsymbol{k}) = E_n(\boldsymbol{k})\,\boldsymbol{D}_n(\boldsymbol{k}) \quad , \tag{4.290}$$

Thus, for each $\boldsymbol{k}$ the eigenvectors $\big\{\boldsymbol{D}_a(\boldsymbol{k})\big\}$ of the Hermitian matrix $\widetilde{H}(\boldsymbol{k}) \equiv S^{-1/2}(\boldsymbol{k})\,H(\boldsymbol{k})\,S^{-1/2}(\boldsymbol{k})$ yield, upon multiplication by $S^{1/2}(\boldsymbol{k})$, the coefficients $C_n(\boldsymbol{k})$. Here we imagine that the indices n and n' are restricted to the core levels alone. This obviates the subtle problem of overcompleteness of the atomic levels arising from the existence of scattering states. Furthermore, we may write

$$\begin{aligned} S_{nn'}(\boldsymbol{k}) &= \delta_{nn'} + \Sigma_{nn'}(\boldsymbol{k}) \\ \Sigma_{nn'}(\boldsymbol{k}) &= \sum_{\boldsymbol{R}\neq 0} S_{nn'}(\boldsymbol{R})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \end{aligned} \tag{4.291}$$

Now since we are talking about core levels, the contribution $\Sigma_{nn'}(\boldsymbol{k})$, which involves overlaps on different sites, is very small. This means the inverse square root, $S^{-1/2} \approx 1 - \frac{1}{2}\Sigma + \mathcal{O}(\Sigma^2)$, can be well-approximated by the first two terms in the expansion in powers of Σ.

We will write $|\,\mathsf{a}\boldsymbol{k}\,\rangle$ for a core Bloch state in band a, and $|\,\mathsf{a}\boldsymbol{R}\,\rangle$ for the corresponding core Wannier state[48]. Now define the projector,

$$\Pi = \sum_{\mathsf{a},\boldsymbol{k}} |\,\mathsf{a}\boldsymbol{k}\,\rangle\langle\,\mathsf{a}\boldsymbol{k}\,| = \sum_{\mathsf{a},\boldsymbol{R}} |\,\mathsf{a}\boldsymbol{R}\,\rangle\langle\,\mathsf{a}\boldsymbol{R}\,| \quad . \tag{4.292}$$

Note that $\Pi^2 = \Pi$, which is a property of projection operators. We now define

$$|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle \equiv \big(1-\Pi\big)|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle \quad , \tag{4.293}$$

where $|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle$ is the plane wave, for which $\langle\,\boldsymbol{r}\,|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle = V^{-1/2}\,e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}$. It is important to note that we continue to restrict $\boldsymbol{k} \in \widehat{\mathcal{L}}$ to the first Brillouin zone. Note that $\Pi\,|\phi_{\boldsymbol{G}+\boldsymbol{k}}\rangle = 0$, *i.e.*

[48]Recall that the Wannier states in a given band are somewhat arbitrary as they depend on a choice of phase.

the state $|\phi_{\boldsymbol{G}+\boldsymbol{k}}\rangle$ has been *orthogonalized* to all the core orbitals. Accordingly, we call

$$\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r}) = \frac{e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}}{\sqrt{V}} \left\{ 1 - \sum_{\mathsf{a}} u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\, e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \int d^d r'\, u^*_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r}')\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}'} \right\} \tag{4.294}$$

an *orthogonalized plane wave* (OPW).

As an example, consider the case of only a single core 1s orbital, whose atomic wavefunction is given by the hydrogenic form $\varphi(\boldsymbol{r}) = \frac{\alpha^{3/2}}{\sqrt{\pi}}\, e^{-\alpha r}$. The core cell function is then approximately

$$u_{\boldsymbol{k}}(\boldsymbol{r}) \approx \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} \varphi(\boldsymbol{r}-\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})} \quad . \tag{4.295}$$

We then have

$$\int d^3 r'\, u^*_{\boldsymbol{k}}(\boldsymbol{r}')\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}'} \approx \sqrt{N}\, \big[\widehat{\varphi}(\boldsymbol{G}+\boldsymbol{k})\big]^* = \frac{8\pi^{1/2}\alpha^{5/2}N^{1/2}}{\big[\alpha^2 + (\boldsymbol{G}+\boldsymbol{k})^2\big]^2} \tag{4.296}$$

and then

$$\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r}) = \frac{e^{i(\boldsymbol{G}+\boldsymbol{k})\cdot\boldsymbol{r}}}{\sqrt{V}} \left\{ 1 - \frac{8\,\alpha^4}{\big[\alpha^2 + (\boldsymbol{G}+\boldsymbol{k})^2\big]^2} \sum_{\boldsymbol{R}} e^{-\alpha|\boldsymbol{r}-\boldsymbol{R}|}\, e^{-i(\boldsymbol{G}+\boldsymbol{k})\cdot(\boldsymbol{r}-\boldsymbol{R})} \right\} \quad . \tag{4.297}$$

For $\boldsymbol{G}+\boldsymbol{k} = 0$, we have, in the $\boldsymbol{R} = 0$ cell,

$$\phi_0(\boldsymbol{r}) \approx \frac{1}{\sqrt{V}} \left(1 - 8\, e^{-\alpha r}\right) \quad . \tag{4.298}$$

Note that the OPW states are not normalized. Indeed, we have[49]

$$\langle\, \phi_{\boldsymbol{G}+\boldsymbol{k}} \,|\, \phi_{\boldsymbol{G}+\boldsymbol{k}} \,\rangle = \int d^d r\, \big|\phi_{\boldsymbol{G}+\boldsymbol{k}}(\boldsymbol{r})\big|^2 = 1 - \frac{1}{V} \sum_{\mathsf{a}} \left| \int d^d r\, u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\, e^{-i\boldsymbol{G}\cdot\boldsymbol{r}'} \right|^2 \quad . \tag{4.299}$$

The energy eigenvalues are then obtained by solving the equation $\det M_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}, E) = 0$ for $E_n(\boldsymbol{k})$, where

$$M_{\boldsymbol{G}\boldsymbol{G}'}(\boldsymbol{k}, E) = \langle\, \phi_{\boldsymbol{G}+\boldsymbol{k}} \,|\, H \,|\, \phi_{\boldsymbol{G}'+\boldsymbol{k}} \,\rangle \tag{4.300}$$

$$= \left[\frac{\hbar^2(\boldsymbol{G}+\boldsymbol{k})^2}{2m} - E \right] \delta_{\boldsymbol{G}\boldsymbol{G}'} + V_{\boldsymbol{G}-\boldsymbol{G}'} + \sum_{\mathsf{a}} \big[E - E_{\mathsf{a}}(\boldsymbol{k})\big]\, \langle\, \boldsymbol{G}+\boldsymbol{k} \,|\, \mathsf{a}\boldsymbol{k} \,\rangle\, \langle\, \mathsf{a}\boldsymbol{k} \,|\, \boldsymbol{G}'+\boldsymbol{k} \,\rangle \quad .$$

The overlap of the plane wave state $|\, \boldsymbol{G}+\boldsymbol{k} \,\rangle$ and the core Bloch state $|\, \mathsf{a}\boldsymbol{k} \,\rangle$ is given by

$$\langle\, \boldsymbol{G}+\boldsymbol{k} \,|\, \mathsf{a}\boldsymbol{k} \,\rangle = N^{1/2}\, v_0^{-1/2} \int\limits_{\Omega} d^d r\, u_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\, e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} \quad . \tag{4.301}$$

[49] Note that $0 \leqslant \langle\, \phi_{\boldsymbol{k}} \,|\, \phi_{\boldsymbol{k}} \,\rangle \leqslant 1$.

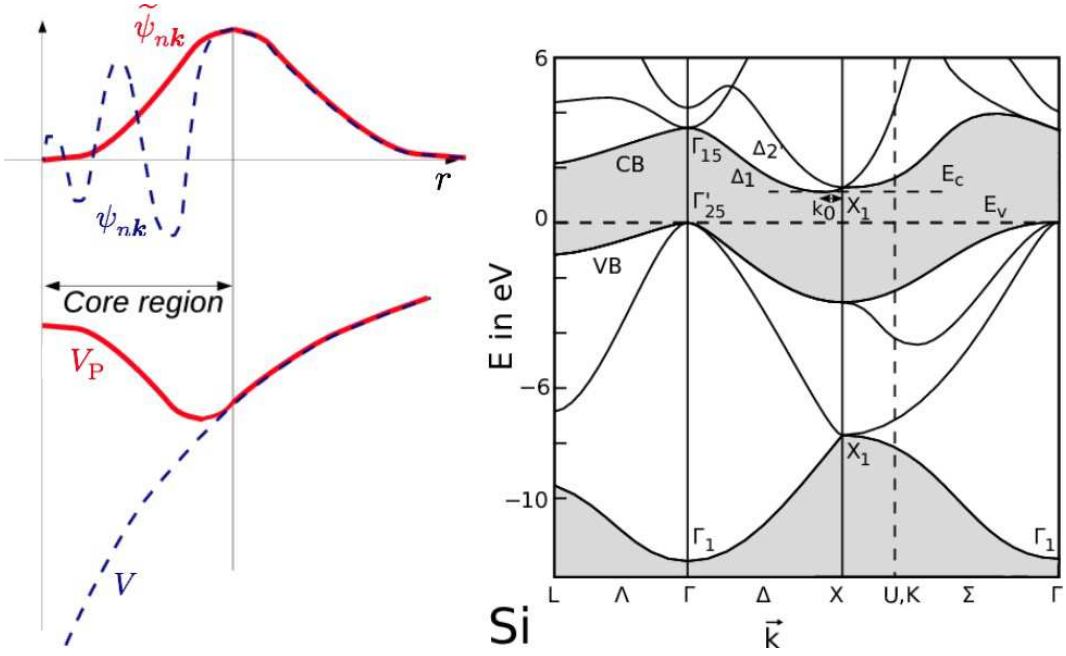


Figure 4.23: Left: Blue curves sketch the potential V in the vicinity of a nucleus, and corresponding valence or conduction band wavefunction $\psi_{n\boldsymbol{k}}$. Note that $\psi_{n\boldsymbol{k}}$ wiggles significantly in the vicinity of the nucleus because it must be orthogonal to the core atomic orbitals. Red curves sketch the pseudopotential $V_{\rm P}$ and corresponding pseudo-wavefunction $\widetilde{\psi}_{n\boldsymbol{k}}$. Right: pseudopotential calculation for the band structure of Si.

4.6.2 The pseudopotential

Equivalently, we may use the $|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle$ in linear combinations to build our Bloch states, *viz.*

$$|\,\psi_{n\boldsymbol{k}}\,\rangle = \sum_{\boldsymbol{G}} C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})\,|\,\phi_{\boldsymbol{G}+\boldsymbol{k}}\,\rangle \equiv (1-\Pi)\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle \quad , \tag{4.302}$$

where $|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle = \sum_{\boldsymbol{G}} C^{(n)}_{\boldsymbol{G}}(\boldsymbol{k})\,|\,\boldsymbol{G}+\boldsymbol{k}\,\rangle$. Note that the *pseudo-wavefunction* $\widetilde{\psi}_{n\boldsymbol{k}}(\boldsymbol{r})$ is a sum over plane waves. What equation does it satisfy? From $H|\,\psi_{n\boldsymbol{k}}\,\rangle = E\,|\,\psi_{n\boldsymbol{k}}\,\rangle$, we derive

$$H|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle + \overbrace{(E-H)\,\Pi}^{V_{\rm R}}\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle = E\,|\,\widetilde{\psi}_{n\boldsymbol{k}}\,\rangle \quad . \tag{4.303}$$

Note that

$$V_{\rm R} = (E-H)\,\Pi = \sum_{\rm a}\big[E - E_{\rm a}(\boldsymbol{k})\big]\,|\,{\rm a}\boldsymbol{k}\,\rangle\langle\,{\rm a}\boldsymbol{k}\,| \tag{4.304}$$

and

$$\langle\,\boldsymbol{r}\,|\,V_{\rm R}\,|\,\widetilde{\psi}\,\rangle = \int\! d^dr'\,V_{\rm R}(\boldsymbol{r},\boldsymbol{r}')\,\widetilde{\psi}(\boldsymbol{r}') \quad , \tag{4.305}$$

where

$$\begin{aligned} V_{\rm R}(\boldsymbol{r},\boldsymbol{r}') &= \sum_{\mathsf{a},\boldsymbol{k}} \big[E - E_{\mathsf{a}}(\boldsymbol{k})\big]\,\psi_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,\psi^*_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r}') \\ &\approx \sum_{\mathsf{a},\boldsymbol{R}} (E - E_{\mathsf{a}})\,\varphi_{\mathsf{a}}(\boldsymbol{r}-\boldsymbol{R})\,\varphi^*_{\mathsf{a}}(\boldsymbol{r}'-\boldsymbol{R}) \quad , \end{aligned} \tag{4.306}$$

where $\varphi_{\mathsf{a}}(\boldsymbol{r})$ is the $\mathsf{a}^{\rm th}$ atomic core wavefunction. The approximation in the second line above is valid in the limit where the core energy bands are dispersionless, and $E_{\mathsf{a}}(\boldsymbol{k})$ is replaced by the $\boldsymbol{k}$-independent atomic eigenvalue E_{a}. We then use $\sum_{\boldsymbol{k}} |\,\mathsf{a}\boldsymbol{k}\,\rangle\langle\,\mathsf{a}\boldsymbol{k}\,| = \sum_{\boldsymbol{R}} |\,\mathsf{a}\boldsymbol{R}\,\rangle\langle\,\mathsf{a}\boldsymbol{R}\,|$, with $\langle\,\boldsymbol{r}\,|\,\mathsf{a}\boldsymbol{R}\,\rangle \approx \varphi_{\mathsf{a}}(\boldsymbol{r}-\boldsymbol{R})$. Because the atomic levels are highly localized, this means $V_{\boldsymbol{R}}(\boldsymbol{r},\boldsymbol{r}')$ is very small unless both $\boldsymbol{r}$ and $\boldsymbol{r}'$ lie within the same core region. Thus, $V_{\rm R}(\boldsymbol{r},\boldsymbol{r}')$ is "almost diagonal" in $\boldsymbol{r}$ and $\boldsymbol{r}'$. Since the energies of interest satisfy $E > E_{\rm a}$, the term $V_{\boldsymbol{R}}$ tends to add to what is a negative (attractive) potential V, and the combination $V_{\rm P} = V + V_{\rm R}$, known as the *pseudopotential*, is in general weaker the original potential. As depicted in the left panel of Fig. 4.23, whereas the actual valence or conduction band Bloch states $\psi_{{\rm c/v},\boldsymbol{k}}(\boldsymbol{r})$ must wiggle rapidly in the vicinity of each nucleus, in order to be orthogonal to the atomic core states and thereby necessitating the contribution of a large number of high wavevector plane wave components, each pseudo-wavefunction $\widetilde{\psi}_{n\boldsymbol{k}}(\boldsymbol{r})$ is unremarkable in the core region, and can be described using far less information.

In fact, there is a great arbitrariness in defining the operator $V_{\rm R}$. Suppose we take $\widetilde{V}_{\rm R} = \Pi\, W$, where W is *any* operator. Note that

$$\widetilde{V}_{\rm R}(\boldsymbol{r},\boldsymbol{r}') = \langle\,\boldsymbol{r}\,|\,\Pi\,W\,|\,\boldsymbol{r}'\,\rangle = \sum_{\mathsf{a},\boldsymbol{k}} \psi_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,\langle\,\mathsf{a}\boldsymbol{k}\,|\,W\,|\,\boldsymbol{r}'\,\rangle \equiv \sum_{\mathsf{a},\boldsymbol{k}} \psi_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r})\,W^*_{\mathsf{a}\boldsymbol{k}}(\boldsymbol{r}') \quad . \tag{4.307}$$

This needn't even be Hermitian! The point is that $H = T + V$ and $\widetilde{H} = T + V + \widetilde{V}_{\rm R}$ have the same eigenvalues so long as they are acting outside the space of core wavefunctions. To see this, let us suppose

$$H\,|\,\psi\,\rangle = E\,|\,\psi\,\rangle \qquad , \qquad \widetilde{H}\,|\,\xi\,\rangle = \widetilde{E}\,|\,\xi\,\rangle \quad . \tag{4.308}$$

Then

$$\begin{aligned} \widetilde{E}\,\langle\,\psi\,|\,\xi\,\rangle = \langle\,\psi\,|\,\widetilde{H}\,|\,\xi\,\rangle &= \langle\,\psi\,|\,(H + \Pi\,W)\,|\,\xi\,\rangle \\ &= E\,\langle\,\psi\,|\,\xi\,\rangle + \langle\,\psi\,|\,\Pi\,W\,|\,\xi\,\rangle \quad . \end{aligned} \tag{4.309}$$

Now if $|\,\psi\,\rangle$ lies in the complement of that part of the Hilbert space spanned by the core states, then $\Pi\,|\,\psi\,\rangle = 0$, and it follows that $(E-\widetilde{E})\,\langle\,\psi\,|\,\xi\,\rangle = 0$, hence $E = \widetilde{E}$, so long as $\langle\,\psi\,|\,\xi\,\rangle \neq 0$. If we want $V_{\rm P}$ to be Hermitian, a natural choice might be $\widetilde{V}_{\rm R} = -\Pi\,V\,\Pi$, which gives $V_{\rm P} = V - \Pi\,V\,\Pi$. This effectively removes from the potential V any component which can be constructed from core states alone.

4.7 Appendix I : Gauss-Bonnet and Pontrjagin

4.7.1 Gauss-Bonnet theorem

There is a deep result in mathematics, the Gauss-Bonnet theorem, which connects the *local geometry* of a two-dimensional manifold to its *global topology*. The content of the theorem is as follows:

$$\int\limits_{\mathcal{M}} dS\ K = 2\pi\,\chi(\mathcal{M}) = 2\pi \sum_i \mathsf{ind}\,(\boldsymbol{V})\ , \tag{4.310}$$

where $\mathcal{M}$ is a 2-manifold (a topological space locally homeomorphic to $\mathbb{R}^2$), K is the local *Gaussian curvature* of $\mathcal{M}$, given by $K = (R_1\,R_2)^{-1}$, where $R_{1,2}$ are the principal radii of curvature at a given point, and dS is the differential area element. Here $\boldsymbol{V}(\boldsymbol{x})$ is a vector field on $\mathcal{M}$, and $\mathsf{ind}_{\boldsymbol{x}_i}(\boldsymbol{V})$ refers to the *index* of $\boldsymbol{V}$ at its $i^{\rm th}$ singularity $\boldsymbol{x}_i$. The index is in general defined relative to any closed curve in $\mathcal{M}$, and is given by the winding number of $\boldsymbol{V}(\boldsymbol{x})$ around the curve, *viz.*

$$\mathop{\mathsf{ind}}_{\mathcal{C}}(\boldsymbol{V}) = \oint\limits_{\mathcal{C}} d\boldsymbol{x}\cdot\boldsymbol{\nabla}\tan^{-1}\left(\frac{V_2(\boldsymbol{x})}{V_1(\boldsymbol{x})}\right)\quad . \tag{4.311}$$

If $\mathcal{C}$ encloses no singularities, then the index necessarily vanishes, but if $\mathcal{C}$ encloses one or more singularities, the index is an integer, given by the winding number of $\boldsymbol{V}$ around the curve $\mathcal{C}$.

The quantity $\chi(\mathcal{M})$ is called the *Euler characteristic* of $\mathcal{M}$ and is given by $\chi(\mathcal{M}) = 2-2g$, where g is the *genus* of $\mathcal{M}$, which is the number of holes (or handles) of $\mathcal{M}$. Furthermore, $\boldsymbol{V}(\boldsymbol{x})$ can be *any* smooth vector field on $\mathcal{M}$, with $\boldsymbol{x}_i$ the singularity points of that vector field[50].

To apprehend the content of the Gauss-Bonnet theorem, it is helpful to consider an example. Let $\mathcal{M} = \mathbb{S}^2$ be the unit 2-sphere, as depicted in fig. 4.24. At any point on the unit 2-sphere, the radii of curvature are degenerate and both equal to $R = 1$, hence $K = 1$. If we integrate the Gaussian curvature over the sphere, we thus get $4\pi = 2\pi\,\chi\big(\mathbb{S}^2\big)$, which says $\chi(\mathbb{S}^2) = 2-2g = 2$, which agrees with $g = 0$ for the sphere. Furthermore, the Gauss-Bonnet theorem says that *any* smooth vector field on $\mathbb{S}^2$ *must* have a singularity or singularities, with the total index summed over the singularities equal to $+2$. The vector field sketched in the left panel of fig. 4.24 has two index $+1$ singularities, which could be taken at the north and south poles, but which could be anywhere. Another possibility, depicted in the right panel of fig. 4.24, is that there is a one singularity with index $+2$.

In fig. 4.25 we show examples of manifolds with genii $g = 1$ and $g = 2$. The case $g = 1$ is the familiar 2-torus, which is topologically equivalent to a product of circles: $\mathbb{T}^2 \cong \mathbb{S}^1 \times \mathbb{S}^1$, and is thus coordinatized by two angles θ_1 and θ_2. A smooth vector field pointing in the direction of increasing θ_1 never vanishes, and thus has no singularities, consistent with $g = 1$ and $\chi\big(\mathbb{T}^2\big) = 0$.

[50] The singularities $\boldsymbol{x}_i$ are fixed points of the dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$.

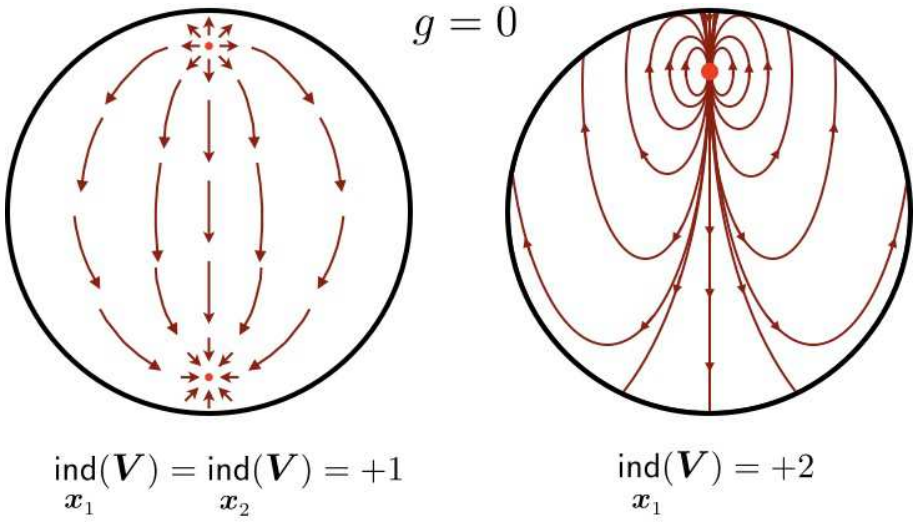


Figure 4.24: Two smooth vector fields on the sphere $\mathbb{S}^2$, which has genus $g = 0$. Left panel: two index $+1$ singularities. Right panel: one index $+2$ singularity.

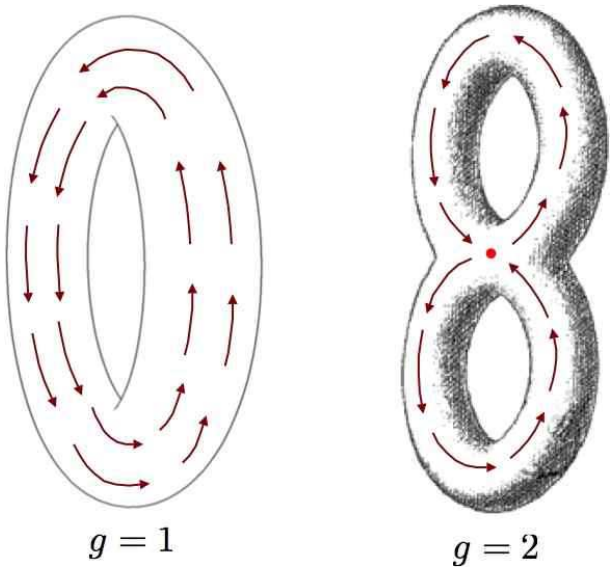


Figure 4.25: Smooth vector fields on the torus $\mathbb{T}^2$, and on a 2-manifold $\mathcal{M}$ of genus $g = 2$

Topologically, one can define a torus as the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, or as a square with opposite sides identified. This is what mathematicians call a 'flat torus' – one with curvature $K = 0$ everywhere. Of course, such a torus cannot be embedded in three-dimensional Euclidean space; a two-dimensional figure embedded in a three-dimensional Euclidean space inherits a metric due to the embedding, and for a physical torus, like the surface of a bagel, the Gaussian curvature is only zero *on average*.

The $g = 2$ surface $\mathcal{M}$ shown in the right panel of fig. 4.25 has Euler characteristic $\chi(\mathcal{M}) = -2$,

which means that any smooth vector field on $\mathcal{M}$ must have singularities with indices totalling -2. One possibility, depicted in the figure, is to have two saddle points with index -1; one of these singularities is shown in the figure (the other would be on the opposite side).

4.7.2 The Pontrjagin index

Consider an N-dimensional vector field $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$, and let $\hat{\boldsymbol{n}}(\boldsymbol{x})$ be the unit vector field defined by $\hat{\boldsymbol{n}}(\boldsymbol{x}) = \boldsymbol{V}(\boldsymbol{x})\big/|\boldsymbol{V}(\boldsymbol{x})|$. Consider now a unit sphere in $\hat{\boldsymbol{n}}$ space, which is of dimension $(N-1)$. If we integrate over this surface, we obtain

$$\Omega_N = \oint d\sigma_a \, n^a = \frac{2\pi^{(N-1)/2}}{\Gamma\left(\frac{N-1}{2}\right)} \,, \tag{4.312}$$

which is the surface area of the unit sphere $\mathbb{S}^{N-1}$. Thus, $\Omega_2 = 2\pi$, $\Omega_3 = 4\pi$, $\Omega_4 = 2\pi^2$, *etc.*

Now consider a change of variables to those over the surface of the sphere, $(\xi_1, \ldots, \xi_{N-1})$. We then have

$$\Omega_N = \oint\limits_{\mathbb{S}^{N-1}} d\sigma_a \, n^a = \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \, \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{4.313}$$

The topological charge is then

$$Q = \frac{1}{\Omega_N} \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \, \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{4.314}$$

The quantity Q is an *integer topological invariant* which characterizes the map from the surface $(\xi_1, \ldots, \xi_{N-1})$ to the unit sphere $|\hat{\boldsymbol{n}}| = 1$. In mathematical parlance, Q is known as the *Pontrjagin index* of this map.

This analytical development recapitulates some basic topology. Let $\mathcal{M}$ be a topological space and consider a map from the circle $\mathbb{S}^1$ to $\mathcal{M}$. We can compose two such maps by merging the two circles, as shown in fig. 4.26. Two maps are said to be *homotopic* if they can be smoothly deformed into each other. Any two homotopic maps are said to belong to the same *equivalence class* or *homotopy class*. For general $\mathcal{M}$, the homotopy classes may be multiplied using the composition law, resulting in a group structure. The group is called the *fundamental group* of the manifold $\mathcal{M}$, and is abbreviated $\pi_1(\mathcal{M})$. If $\mathcal{M} = \mathbb{S}^2$, then any such map can be smoothly contracted to a point on the 2-sphere, which is to say a trivial map. We then have $\pi_1(\mathcal{M}) = 0$. If $\mathcal{M} = \mathbb{S}^1$, the maps can wind nontrivially, and the homotopy classes are labeled by a single integer winding number: $\pi_1(\mathbb{S}^1) = \mathbb{Z}$. The winding number of the composition of two such maps is the sum of their individual winding numbers. If $\mathcal{M} = \mathbb{T}^2$, the maps can wind nontrivially around either of the two cycles of the 2-torus. We then have $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$, and in general $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$. This makes good sense, since an n-torus is topologically equivalent to

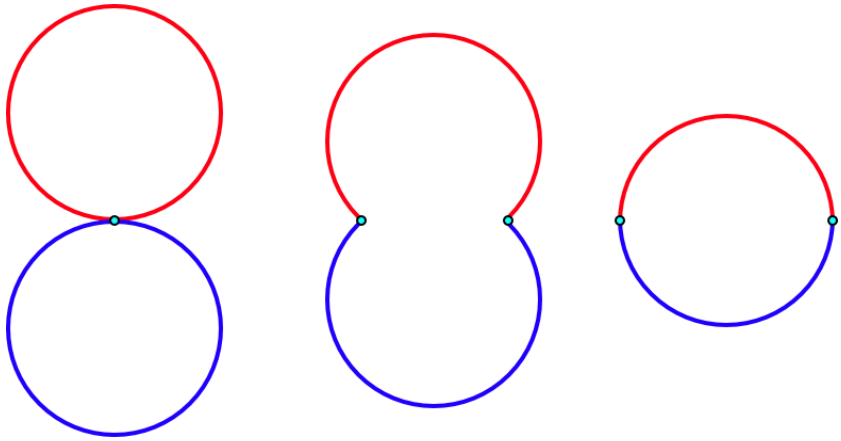


Figure 4.26: Composition of two circles. The same general construction applies to the merging of n-spheres $\mathbb{S}^n$, called the *wedge sum*.

a product of n circles. In some cases, $\pi_1(\mathcal{M})$ can be nonabelian, as is the case when $\mathcal{M}$ is the genus $g = 2$ structure shown in the right hand panel of fig. 4.25.

In general we define the $n^{\rm th}$ *homotopy group* $\pi_n(\mathcal{M})$ as the group under composition of maps from $\mathbb{S}^n$ to $\mathcal{M}$. For $n \geqslant 2$, $\pi_n(\mathcal{M})$ is abelian. If $\dim(\mathcal{M}) < n$, then $\pi_n(\mathcal{M}) = 0$. In general, $\pi_n(\mathbb{S}^n) = \mathbb{Z}$. These $n^{\rm th}$ homotopy classes of the n-sphere are labeled by their associated Pontrjagin index Q.

4.8 Appendix II : Derivation of Eqn. 4.268

Consider the Hamiltonian $H(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} H e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, for which $H(\boldsymbol{k})\,|u_{n\boldsymbol{k}}\rangle = E_n(\boldsymbol{k})\,|u_{n\boldsymbol{k}}\rangle$, where $u_{n\boldsymbol{k}}(\boldsymbol{r}) = \langle\,\boldsymbol{r}\,|\,u_{n\boldsymbol{k}}\,\rangle = e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\,\psi_{n\boldsymbol{k}}(\boldsymbol{r})$ are the cell functions. We consider the wavevector $\boldsymbol{k} = \boldsymbol{k}(t)$ to be an adiabatic parameter which is slowly varying with time. In this setting, the *time-dependent* solutions to the Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}\,|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle = H(\boldsymbol{k})\,|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle \quad , \tag{4.315}$$

are, per Eqn. 4.262, given by

$$|\,\tilde{u}_{n\boldsymbol{k}}(t)\,\rangle = e^{i\phi_{n\boldsymbol{k}}(t)}\,e^{i\gamma_{n\boldsymbol{k}}}\left\{|\,u_{n\boldsymbol{k}}\,\rangle - i\hbar{\sum_{\ell}}'\frac{|\,u_{\ell\boldsymbol{k}}\,\rangle\,\langle\,u_{\ell\boldsymbol{k}}\,|\,d_t\,|\,u_{n\boldsymbol{k}}\,\rangle}{E_\ell(\boldsymbol{k}) - E_n(\boldsymbol{k})}\right\} + \ldots \quad , \tag{4.316}$$

where the prime on the sum indicates that the state $|u_{n\boldsymbol{k}}\rangle$ is excluded. Higher order terms are negligible in the adiabatic limit. Here

$$\phi_{n\boldsymbol{k}}(t) = -\frac{1}{\hbar}\int^t dt'\;E_n\big(\boldsymbol{k}(t')\big) \quad , \tag{4.317}$$

where $\boldsymbol{k} = \boldsymbol{k}(t)$, and

$$\frac{d\gamma_{n\boldsymbol{k}}}{dt} = i \left\langle u_{n\boldsymbol{k}} \right| \frac{d}{dt} \left| u_{n\boldsymbol{k}} \right\rangle = i \left\langle u_{n\boldsymbol{k}} \right| \frac{\partial}{\partial k^\alpha} \left| u_{n\boldsymbol{k}} \right\rangle \frac{dk^\alpha}{dt} \quad . \tag{4.318}$$

We now compute the expectation of the velocity operator, $v^\alpha = \hbar^{-1} \partial_\alpha H(\boldsymbol{k})$, where we abbreviate $\partial_\alpha \equiv \partial/\partial k^\alpha$:

$$\left\langle \tilde{u}_{n\boldsymbol{k}}(t) \right| v^\alpha \left| \tilde{u}_{n\boldsymbol{k}}(t) \right\rangle = \left\langle u_{n\boldsymbol{k}} \right| \hbar^{-1} \partial_\alpha H(\boldsymbol{k}) \left| u_{n\boldsymbol{k}} \right\rangle \tag{4.319}$$
$$- i \sum_\ell{}' \frac{\left\langle u_{n\boldsymbol{k}} \right| \partial_\alpha H(\boldsymbol{k}) \left| u_{\ell\boldsymbol{k}} \right\rangle \left\langle u_{\ell\boldsymbol{k}} \right| d_t \left| u_{n\boldsymbol{k}} \right\rangle - \left\langle u_{n\boldsymbol{k}} \right| d_t \left| u_{\ell\boldsymbol{k}} \right\rangle \left\langle u_{\ell\boldsymbol{k}} \right| \partial_\alpha H(\boldsymbol{k}) \left| u_{n\boldsymbol{k}} \right\rangle}{E_\ell(\boldsymbol{k}) - E_n(\boldsymbol{k})} \quad .$$

Now

$$\frac{\partial}{\partial k^\alpha} \left\langle u_{n\boldsymbol{k}} \right| H(\boldsymbol{k}) \left| u_{\ell\boldsymbol{k}} \right\rangle = \frac{\partial E_n(\boldsymbol{k})}{\partial k^\alpha} \delta_{n\ell} \tag{4.320}$$
$$= E_\ell(\boldsymbol{k}) \left\langle \frac{\partial u_{n\boldsymbol{k}}}{\partial k^\alpha} \middle| u_{\ell\boldsymbol{k}} \right\rangle + \left\langle u_{n\boldsymbol{k}} \middle| \frac{\partial H(\boldsymbol{k})}{\partial k^\alpha} \middle| u_{\ell\boldsymbol{k}} \right\rangle + E_n(\boldsymbol{k}) \left\langle u_{n\boldsymbol{k}} \middle| \frac{\partial u_{\ell\boldsymbol{k}}}{\partial k^\alpha} \right\rangle \quad ,$$

which says

$$\left\langle u_{n\boldsymbol{k}} \right| \partial_\alpha H(\boldsymbol{k}) \left| u_{\ell\boldsymbol{k}} \right\rangle = \partial_\alpha E_n(\boldsymbol{k}) \, \delta_{n\ell} + \big(E_n(\boldsymbol{k}) - E_\ell(\boldsymbol{k})\big) \left\langle u_{n\boldsymbol{k}} \right| \partial_\alpha \left| u_{\ell\boldsymbol{k}} \right\rangle \quad . \tag{4.321}$$

Thus, we have

$$\left\langle \tilde{u}_{n\boldsymbol{k}}(t) \right| v^\alpha \left| \tilde{u}_{n\boldsymbol{k}}(t) \right\rangle = \hbar^{-1} \partial_\alpha E_n(\boldsymbol{k}) + i \sum_\ell{}' \Big(\left\langle u_{n\boldsymbol{k}} \right| \partial_\alpha \left| u_{\ell\boldsymbol{k}} \right\rangle \left\langle u_{\ell\boldsymbol{k}} \right| d_t \left| u_{n\boldsymbol{k}} \right\rangle - \text{c.c.} \Big) \tag{4.322}$$
$$= \hbar^{-1} \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\left\langle \partial_\alpha u_{n\boldsymbol{k}} \right| d_t u_{n\boldsymbol{k}} \left\rangle - \left\langle \partial_\alpha u_{n\boldsymbol{k}} \right| u_{n\boldsymbol{k}} \right\rangle \left\langle u_{n\boldsymbol{k}} \right| d_t u_{n\boldsymbol{k}} \left\rangle - \text{c.c.} \Big)$$
$$= \hbar^{-1} \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\left\langle \partial_\alpha n \right| \partial_\beta n \left\rangle \dot{k}^\beta - \left\langle \partial_\alpha n \right| n \left\rangle \left\langle n \right| \partial_\beta n \left\rangle \dot{k}^\beta - \text{c.c.} \Big)$$
$$= \hbar^{-1} \partial_\alpha E_n(\boldsymbol{k}) - i \Big(\partial_\alpha \left\langle n \right| \partial_\beta n \left\rangle - \partial_\beta \left\langle n \right| \partial_\alpha n \left\rangle \Big) \dot{k}^\beta \quad ,$$

where $\left| \partial_\alpha u_{n\boldsymbol{k}} \right\rangle \equiv \left| \partial_\alpha n \right\rangle$, where we ultimately suppress the wavevector label $\boldsymbol{k}$ for notational convenience. Recall now the definitions in Eqn. 4.269,

$$\mathcal{A}_n^\alpha(\boldsymbol{k}) = i \left\langle u_{n\boldsymbol{k}} \right| \frac{\partial}{\partial k^\alpha} \left| u_{n\boldsymbol{k}} \right\rangle \qquad , \qquad \Omega_n^\alpha(\boldsymbol{k}) = \epsilon_{\alpha\beta\gamma} \frac{\partial \mathcal{A}_n^\gamma(\boldsymbol{k})}{\partial k^\beta} \quad . \tag{4.323}$$

In our abbreviated notation, $\mathcal{A}_n^\alpha = i \left\langle n \right| \partial_\alpha \left| n \right\rangle$, and we obtain

$$\frac{dx^\alpha}{dt} = \left\langle \psi_{n\boldsymbol{k}}(t) \right| v^\alpha \left| \psi_{n\boldsymbol{k}}(t) \right\rangle = \hbar^{-1} \partial_\alpha E_n(\boldsymbol{k}) - \epsilon_{\alpha\beta\gamma} \, \dot{k}^\beta \, \Omega_n^\gamma(\boldsymbol{k}) \quad , \tag{4.324}$$

since

$$\epsilon_{\alpha\beta\gamma} \, \dot{k}^\beta \, \Omega_n^\gamma = \epsilon_{\alpha\beta\gamma} \, \dot{k}^\beta \, \epsilon_{\gamma\alpha\beta} \, \partial_\alpha \, \mathcal{A}_n^\beta = \epsilon_{\alpha\beta\gamma} \, \epsilon_{\alpha\beta\beta} \, \dot{k}^\beta \, \partial_\alpha \, \mathcal{A}_n^\beta = \big(\delta_{\alpha\alpha} \, \delta_{\beta\beta} - \delta_{\alpha\beta} \, \delta_{\beta\alpha} \big) \, \dot{k}^\beta \, \partial_\alpha \, \mathcal{A}_n^\beta$$
$$= \big(\partial_\alpha \mathcal{A}_n^\beta - \partial_\beta \mathcal{A}_n^\alpha \big) \, \dot{k}^\beta = i \Big(\partial_\alpha \left\langle n \right| \partial_\beta \, n \left\rangle - \partial_\beta \left\langle n \right| \partial_\alpha \, n \left\rangle \Big) \, \dot{k}^\beta \quad . \tag{4.325}$$

Chapter 5

Metals

5.1 Introduction

Metals are characterized by a finite electronic density of states $f(\varepsilon_{\mathsf{F}})$ at the Fermi level at zero temperature. This entails a number of salient features, such as thermodynamic, electrodynamic, and transport properties.

5.2 $T = 0$ and the Fermi Surface

5.2.1 Definition of the Fermi surface

The Pauli principle says that each fermionic energy state can accommodate either zero or one electrons[1]. At zero temperature, the ground state of a noninteracting Fermi gas is obtained by filling up all the distinct eigenstates in order of energy, starting from the bottom of the spectrum, until all the fermions are used up. The energy of the last level to be filled is called the *Fermi energy*, and is written ε_{F}. The energy distribution function at $T = 0$ is thus $n(\varepsilon) = \Theta(\varepsilon_{\mathsf{F}} - \varepsilon)$, which says that all single particle energy states up to $\varepsilon = \mu$ are filled, and all energy states above $\varepsilon = \mu$ are empty. As we shall see in the next section, the Fermi energy is the zero temperature value of the chemical potential: $\varepsilon_{\mathsf{F}} = \mu(T = 0)$. If the single particle dispersion $\varepsilon(\boldsymbol{k})$ depends only on the wavevector $\boldsymbol{k}$, then the locus of points in $\boldsymbol{k}$-space for which $\varepsilon(\boldsymbol{k}) = \varepsilon_{\mathsf{F}}$ is called the *Fermi surface*. For isotropic systems, $\varepsilon(\boldsymbol{k}) = \varepsilon(k)$ is a function only of the magnitude $k = |\boldsymbol{k}|$, and the Fermi surface is a sphere in $d = 3$ or a circle in $d = 2$. The radius of this circle is the *Fermi wavevector*, k_{F}. When there is internal (*e.g.* spin) degree of freedom, there is a Fermi surface and Fermi wavevector (for isotropic systems) for each polarization state of the internal degree of

[1] We consider two degenerate energy states with different spin polarizations $\uparrow$ and $\downarrow$ to be distinct quantum states.

freedom.

Let's compute the Fermi wavevector $k_{\rm F}$ and Fermi energy $\varepsilon_{\rm F}$ for the IFG with a ballistic dispersion $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$. We allow for a common degeneracy g for each of the $\boldsymbol{k}$ states, *e.g.*, due to spin, for which $\mathsf{g} = 2S+1$, with $S = \frac{1}{2}$ for electrons. The number density is

$$n = \mathsf{g}\int d^dk\;\Theta(k_{\rm F}-k) = \frac{\mathsf{g}\,\Omega_d}{(2\pi)^d}\cdot\frac{k_{\rm F}^d}{d} = \begin{cases} \mathsf{g}\,k_{\rm F}/\pi & (d=1) \\ \mathsf{g}\,k_{\rm F}^2/4\pi & (d=2) \\ \mathsf{g}\,k_{\rm F}^3/6\pi^2 & (d=3) \end{cases}, \tag{5.1}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in d space dimensions ($\Omega_1 = 2$, $\Omega_2 = 2\pi$, $\Omega_3 = 4\pi$, *etc.*). Note that the form of $n(k_{\rm F})$ is independent of the dispersion relation, so long as it remains isotropic. Inverting the above expressions, we obtain $k_{\rm F}(n)$:

$$k_{\rm F} = 2\pi\left(\frac{d\,n}{\mathsf{g}\,\Omega_d}\right)^{1/d} = \begin{cases} \pi n/\mathsf{g} & (d=1) \\ (4\pi n/\mathsf{g})^{1/2} & (d=2) \\ (6\pi^2 n/\mathsf{g})^{1/3} & (d=3) \end{cases}. \tag{5.2}$$

The Fermi energy in each case, for ballistic dispersion, is therefore

$$\varepsilon_{\rm F} = \frac{\hbar^2 k_{\rm F}^2}{2m} = \frac{2\pi^2\hbar^2}{m}\left(\frac{d\,n}{\mathsf{g}\,\Omega_d}\right)^{2/d} = \begin{cases} \frac{\pi^2\hbar^2 n^2}{2\mathsf{g}^2 m} & (d=1) \\ \frac{2\pi\hbar^2\,n}{\mathsf{g}\,m} & (d=2) \\ \frac{\hbar^2}{2m}\left(\frac{6\pi^2 n}{\mathsf{g}}\right)^{2/3} & (d=3) \end{cases}. \tag{5.3}$$

Another useful result for the ballistic dispersion, which follows from the above, is that the density of states at the Fermi level is given by

$$g(\varepsilon_{\rm F}) = \frac{\mathsf{g}\,\Omega_d}{(2\pi)^d}\cdot\frac{m k_{\rm F}^{d-2}}{\hbar^2} = \frac{d}{2}\cdot\frac{n}{\varepsilon_{\rm F}}\quad . \tag{5.4}$$

For the electron gas, we have $\mathsf{g} = 2$. In a metal, one typically has $k_{\rm F} \sim 0.5\,\text{Å}^{-1}$ to $2\,\text{Å}^{-1}$, and $\varepsilon_{\rm F} \sim 1\,\text{eV} - 10\,\text{eV}$. Due to the effects of the crystalline lattice, electrons in a solid behave as if they had an *effective mass* m^* which is typically on the order of the electron mass but very often about an order of magnitude smaller, particularly in semiconductors.

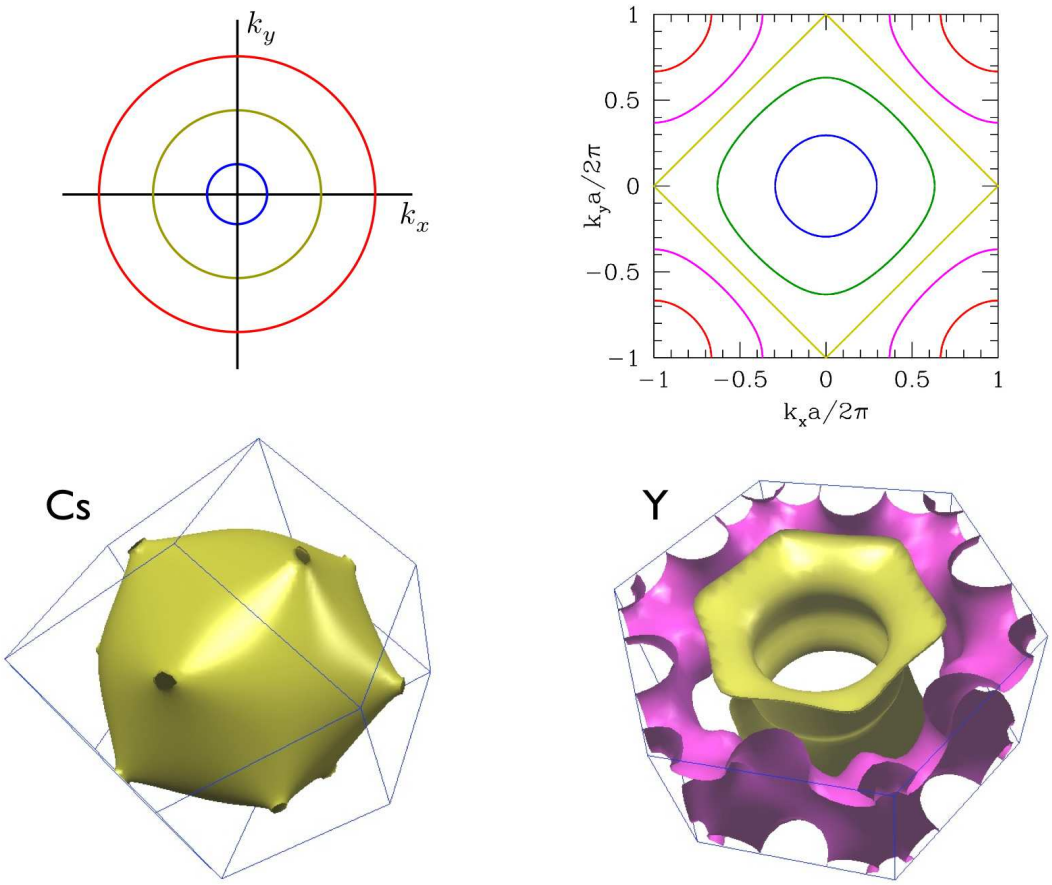


Figure 5.1: Fermi surfaces for two and three-dimensional structures. Upper left: free particles in two dimensions. Upper right: 'tight binding' electrons on a square lattice. Lower left: Fermi surface for cesium, which is predominantly composed of electrons in the $6s$ orbital shell. Lower right: the Fermi surface of yttrium has two parts. One part (yellow) is predominantly due to $5s$ electrons, while the other (pink) is due to $4d$ electrons. (Source: www.phys.ufl.edu/fermisurface/)

In solids, the dispersions $\varepsilon(\boldsymbol{k})$ are in general anisotropic, and give rise to non-spherical Fermi surfaces. The simplest example is that of a two-dimensional tight-binding model of electrons hopping on a square lattice, as may be appropriate in certain layered materials. The dispersion relation is then

$$\varepsilon(k_x, k_y) = -2t\,\cos(k_x a) - 2t\,\cos(k_y a) \quad , \tag{5.5}$$

where k_x and k_y are confined to the interval $\left[-\frac{\pi}{a}, \frac{\pi}{a}\right]$. The quantity t has dimensions of energy and is known as the *hopping integral*. The Fermi surface is the set of points (k_x, k_y) which satisfies $\varepsilon(k_x, k_y) = \varepsilon_{\rm F}$. When $\varepsilon_{\rm F}$ achieves its minimum value of $\varepsilon_{\rm F}^{\rm min} = -4t$, the Fermi surface collapses to a point at $(k_x, k_y) = (0, 0)$. For energies just above this minimum value, we can

expand the dispersion in a power series, writing

$$\varepsilon(k_x, k_y) = -4t + ta^2\left(k_x^2 + k_y^2\right) - \tfrac{1}{12}\,ta^4\left(k_x^4 + k_y^4\right) + \ldots \quad . \tag{5.6}$$

If we only work to quadratic order in k_x and k_y, the dispersion is isotropic, and the Fermi surface is a circle, with $k_{\rm F}^2 = (\varepsilon_{\rm F} + 4t)/ta^2$. As the energy increases further, the continuous $\mathrm{O}(2)$ rotational invariance is broken down to the discrete group of rotations of the square, C_{4v}. The Fermi surfaces distort and eventually, at $\varepsilon_{\rm F} = 0$, the Fermi surface is itself a square. As $\varepsilon_{\rm F}$ increases further, the square turns back into a circle, but centered about the point $\left(\frac{\pi}{a}, \frac{\pi}{a}\right)$. Note that everything is periodic in k_x and k_y modulo $\frac{2\pi}{a}$. The Fermi surfaces for this model are depicted in the upper right panel of Fig. 5.1.

Fermi surfaces in three dimensions can be very interesting indeed, and of great importance in understanding the electronic properties of solids. Two examples are shown in the bottom panels of Fig. 5.1. The electronic configuration of cesium (Cs) is $[\mathrm{Xe}]\,6s^1$. The $6s$ electrons 'hop' from site to site on a body centered cubic (BCC) lattice, a generalization of the simple two-dimensional square lattice hopping model discussed above. The elementary unit cell in $\boldsymbol{k}$ space, known as the *first Brillouin zone*, turns out to be a dodecahedron. In yttrium, the electronic structure is $[\mathrm{Kr}]\,5s^2\,4d^1$, and there are two electronic energy bands at the Fermi level, meaning two Fermi surfaces. Yttrium forms a hexagonal close packed (HCP) crystal structure, and its first Brillouin zone is shaped like a hexagonal pillbox.

5.2.2 Fermi surface vs. Brillouin zone

The construction of the first Brillouin zone proceeds as follows. Draw the bisecting planes ($d = 3$) or lines ($d = 2$) for each of the reciprocal lattice vectors $\boldsymbol{G} = \sum_{\mu=1}^{d} n_\mu\,\boldsymbol{b}_\mu$. The region bounded by these bisectors which contains the origin is the first Brillouin zone. The regions for which a minimum of one bisector is crossed in order to get to the first zone defines the second zone. Points for which a minimum of $(j-1)$ bisectors must be crossed to arrive in the first zone comprise the $j^{\rm th}$ zone. For the square lattice, this scheme is depicted in Fig. 5.2. By shifting all the various fragments of the $j^{\rm th}$ zone by reciprocal lattice vectors, one can completely cover the first zone, with no overlapping areas. Thus, the volume of each of the zones is always $\hat{v}_0$.

Suppose there are Z electrons per unit cell. The Fermi wavevector $k_{\rm F}$ is determined by $k_{\rm F} = (2\pi n)^{1/2}$ with $na^2 = Z$. The side length of the Brillouin zone is $b = 2\pi/a$. Thus, the ratio of the diameter of the free electron Fermi circle to the elementary RLV is

$$r \equiv \frac{2k_{\rm F}}{b} = \sqrt{\frac{2Z}{\pi}} \quad . \tag{5.7}$$

If $r < 1$, the Fermi circle lies entirely within the first Brillouin zone $\hat{\Omega}$. This is the case for $Z = 1$, when $r = 0.798$, but for $Z = 2$ the area of the Fermi circle is precisely the Brillouin zone area,

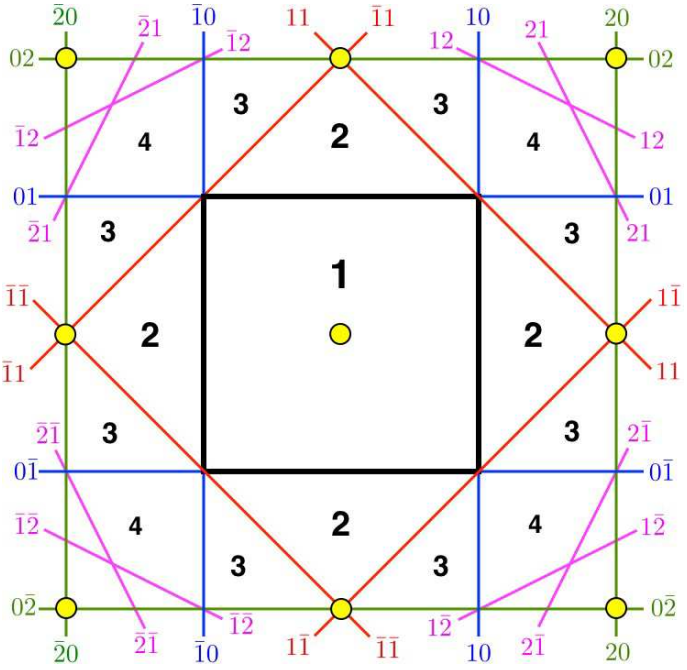


Figure 5.2: Extended zones and their folding for the square lattice. Reciprocal lattice points are shown as yellow dots. The central square, defined by the thick black line, is the first Brillouin zone. The colored lines denote bisectors of reciprocal lattice vectors $\boldsymbol{G} = n_1\boldsymbol{b}_1 + n_2\boldsymbol{b}_2$ and are labeled by n_1 and n_2. In general, a minimum of $j-1$ such bisectors must be crossed in going from the $j^{\rm th}$ Brillouin zone to the first Brillouin zone., with j labeled in black. (Not all of the fourth zone is labeled.)

and $r = (4/\pi)^{1/2} = 1.128$, so the Fermi circle spills over into the second zone. The situation is depicted in the left panel of Fig. 5.3. Since $r < \sqrt{2}$, it does not cross any of the red lines in the left panel, *i.e.* the Fermi circle is confined to the first and second zones. The effect of a weak crystalline potential, as we have seen, is to introduce energy gaps along the Brillouin zone boundaries. If the crystalline potential is strong enough, it can pull all of the states from the second zone into the first zone, completely filling it , thereby resulting in a band insulator.

When $Z = 3$, find $r = 1.382 < \sqrt{2}$, so again the Fermi surface lies only within the first and second zones. For $Z = 4$, $r = 1.596 > \sqrt{2}$, and as we see in the right panel of the figure, the Fermi sea completely encloses the first zone, and spills over into zones two, three, and four.

What happens in $d = 3$ dimensions? Fig. 5.4 shows some examples. Sodium (Na) is monovalent, and the volume of its free electron Fermi sphere is half that of the Brillouin zone and fits entirely within $\hat{\Omega}$. The crystal structure is bcc and the first Brillouin zone has the shape of a rhombic dodecahedron. Copper (Cu) is also monovalent, but the crystalline potential is stronger and leads to the eight Fermi surface 'necks' shown in the figure. The crystal structure is fcc, and the Brillouin zone has the shape of a truncated octahedron. The necks straddle the eight hexagonal faces of the first zone. Calcium (Ca) is divalent, hence the free electron Fermi

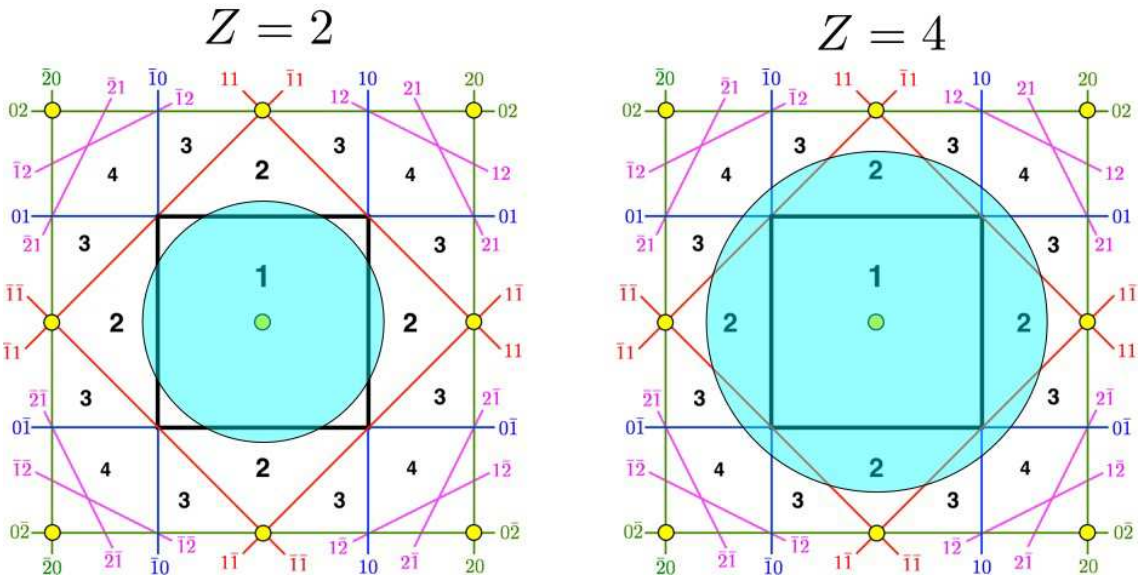


Figure 5.3: Brillouin zones and free electron Fermi seas (in blue) for the square lattice. Left: $Z = 2$ electrons per cell. The Fermi surface has area $\hat{v}_0$, and the free electron Fermi sea extends into the second zone. Right: $Z = 4$ electrons per cell. The Fermi surface has area $2\hat{v}_0$. and the free electron Fermi sea completely covers the first zone, and extends into portions of the second, third, and fourth zones.

sphere has exactly the same volume as that of the first Brillouin zone. Thus, this sphere must cut across the Brillouin zone boundaries, resulting in two bands, the Fermi surface in the first of which is depicted in the figure. The lattice potential pulls most but not all of the states in the second zone into the first zone.

5.2.3 Spin-split Fermi surfaces

Consider an electron gas in an external magnetic field H. The single particle Hamiltonian is then

$$\hat{H} = \frac{\boldsymbol{p}^2}{2m} + \mu_{\rm B} H \sigma \quad , \tag{5.8}$$

where $\mu_{\rm B}$ is the *Bohr magneton*,

$$\begin{aligned} \mu_{\rm B} &= \frac{e\hbar}{2mc} = 5.788 \times 10^{-9}\,{\rm eV/G} \\ \mu_{\rm B}/k_{\rm B} &= 6.717 \times 10^{-5}\,{\rm K/G} \quad , \end{aligned} \tag{5.9}$$

where m is the electron mass. What happens at $T = 0$ to a noninteracting electron gas in a magnetic field?

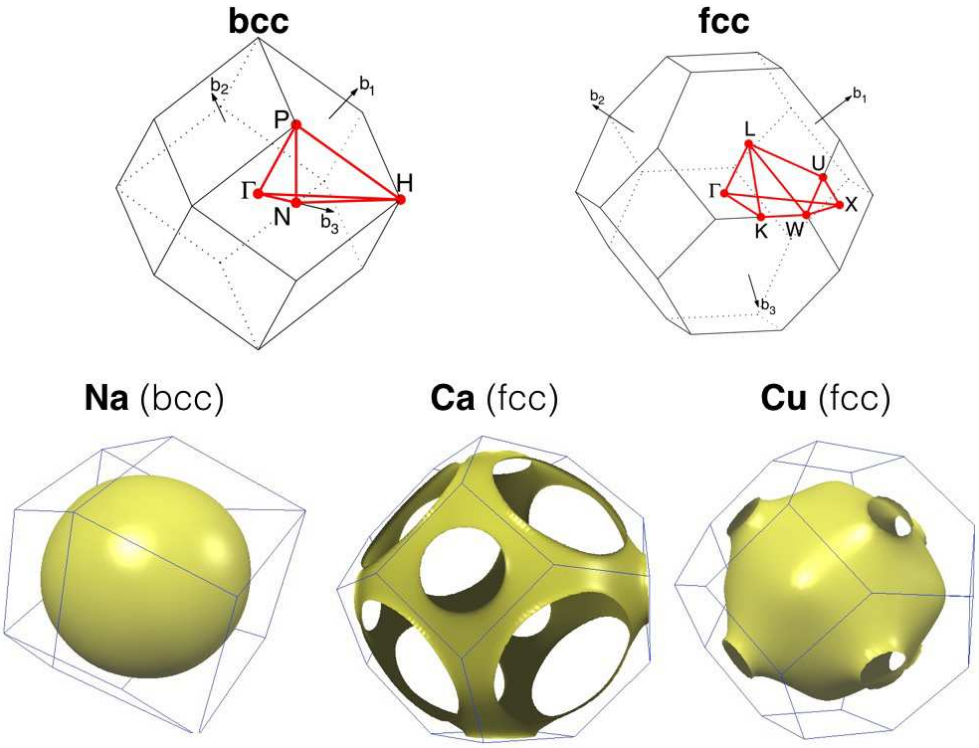


Figure 5.4: Top: First Brillouin zones for bcc (left) and fcc (right) lattice solids. Bottom: Fermi surfaces for Na (left), Ca (center), and Cu (right). (Source: www.phys.ufl.edu/fermisurface/)

Electrons of each spin polarization form their own Fermi surfaces. That is, there is an up spin Fermi surface, with Fermi wavevector $k_{\rm F\uparrow}$, and a down spin Fermi surface, with Fermi wavevector $k_{\rm F\downarrow}$. The individual Fermi energies, on the other hand, must be equal, hence

$$\frac{\hbar^2 k_{\rm F\uparrow}^2}{2m} + \mu_{\rm B} H = \frac{\hbar^2 k_{\rm F\downarrow}^2}{2m} - \mu_{\rm B} H \quad , \tag{5.10}$$

which says

$$k_{\rm F\downarrow}^2 - k_{\rm F\uparrow}^2 = \frac{2eH}{\hbar c} \quad . \tag{5.11}$$

The total density is

$$n = \frac{k_{\rm F\uparrow}^3}{6\pi^2} + \frac{k_{\rm F\downarrow}^3}{6\pi^2} \qquad \Longrightarrow \qquad k_{\rm F\uparrow}^3 + k_{\rm F\downarrow}^3 = 6\pi^2 n \quad . \tag{5.12}$$

Clearly the down spin Fermi surface grows and the up spin Fermi surface shrinks with increasing H. Eventually, the minority spin Fermi surface vanishes altogether. This happens for the up spins when $k_{\rm F\uparrow} = 0$. Solving for the critical field, we obtain

$$H_{\rm c} = \frac{\hbar c}{2e} \cdot \left(6\pi^2 n\right)^{1/3} \quad . \tag{5.13}$$

In real magnetic solids, like cobalt and nickel, the spin-split Fermi surfaces are not spheres, just like the case of the (spin degenerate) Fermi surfaces for Cs and Y shown in Fig. 5.1.

5.3 Quantum Thermodynamics of the Electron Gas

Electrons are fermions, and from this flows some universal features of their thermodynamic properties. We shall assume for the moment that the electrons are noninteracting, or that their mutual interactions can be treated within a "mean field" scheme. In this case, the grand potential of the electron gas is given by

$$\begin{aligned}\Omega(T,V,\mu) &= -Vk_{\rm B}T\sum_\alpha \ln\Big(1+e^{\mu/k_{\rm B}T}\,e^{-\varepsilon_\alpha/k_{\rm B}T}\Big) \\ &= -Vk_{\rm B}T\int\limits_{-\infty}^{\infty}\!d\varepsilon\;g(\varepsilon)\;\ln\Big(1+e^{(\mu-\varepsilon)/k_{\rm B}T}\Big) \quad .\end{aligned} \tag{5.14}$$

The average number of particles in a state with energy ε is

$$n(\varepsilon) = \frac{1}{e^{(\varepsilon-\mu)/k_{\rm B}T}+1} \quad , \tag{5.15}$$

hence the total number of particles is

$$N = V\int\limits_{-\infty}^{\infty}\!d\varepsilon\;g(\varepsilon)\,\frac{1}{e^{(\varepsilon-\mu)/k_{\rm B}T}+1} \quad . \tag{5.16}$$

5.3.1 Fermi distribution

We define the function $f(x) = 1/(e^{\beta x}+1)$, known as the *Fermi distribution*. In the $T\to\infty$ limit, $f(\epsilon)\to\frac{1}{2}$ for all finite values of ε. As $T\to 0$, $f(\epsilon)$ approaches a step function $\Theta(-\epsilon)$. The average number of particles in a state of energy ε in a system at temperature T and chemical potential μ is $n(\varepsilon) = f(\varepsilon-\mu)$. In Fig. 5.5 we plot $f(\varepsilon-\mu)$ *versus* ε for three representative temperatures.

5.3.2 Sommerfeld expansion

In dealing with the ideal Fermi gas, we will repeatedly encounter integrals of the form

$$\mathcal{I}(T,\mu) \equiv \int\limits_{-\infty}^{\infty}\!d\varepsilon\;f(\varepsilon-\mu)\,\phi(\varepsilon) \quad . \tag{5.17}$$

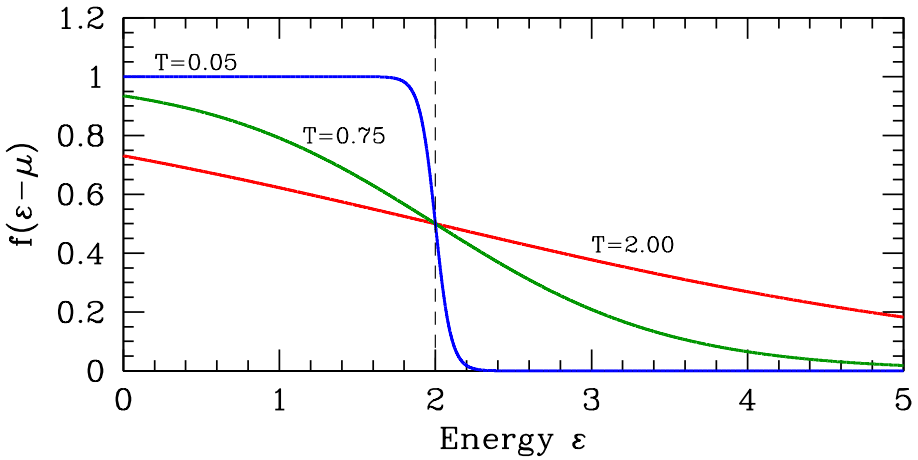


Figure 5.5: The Fermi distribution, $f(\varepsilon-\mu) = \big[\exp\big((\varepsilon-\mu)/k_{\rm B}T\big)+1\big]^{-1}$. Here we have set $k_{\rm B}=1$ and taken $\mu=2$, with $T=\frac{1}{20}$ (blue), $T=\frac{3}{4}$ (green), and $T=2$ (red). In the $T\to 0$ limit, $f(\varepsilon-\mu)$ approaches a step function $\Theta(\mu-\varepsilon)$.

The Sommerfeld expansion provides a systematic way of expanding these expressions in powers of T and is an important analytical tool in analyzing the low temperature properties of the ideal Fermi gas (IFG). We start by defining

$$\Phi(\varepsilon) \equiv \int\limits_{-\infty}^{\varepsilon} d\varepsilon'\,\phi(\varepsilon') \tag{5.18}$$

so that $\phi(\varepsilon)=\Phi'(\varepsilon)$. We then have

$$\mathcal{I} = \int\limits_{-\infty}^{\infty} d\varepsilon\, f(\varepsilon-\mu)\,\frac{d\Phi}{d\varepsilon} = -\int\limits_{-\infty}^{\infty} d\varepsilon\, f'(\varepsilon)\,\Phi(\mu+\varepsilon) \quad , \tag{5.19}$$

where we assume $\Phi(-\infty)=0$. Next, we invoke Taylor's theorem, to write

$$\Phi(\mu+\varepsilon) = \sum_{n=0}^{\infty}\frac{\varepsilon^n}{n\,!}\,\frac{d^n\Phi}{d\mu^n} = \exp\left(\varepsilon\,\frac{d}{d\mu}\right)\Phi(\mu) \quad . \tag{5.20}$$

This last expression involving the exponential of a differential operator may appear overly formal but it proves extremely useful. Since

$$f'(\varepsilon) = -\frac{1}{k_{\rm B}T}\,\frac{e^{\varepsilon/k_{\rm B}T}}{\big(e^{\varepsilon/k_{\rm B}T}+1\big)^2} \quad , \tag{5.21}$$

we can write

$$\mathcal{I} = \int\limits_{-\infty}^{\infty} dv\,\frac{e^{vD}}{(e^v+1)(e^{-v}+1)}\,\Phi(\mu) \quad , \tag{5.22}$$

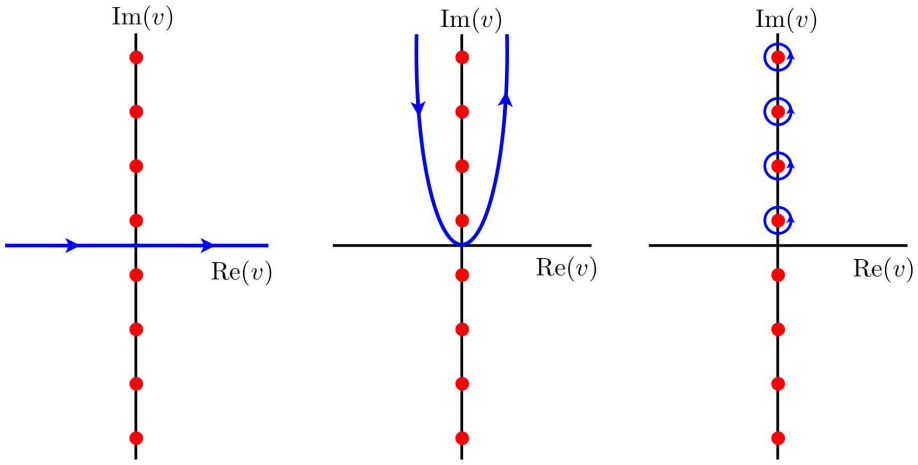


Figure 5.6: Deformation of the complex integration contour in Eqn. 5.23.

with $v = \varepsilon/k_{\rm B}T$, where $D = k_{\rm B}T\,\frac{d}{d\mu}$ is a dimensionless differential operator. The integral can now be done using the methods of complex integration:[2]

$$\begin{aligned}\int\limits_{-\infty}^{\infty}\!dv\;\frac{e^{vD}}{(e^v+1)(e^{-v}+1)} &= 2\pi i\sum_{n=1}^{\infty}\mathrm{Res}\left[\frac{e^{vD}}{(e^v+1)(e^{-v}+1)}\right]_{v=(2n+1)i\pi} \\ &= -2\pi i\sum_{n=0}^{\infty}D\,e^{(2n+1)i\pi D} = -\frac{2\pi i D\,e^{i\pi D}}{1-e^{2\pi i D}} = \pi D\,\csc\pi D \quad .\end{aligned} \tag{5.23}$$

Thus,

$$\mathcal{I}(T,\mu) = \pi D\,\csc(\pi D)\,\Phi(\mu) \quad , \tag{5.24}$$

which is to be understood as the differential operator $\pi D\,\csc(\pi D) = \pi D/\sin(\pi D)$ acting on the function $\Phi(\mu)$. Appealing once more to Taylor's theorem, we have

$$\pi D\,\csc(\pi D) = 1 + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\frac{d^2}{d\mu^2} + \frac{7\pi^4}{360}\,(k_{\rm B}T)^4\,\frac{d^4}{d\mu^4} + \ldots \quad . \tag{5.25}$$

Thus,

$$\mathcal{I}(T,\mu) = \int\limits_{-\infty}^{\infty}\!d\varepsilon\;f(\varepsilon-\mu)\,\phi(\varepsilon) = \int\limits_{-\infty}^{\mu}\!d\varepsilon\;\phi(\varepsilon) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\phi'(\mu) + \frac{7\pi^4}{360}\,(k_{\rm B}T)^4\,\phi'''(\mu) + \ldots \quad . \tag{5.26}$$

If $\phi(\varepsilon)$ is a polynomial function of its argument, then each derivative effectively reduces the order of the polynomial by one degree, and the dimensionless parameter of the expansion is $(T/\mu)^2$. This procedure is known as the *Sommerfeld expansion*.

[2]Note that writing $v = (2n+1)\,i\pi + \epsilon$ we have $e^{\pm v} = -1 \mp \epsilon - \frac{1}{2}\epsilon^2 + \ldots$, so $(e^v+1)(e^{-v}+1) = -\epsilon^2 + \ldots$ We then expand $e^{vD} = e^{(2n+1)i\pi D}\big(1 + \epsilon D + \ldots\big)$ to find the residue: $\mathrm{Res} = -D\,e^{(2n+1)i\pi D}$.

5.3.3 Chemical potential shift

As our first application of the Sommerfeld expansion formalism, let us compute $\mu(n,T)$ for the ideal Fermi gas. The number density $n(T,\mu)$ is

$$n = \int\limits_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, f(\varepsilon-\mu) = \int\limits_{-\infty}^{\mu} d\varepsilon\, g(\varepsilon) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\, g'(\mu) + \ldots \quad . \tag{5.27}$$

Let us write $\mu = \varepsilon_{\rm F} + \delta\mu$, where $\varepsilon_{\rm F} = \mu(T=0,n)$ is the Fermi energy, which is the chemical potential at $T=0$. We then have

$$\begin{aligned} n &= \int\limits_{-\infty}^{\varepsilon_{\rm F}+\delta\mu} d\varepsilon\, g(\varepsilon) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\, g'(\varepsilon_{\rm F}+\delta\mu) + \ldots \\ &= \int\limits_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\, g(\varepsilon) + g(\varepsilon_{\rm F})\,\delta\mu + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\, g'(\varepsilon_{\rm F}) + \ldots \quad , \end{aligned} \tag{5.28}$$

from which we derive

$$\delta\mu = -\frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\frac{g'(\varepsilon_{\rm F})}{g(\varepsilon_{\rm F})} + \mathcal{O}(T^4) \quad . \tag{5.29}$$

Note that $g'/g = (\ln g)'$. For a ballistic dispersion, assuming $\mathsf{g} = 2$,

$$g(\varepsilon) = 2\int \frac{d^3k}{(2\pi)^3}\,\delta\!\left(\varepsilon - \frac{\hbar^2k^2}{2m}\right) = \left.\frac{m\,k(\varepsilon)}{\pi^2\hbar^2}\right|_{k(\varepsilon)=\frac{1}{\hbar}\sqrt{2m\varepsilon}} \tag{5.30}$$

Thus, $g(\varepsilon) \propto \varepsilon^{1/2}$ and $(\ln g)' = \frac{1}{2}\,\varepsilon^{-1}$, so

$$\mu(n,T) = \varepsilon_{\rm F} - \frac{\pi^2}{12}\,\frac{(k_{\rm B}T)^2}{\varepsilon_{\rm F}} + \ldots \quad , \tag{5.31}$$

where $\varepsilon_{\rm F}(n) = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3}$.

5.3.4 Specific heat

The energy of the electron gas is

$$\begin{aligned}\frac{E}{V} &= \int_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\,\varepsilon\, f(\varepsilon-\mu) = \int_{-\infty}^{\mu} d\varepsilon\, g(\varepsilon)\,\varepsilon + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\frac{d}{d\mu}\Big(\mu\, g(\mu)\Big) + \ldots \\ &= \int_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\, g(\varepsilon)\,\varepsilon + g(\varepsilon_{\rm F})\,\varepsilon_{\rm F}\,\delta\mu + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\varepsilon_{\rm F}\,g'(\varepsilon_{\rm F}) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,g(\varepsilon_{\rm F}) + \ldots \\ &= {\rm e}_0 + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,g(\varepsilon_{\rm F}) + \ldots \quad ,\end{aligned} \tag{5.32}$$

where ${\rm e}_0 = \int_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\; g(\varepsilon)\,\varepsilon$ is the ground state energy density (*i.e.* ground state energy per unit volume). Thus,

$$C_{V,N} = \left(\frac{\partial E}{\partial T}\right)_{V,N} = \frac{\pi^2}{3}\,V k_{\rm B}^2\,T\,g(\varepsilon_{\rm F}) \equiv V\gamma\,T \quad , \tag{5.33}$$

where

$$\gamma = \frac{\pi^2}{3}\,k_{\rm B}^2\,g(\varepsilon_{\rm F}) \quad . \tag{5.34}$$

Note that the molar heat capacity is

$$c_V = \frac{N_{\rm A}}{N}\cdot C_V = \frac{\pi^2}{3}R\cdot\frac{k_{\rm B}T\,g(\varepsilon_{\rm F})}{n} = \frac{\pi^2}{2}\left(\frac{k_{\rm B}T}{\varepsilon_{\rm F}}\right)R \quad , \tag{5.35}$$

where in the last expression on the RHS we have assumed a ballistic dispersion, for which

$$\frac{g(\varepsilon_{\rm F})}{n} = \frac{{\rm g}\,m k_{\rm F}}{2\pi^2\hbar^2}\cdot\frac{6\pi^2}{{\rm g}\,k_{\rm F}^3} = \frac{3}{2\,\varepsilon_{\rm F}} \quad . \tag{5.36}$$

The molar heat capacity in Eqn. 5.35 is to be compared with the classical ideal gas value of $\frac{3}{2}R$. Relative to the classical ideal gas, the IFG value is reduced by a fraction of $(\pi^2/3)\times(k_{\rm B}T/\varepsilon_{\rm F})$, which in most metals is very small and even at room temperature is only on the order of 10^{-2}. Most of the heat capacity of metals at room temperature is due to the energy stored in lattice vibrations.

5.4 Effects of External Magnetic Fields

5.4.1 Magnetic susceptibility and Pauli paramagnetism

Magnetism has two origins: (i) orbital currents of charged particles, and (ii) intrinsic magnetic moment. The intrinsic magnetic moment $\boldsymbol{m}$ of a particle is related to its quantum mechanical

spin via

$$\boldsymbol{m} = \overline{g}\mu_0 \boldsymbol{S}/\hbar \quad , \qquad \mu_0 = \frac{q\hbar}{2mc} = \text{magneton} \quad , \tag{5.37}$$

where $\overline{g}$ is the particle's g-factor[3], μ_0 its magnetic moment, and $\boldsymbol{S}$ is the vector of quantum mechanical spin operators satisfying $\left[S^\alpha\,,\,S^\beta\right] = i\hbar\epsilon_{\alpha\beta\gamma}\,S^\gamma$, *i.e.* SU(2) commutation relations. The Hamiltonian for a single particle is then

$$\hat{H} = \frac{1}{2m^*}\Big(\boldsymbol{p} - \frac{q}{c}\boldsymbol{A}\Big)^2 - \boldsymbol{H}\cdot\boldsymbol{m} = \frac{1}{2m^*}\Big(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}\Big)^2 + \tfrac{1}{2}\overline{g}\,\mu_{\rm B} H\,\sigma \quad , \tag{5.38}$$

where in the last line we've restricted our attention to the electron, for which $q = -e$. The g-factor for an electron is $\overline{g} = 2$ at tree level, and when radiative corrections are accounted for using quantum electrodynamics (QED) one finds $\overline{g} = 2.0023193043617(15)$. For our purposes we can take $\overline{g} = 2$, although we can always absorb the small difference into the definition of $\mu_{\rm B}$, writing $\mu_{\rm B} \to \tilde{\mu}_{\rm B} = \overline{g}e\hbar/4mc$. We've chosen the $\hat{\boldsymbol{z}}$-axis in spin space to point in the direction of the magnetic field, and we wrote the eigenvalues of S^z as $\frac{1}{2}\hbar\sigma$, where $\sigma = \pm 1$. The quantity m^* is the *effective mass* of the electron, here assumed to be isotropic in the vicinity of a band edge. An important distinction is that it is m^* which enters into the kinetic energy term $\boldsymbol{p}^2/2m^*$, but it is the electron mass m itself ($m = 511\,\text{keV}$) which enters into the definition of the Bohr magneton. We shall discuss the consequences of this further below.

In a crystalline semiconductor, the spin-orbit interaction,

$$V_{\rm so} = \frac{\hbar}{4m^2c^2}\,\boldsymbol{p}\cdot\boldsymbol{\sigma}\times\boldsymbol{\nabla}V \quad , \tag{5.39}$$

leads to an effective $\overline{g}$ which is often very far from the free electron value. For cubic systems with a direct band gap, the g-factor in band n is given by[4]

$$\frac{\overline{g}}{2} = 1 + \frac{2}{m}\,\mathsf{Im}\sum_{n'}{}'\frac{\langle\, n\,\Gamma\,|\,p_x\,|\,n'\,\Gamma\,\rangle\langle\, n'\,\Gamma\,|\,p_y\,|\,n\,\Gamma\,\rangle}{E_n(\Gamma) - E_{n'}(\Gamma)} + \ldots \quad , \tag{5.40}$$

where the wavefunctions and the energies are all taken at the zone center Γ. InSb, for example, has $\overline{g} \simeq -44$, while in GaAs $\overline{g} \simeq 0.4$.

In the absence of orbital magnetic coupling, the single particle dispersion is

$$\varepsilon_\sigma(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{k}^2}{2m^*} + \tilde{\mu}_{\rm B} H\,\sigma \quad . \tag{5.41}$$

At $T = 0$, we have the results of §5.2.3. At finite T, we once again use the Sommerfeld expan-

[3]We denote the g-factor by $\overline{g}$ in order to obviate confusion with the density of states function $g(\varepsilon)$.
[4]See, *e.g.*, ch. 14 of C. Kittel, *Quantum Theory of Solids*.

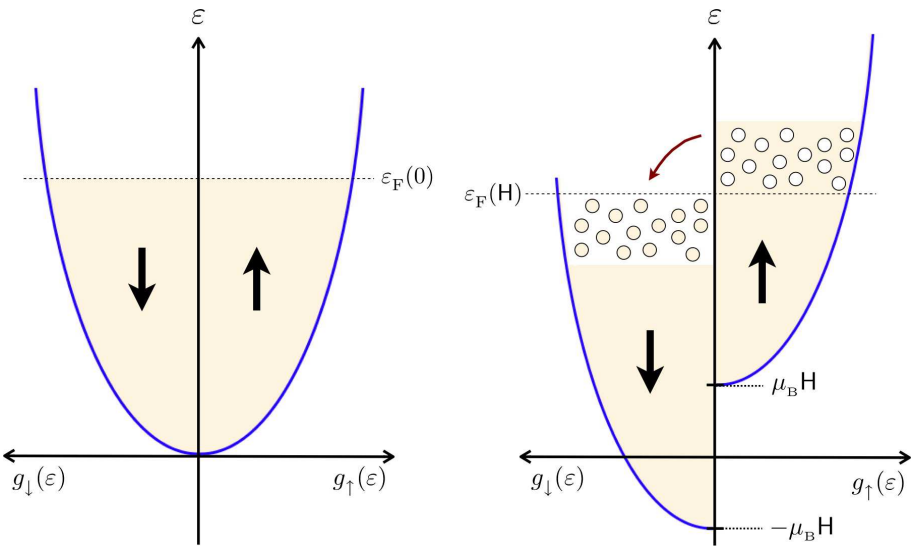


Figure 5.7: Fermi distributions in the presence of an external Zeeman-coupled magnetic field.

sion. We then have

$$\begin{aligned} n &= \int_{-\infty}^{\infty} d\varepsilon\, g_\uparrow(\varepsilon)\, f(\varepsilon-\mu) \;+\; \int_{-\infty}^{\infty} d\varepsilon\, g_\downarrow(\varepsilon)\, f(\varepsilon-\mu) \\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} d\varepsilon\, \Big\{ g(\varepsilon-\tilde\mu_{\rm B}H) + g(\varepsilon+\tilde\mu_{\rm B}H) \Big\}\, f(\varepsilon-\mu) \\ &= \int_{-\infty}^{\infty} d\varepsilon\, \Big\{ g(\varepsilon) + (\tilde\mu_{\rm B}H)^2\, g''(\varepsilon) + \ldots \Big\}\, f(\varepsilon-\mu) \quad . \end{aligned} \tag{5.42}$$

We now invoke the Sommerfeld expension to find the temperature dependence:

$$\begin{aligned} n &= \int_{-\infty}^{\mu} d\varepsilon\, g(\varepsilon) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\, g'(\mu) + (\tilde\mu_{\rm B}H)^2\, g'(\mu) + \ldots \\ &= \int_{-\infty}^{\varepsilon_{\rm F}} d\varepsilon\, g(\varepsilon) + g(\varepsilon_{\rm F})\,\delta\mu + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\, g'(\varepsilon_{\rm F}) + (\tilde\mu_{\rm B}H)^2\, g'(\varepsilon_{\rm F}) + \ldots \quad . \end{aligned} \tag{5.43}$$

Note that the density of states for spin species σ is

$$g_\sigma(\varepsilon) = \tfrac{1}{2}\, g(\varepsilon - \tilde\mu_{\rm B}H\sigma) \quad , \tag{5.44}$$

where $g(\varepsilon)$ is the total density of states per unit volume, for both spin species, in the absence of

a magnetic field. We conclude that the chemical potential shift in an external field is

$$\delta\mu(T,n,H) = -\left\{\frac{\pi^2}{6}\,(k_{\rm B}T)^2 + (\tilde\mu_{\rm B}H)^2\right\}\frac{g'(\varepsilon_{\rm F})}{g(\varepsilon_{\rm F})} + \ldots \quad . \tag{5.45}$$

We next compute the difference $n_\uparrow - n_\downarrow$ in the densities of up and down spin electrons:

$$\begin{aligned} n_\uparrow - n_\downarrow &= \int\limits_{-\infty}^{\infty} d\varepsilon\,\Big\{g_\uparrow(\varepsilon) - g_\downarrow(\varepsilon)\Big\}\,f(\varepsilon-\mu) \\ &= \tfrac{1}{2}\int\limits_{-\infty}^{\infty} d\varepsilon\,\Big\{g(\varepsilon-\tilde\mu_{\rm B}H) - g(\varepsilon+\tilde\mu_{\rm B}H)\Big\}\,f(\varepsilon-\mu) \\ &= -\tilde\mu_{\rm B}H\cdot\pi D\csc(\pi D)\,g(\mu) + \mathcal{O}(H^3) \quad . \end{aligned} \tag{5.46}$$

We needn't go beyond the trivial lowest order term in the Sommerfeld expansion, because H is already assumed to be small. Thus, the magnetization density is

$$M = -\tilde\mu_{\rm B}(n_\uparrow - n_\downarrow) = \tilde\mu_{\rm B}^2\,g(\varepsilon_{\rm F})\,H \quad . \tag{5.47}$$

in which the magnetic susceptibility is

$$\chi = \left(\frac{\partial M}{\partial H}\right)_{T,N} = \tilde\mu_{\rm B}^2\,g(\varepsilon_{\rm F}) \quad . \tag{5.48}$$

This is called the *Pauli paramagnetic susceptibility*.

5.4.2 Landau diamagnetism

When orbital effects are included, the single particle energy levels are given by

$$\varepsilon(n,k_z,\sigma) = (n+\tfrac{1}{2})\hbar\omega_{\rm c} + \frac{\hbar^2k_z^2}{2m^*} + \tilde\mu_{\rm B}H\,\sigma \quad . \tag{5.49}$$

Here n is a Landau level index, and $\omega_{\rm c} = eH/m^*c$ is the *cyclotron frequency*. Note that

$$\frac{\tilde\mu_{\rm B}H}{\hbar\omega_{\rm c}} = \frac{ge\hbar H}{4mc}\cdot\frac{m^*c}{\hbar eH} = \frac{\overline{g}}{4}\cdot\frac{m^*}{m} \quad . \tag{5.50}$$

Accordingly, we define the ratio $r \equiv (\overline{g}/2)\times(m^*/m)$. We can then write

$$\varepsilon(n,k_z,\sigma) = \big(n+\tfrac{1}{2}+\tfrac{1}{2}r\sigma\big)\hbar\omega_{\rm c} + \frac{\hbar^2k_z^2}{2m^*} \quad . \tag{5.51}$$

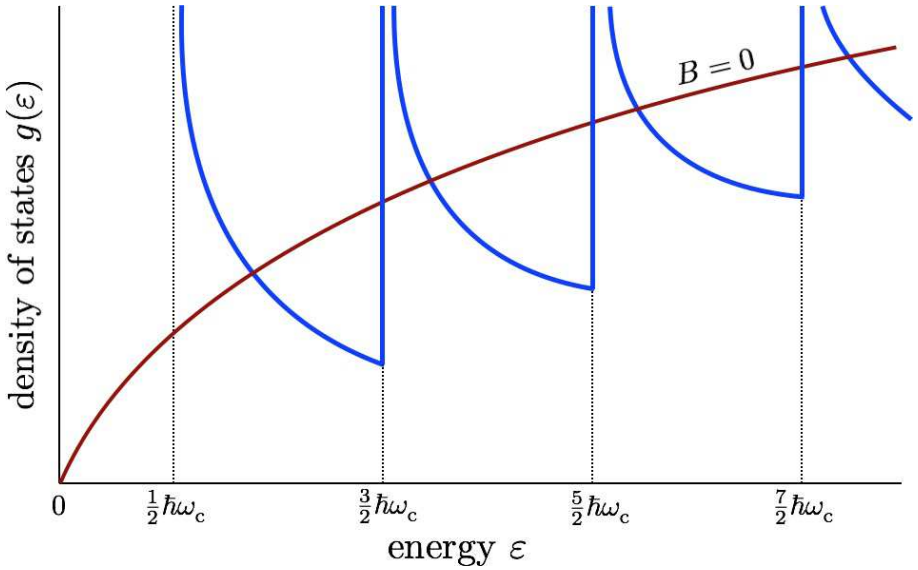


Figure 5.8: Density of states for a three-dimensional free electron gas with $\overline{g} = 0$ in the presence of an external magnetic field (blue), compared with $B = 0$ result (dark red).

The density of states per unit volume is then

$$g(\varepsilon) = \frac{1}{2\pi\ell^2} \sum_{n,\sigma} \int_{-\infty}^{\infty} \frac{dk_z}{2\pi}\, \delta\big(\varepsilon - \varepsilon(n, k_z, \sigma)\big) \quad , \tag{5.52}$$

where $\ell = (\hbar c/eH)^{1/2}$ is the *magnetic length*. The significance of ℓ is that the area per Dirac fluxoid $\phi_0 = hc/e$ is $2\pi\ell^2$.

The grand potential is then given by

$$\Omega = -\frac{HA}{\phi_0} \cdot L_z \cdot k_{\rm B}T \int_{-\infty}^{\infty} \frac{dk_z}{2\pi} \sum_{n=0}^{\infty} \sum_{\sigma=\pm 1} \ln\Big[1 + e^{\mu/k_{\rm B}T}\, e^{-(n+\frac{1}{2}+\frac{1}{2}r\sigma)\,\hbar\omega_{\rm c}/k_{\rm B}T}\, e^{-\hbar^2 k_z^2/2m^* k_{\rm B}T}\Big] \quad . \tag{5.53}$$

A few words are in order here regarding the prefactor. In the presence of a uniform magnetic field, the energy levels of a two-dimensional ballistic charged particle collapse into Landau levels. The number of states per Landau level scales with the area of the system, and is equal to the number of flux quanta through the system: $N_\phi = HA/\phi_0$, where $\phi_0 = hc/e$ is the Dirac flux quantum. Note that

$$\frac{HA}{\phi_0} \cdot L_z \cdot k_{\rm B}T = \hbar\omega_{\rm c} \cdot \frac{V}{\lambda_T^3} \quad , \tag{5.54}$$

hence we can write

$$\Omega(T, V, \mu, H) = \hbar\omega_{\rm c} \sum_{n=0}^{\infty} \sum_{\sigma=\pm 1} Q\big((n + \tfrac{1}{2} + \tfrac{1}{2}r\sigma)\,\hbar\omega_{\rm c} - \mu\big) \quad , \tag{5.55}$$

where we have defined the dimensionless function

$$Q(\varepsilon) = -\frac{V}{\lambda_T^2}\int\limits_{-\infty}^{\infty}\frac{dk_z}{2\pi}\,\ln\Big[1+e^{-\varepsilon/k_{\rm B}T}e^{-\hbar^2k_z^2/2m^*k_{\rm B}T}\Big] \quad . \tag{5.56}$$

We now invoke the Euler-MacLaurin formula,

$$\sum_{n=0}^{\infty}F(n) = \int\limits_0^{\infty}dx\,F(x) + \tfrac{1}{2}\,F(0) - \tfrac{1}{12}\,F'(0) + \ldots \quad , \tag{5.57}$$

resulting in

$$\begin{aligned}\Omega(T,V,\mu,H) = \sum_{\sigma=\pm1}\Bigg\{&\int\limits_{\frac{1}{2}(1+r\sigma)\hbar\omega_{\rm c}}^{\infty}d\varepsilon\,Q(\varepsilon-\mu) + \tfrac{1}{2}\,\hbar\omega_{\rm c}\,Q\big(\tfrac{1}{2}(1+r\sigma)\,\hbar\omega_{\rm c}-\mu\big)\\ &\qquad - \tfrac{1}{12}\,(\hbar\omega_{\rm c})^2\,Q'\big(\tfrac{1}{2}(1+r\sigma)\,\hbar\omega_{\rm c}-\mu\big) + \ldots\Bigg\} \quad .\end{aligned} \tag{5.58}$$

We next expand in powers of the magnetic field H to obtain

$$\Omega(T,V,\mu,H) = 2\int\limits_0^{\infty}d\varepsilon\,Q(\varepsilon-\mu) + \tfrac{1}{4}\big(r^2-\tfrac{1}{3}\big)\,(\hbar\omega_{\rm c})^2\,Q'(-\mu) + \ldots \quad . \tag{5.59}$$

Thus, the magnetic susceptibility is

$$\begin{aligned}\chi &= -\frac{1}{V}\frac{\partial^2\Omega}{\partial H^2} = \big(r^2-\tfrac{1}{3}\big)\cdot\tilde\mu_{\rm B}^2\cdot\big(m/m^*\big)^2\cdot\Big(-\frac{2}{V}\,Q'(-\mu)\Big)\\ &= \left(\frac{\overline{g}^2}{4}-\frac{m^2}{3m^{*2}}\right)\cdot\tilde\mu_{\rm B}^2\cdot n^2\kappa_T \quad ,\end{aligned} \tag{5.60}$$

where κ_T is the isothermal compressibility[5], which at $T=0$ is related to the density of states by $\kappa_T(T=0,n) = n^{-2}g(\varepsilon_{\rm F})$. In most metals we have $m^*\approx m$ and the term in brackets is positive (recall $\overline{g}\approx 2$). In semiconductors, however, we can have $m^*\ll m$; for example in GaAs we have $m^* = 0.067\,m$ and $\overline{g}=0.4$. Thus, semiconductors can have a *diamagnetic* response. If we take $\overline{g}=2$ and $m^*=m$, we see that the orbital currents give rise to a diamagnetic contribution to the magnetic susceptibility which is exactly $-\frac{1}{3}$ times as large as the contribution arising from Zeeman coupling. The net result is then paramagnetic ($\chi>0$) and $\frac{2}{3}$ as large as the Pauli susceptibility. The orbital currents can be understood within the context of *Lenz's law*.

[5]We've used $-\frac{2}{V}\,Q'(\mu) = -\frac{1}{V}\frac{\partial^2\Omega}{\partial\mu^2} = n^2\kappa_T$.

5.4.3 de Haas-van Alphen oscillations

The Landau level structure in the density of states (see Fig. 5.8) is responsible for striking behavior in metals when subjected to an external magnetic field. For weak fields, the magnetization density is $m = \chi H$, but at stronger fields we have

$$m(T,H,\mu) = -\frac{1}{V}\frac{\partial\Omega}{\partial H} = \left(\frac{e^2}{4\pi^2\hbar}\right)\left(\frac{2\mu}{m^*}\right)^{1/2} H\cdot\Bigg\{\left(r^2-\tfrac{1}{3}\right) + \\ + \left(\frac{2\pi k_{\rm B}T}{\hbar\omega_{\rm c}}\right)\left(\frac{2\mu}{\hbar\omega_{\rm c}}\right)^{1/2}\sum_{l=1}^{\infty}\frac{(-1)^l}{\sqrt{l}}\,\frac{\sin\left(\frac{2\pi l\mu}{\hbar\omega_{\rm c}}-\frac{\pi}{4}\right)\cos\left(l\pi r\right)}{\sinh\left(2\pi^2 l k_{\rm B}T/\hbar\omega_{\rm c}\right)}\Bigg\} \quad . \tag{5.61}$$

The electron number density is given by

$$n(T,H,\mu) = -\frac{1}{V}\frac{\partial\Omega}{\partial\mu} = \frac{1}{3\pi^2}\left(\frac{2m^*\mu}{\hbar^2}\right)^{1/2}\cdot\Bigg\{1+\frac{3}{32}\left(\frac{\hbar\omega_{\rm c}}{\mu}\right)\left(r^2-\tfrac{1}{3}\right) + \\ + \left(\frac{3\pi k_{\rm B}T}{\hbar\omega_{\rm c}}\right)\left(\frac{\hbar\omega_{\rm c}}{2\mu}\right)^{3/2}\sum_{l=1}^{\infty}\frac{(-1)^l}{\sqrt{l}}\,\frac{\sin\left(\frac{2\pi l\mu}{\hbar\omega_{\rm c}}-\frac{\pi}{4}\right)\cos\left(l\pi r\right)}{\sinh\left(2\pi^2 l k_{\rm B}T/\hbar\omega_{\rm c}\right)}\Bigg\} \quad . \tag{5.62}$$

These expressions are valid in the limit $\mu \gg \hbar\omega_{\rm c}$ and $\mu \gg k_{\rm B}T$. Under experimental conditions, it is the electron number density n which is held constant, and not the chemical potential μ. Thus, one must invert to obtain $\mu(n,T,H)$ and substitute this in the expression for $m(T,H,\mu)$ to obtain $m(n,T,H)$.

To derive the above results, we integrate Eqn. 5.14 twice by parts to obtain

$$\Omega = -V\int_{-\infty}^{\infty} d\varepsilon\, R(\varepsilon)\left(-\frac{\partial f}{\partial\varepsilon}\right) \quad , \tag{5.63}$$

where $R(\varepsilon)$ is given by

$$R(\varepsilon) = \int_{-\infty}^{\varepsilon} d\varepsilon' \int_{-\infty}^{\varepsilon'} d\varepsilon''\, g(\varepsilon'') \quad , \tag{5.64}$$

i.e. $g(\varepsilon) = R''(\varepsilon)$. In the presence of a uniform magnetic field, the density of states for a ballistic particle with dispersion $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m^*$ is

$$g(\varepsilon) = \frac{1}{2\pi\ell^2}\,\frac{\sqrt{m^*}}{\sqrt{2}\,\pi\hbar}\sum_{n=0}^{\infty}\sum_{\sigma=\pm1}\Big[\varepsilon - \left(n+\tfrac{1}{2}+\tfrac{1}{2}\sigma r\right)\hbar\omega_{\rm c}\Big]_+^{-1/2} \quad , \tag{5.65}$$

where $r = \overline{g}\, m^*/2m$ as before, and $[x]_+ \equiv x\,\Theta(x)$. Thus,

$$R(\varepsilon) = \frac{\sqrt{2}}{3\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2} \sum_{n=0}^{\infty}\sum_{\sigma=\pm1}\Big[\,\varepsilon - \big(n+\tfrac{1}{2}+\tfrac{1}{2}\sigma r\big)\hbar\omega_{\rm c}\,\Big]_+^{3/2} \quad , \tag{5.66}$$

We now invoke the result

$$\sum_{n=0}^{\infty}\phi(n+\tfrac{1}{2}) = \int_0^\infty\! du\;\phi(u) + \tfrac{1}{24}\phi'(0) - \sum_{n=1}^{\infty}\frac{(-1)^l}{2\pi^2 l^2}\int_0^\infty\! du\;\phi''(u)\cos(2\pi l u) \quad , \tag{5.67}$$

which is valid provided $\phi(\infty) = \phi'(\infty) = 0$. This follows from applying the Poisson summation formula[6],

$$\sum_{n=-\infty}^{\infty}\delta(x-n) = \sum_{l=-\infty}^{\infty} e^{2\pi i l x} \quad , \tag{5.68}$$

integrating by parts twice, and using $\sum_{l=1}^{\infty}(-1)^{l+1}/\,l^2 = \frac{\pi^2}{12}$.

$$\begin{aligned} R(\varepsilon) = \frac{2\sqrt{2}}{15\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\frac{\big[\,\varepsilon - \frac{1}{2}r\hbar\omega_{\rm c}\,\big]_+^{5/2}}{\hbar\omega_{\rm c}} - \frac{\sqrt{2}}{48\pi^2}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\hbar\omega_{\rm c}\,\big[\,\varepsilon - \tfrac{1}{2}r\hbar\omega_{\rm c}\,\big]_+^{3/2} + \\ - \frac{1}{8\pi^4}\left(\frac{m^*}{\hbar^2}\right)^{3/2}(\hbar\omega_{\rm c})^{3/2}\sum_{l=1}^{\infty}\frac{(-1)^l}{l^{5/2}}\,\cos\left(\frac{2\pi l\varepsilon}{\hbar\omega_{\rm c}} - \pi l r\sigma - \frac{\pi}{4}\right) \quad . \end{aligned} \tag{5.69}$$

We next integrate over ε, using the Sommerfeld expansion and the result

$$\int_{-\infty}^{\infty}\! d\varepsilon\; e^{is\varepsilon}\left(-\frac{\partial f}{\partial\varepsilon}\right) = \frac{\pi s\, k_{\rm B}T}{\sinh(\pi s\, k_{\rm B}T)} \quad . \tag{5.70}$$

The final result for $\Omega(T,V,\mu,H)$, valid for $\hbar\omega_{\rm c} \ll \mu$ and $k_{\rm B}T \ll \mu$, is

$$\begin{aligned} \Omega(T,V,\mu,H) = -V\cdot\frac{\sqrt{2}}{\pi}\left(\frac{m^*}{\hbar^2}\right)^{3/2}\Bigg\{ \tfrac{4}{15}\left\langle\varepsilon^{5/2}\right\rangle + \tfrac{1}{8}\,(\hbar\omega_{\rm c})^2\left(r^2-\tfrac{1}{3}\right)\left\langle\varepsilon^{1/2}\right\rangle + \\ + \tfrac{1}{2\sqrt{2}}\,(\hbar\omega_{\rm c})^{3/2}k_{\rm B}T\sum_{l=1}^{\infty}\frac{(-1)^l}{l^{3/2}}\,\frac{\cos\big(\frac{2\pi l\mu}{\hbar\omega_{\rm c}} - \frac{\pi}{4}\big)\cos\big(l\pi r\big)}{\sinh\big(2\pi^2 l k_{\rm B}T/\hbar\omega_{\rm c}\big)}\Bigg\} \quad . \end{aligned} \tag{5.71}$$

Here, we have used the notation

$$\left\langle\psi(\varepsilon)\right\rangle \equiv \int_{-\infty}^{\infty}\! d\varepsilon\;\psi(\varepsilon)\left(-\frac{\partial f}{\partial\varepsilon}\right) \quad . \tag{5.72}$$

[6]One first extends the function $\phi(u)$ to the entire real line, symmetrically, so $\phi(-u) = \phi(u)$.

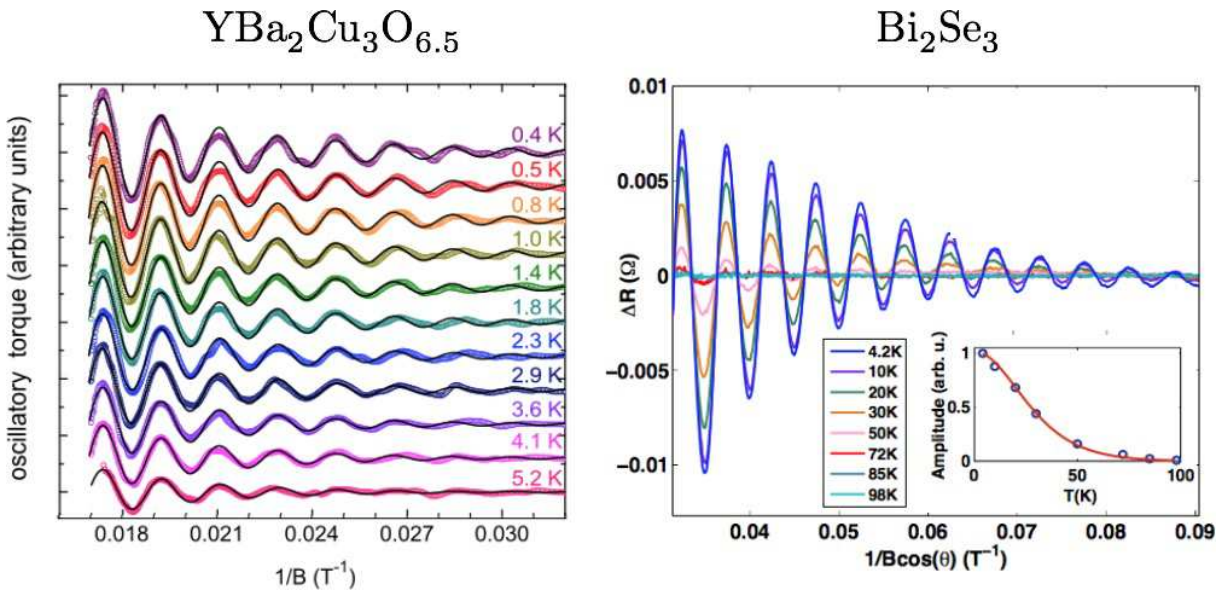


Figure 5.9: Left: de Haas-van Alphen oscillations in the underdoped high temperature superconductor $YBa_2Cu_3O_{6.5}$, measured by C. Jaudet *et al* in *Phys. Rev. Lett.* **100**, 187005 (2008). Right: Schubnikov-de Haas oscillations and their temperature dependence in the three-dimensional topological insulator Bi_2Se_3, indicating the presence of metallic surface states. A smooth polynomial background has been subtracted. Measurements by M. Petrushevsky *et al.* in *Phys. Rev. B* **86**, 045131 (2013).

For homogeneous functions,

$$\left\langle \varepsilon^p \right\rangle = \mu^p + \tfrac{\pi^2}{6}\, p(p-1)\,(k_{\rm B}T)^2\,\mu^{p-2} + \mathcal{O}(T^4) \quad . \tag{5.73}$$

Differentiation of $\Omega(T,V,H)$ with respect to H and μ yields[7], respectively, the results in Eqns. 5.61 and 5.62. From Eqn. 5.71, we see that the oscillating factors are periodic in $1/H$ with periods $\Delta(1/H) = \hbar e/2\pi l \mu m^* c$. In experiments, the magnetization M is typically measured with a torque-magnetometer. The oscillatory nature of $M(H)$ is called the de Haas-van Alphen effect. A related periodicity occurs in the magnetoresistance $R(H)$, where it is called the Schubnikov-de Haas effect. Experimental data for both effects is shown in Fig. 5.9.

Oscillations at $T = 0$: spinless fermions in $d = 2$ dimensions

Apparently the oscillations do not vanish, even at $T = 0$. The prefactor of T which multiplies the oscillating sum in Eqn. 5.61 cancels with the $\sinh(2\pi^2 l k_{\rm B}T/\hbar\omega_{\rm c})$ denominator in the $T \to 0$ limit. Consider the simple case of ballistic spinless electrons in $d = 2$ dimensions. We know

[7]The cyclotron energy $\hbar\omega_{\rm c} = \hbar eH/m^*c$ is linear in the magnetic field H. For $\mu \gg \hbar\omega_{\rm c} \gg k_{\rm B}T$, the dominant contribution to the magnetization comes from differentiating the cosine factor.

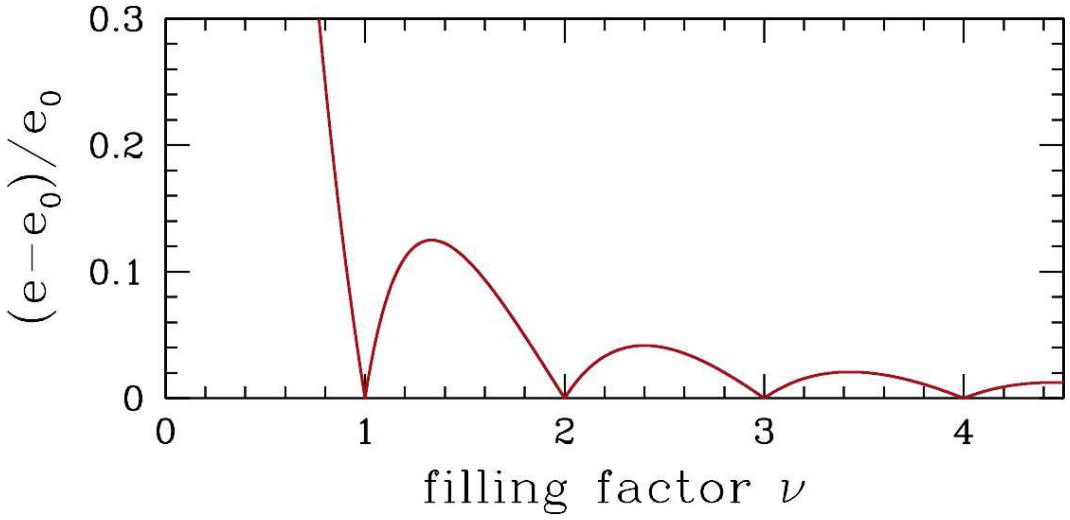


Figure 5.10: Two dimensional oscillations in the energy per unit area $e = E/A$ for a spinless electron gas in a uniform magnetic field $\boldsymbol{H} = H\hat{\boldsymbol{z}}$.

that each Landau level can accommodate $N_{\rm L} = HA/\phi_0$ electrons, where $\phi_0 = hc/e$ is the Dirac flux quantum. It is convenient to define the *filling fraction* ν, as

$$\nu = \frac{N}{N_{\rm L}} = 2\pi\ell^2 n = \frac{2\pi\hbar c}{eH}\,n \quad , \tag{5.74}$$

where $\ell = \sqrt{\hbar c/eH}$ is the magnetic length. The cyclotron energy is

$$\hbar\omega_{\rm c} = \frac{\hbar eH}{m^* c} = \frac{2\pi\hbar^2}{m^*}\,\frac{n}{\nu} \quad . \tag{5.75}$$

We now evaluate the energy per unit area, $\mathrm{e} = E/A$, as a function of n and ν. With the electron number density n fixed, the magnetization per unit area is given by

$$m = \frac{\partial \mathrm{e}}{\partial H} = \frac{\partial \nu}{\partial H}\,\frac{\partial \mathrm{e}}{\partial \nu} = -\frac{\nu}{H}\,\frac{\partial \mathrm{e}}{\partial \nu} \quad . \tag{5.76}$$

Now if $\nu \in [j, j+1]$, the total energy is

$$\begin{aligned} E &= N_{\rm L}\cdot\tfrac{1}{2}\hbar\omega_{\rm c}\cdot\Big(1+3+5+\ldots+(2j-1)\Big) + (N - jN_{\rm L})\cdot(j+\tfrac{1}{2})\,\hbar\omega_{\rm c} \\ &= N_{\rm L}\cdot\tfrac{1}{2}\hbar\omega_{\rm c}\cdot\Big(j^2 + (\nu-j)(2j+1)\Big) \quad . \end{aligned} \tag{5.77}$$

Thus,

$$\mathrm{e}(n,\nu) = \frac{E}{A} = \frac{\pi\hbar^2 n^2}{m^*}\cdot\left\{\frac{(2j+1)\,\nu - j(j+1)}{\nu^2}\right\} \quad . \tag{5.78}$$

Defining, $\mathrm{e}_0(n) = \pi\hbar^2 n^2/m^*$, we have that $\mathrm{e}(n,j) = \mathrm{e}(n,j+1) = \mathrm{e}_0(n)$. Furthermore, since

$$\left.\frac{\partial \mathrm{e}}{\partial \nu}\right|_{\nu=j} = \frac{\mathrm{e}_0(n)}{j^2} > 0 \qquad \text{and} \qquad \left.\frac{\partial \mathrm{e}}{\partial \nu}\right|_{\nu=j+1} = -\frac{\mathrm{e}_0(n)}{j^2} > 0 \quad , \tag{5.79}$$

we see that $\mathrm{e}(n,\nu)$ has a cusp at every integer value of ν. This behavior is depicted in Fig. 5.10. The magnetization density $m(n,\nu)$ therefore *discontinuously changes sign* (from negative to positive) across all integer values of the filling fraction![8] Note also that the periodicity is $\Delta\nu = 1$, hence

$$\Delta\Big(\frac{1}{H}\Big) = \frac{e}{\hbar c n} = \frac{1}{2\pi n \phi_0} \quad . \tag{5.80}$$

5.4.4 de Haas-von Alphen effect for anisotropic Fermi surfaces

We consider a nontopological band structure, for which the semiclassical equations of motion in the presence of a uniform magnetic field are

$$\frac{d\boldsymbol{r}}{dt} = \boldsymbol{v}_n(\boldsymbol{k}) \qquad , \qquad \frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar c}\,\boldsymbol{v}_n(\boldsymbol{k}) \times \boldsymbol{B} \quad . \tag{5.81}$$

These equations entail the conservation of the band energy $E_n(\boldsymbol{k})$:

$$\frac{dE_n(\boldsymbol{k})}{dt} = \frac{\partial E_n(\boldsymbol{k})}{\partial \boldsymbol{k}} \cdot \frac{d\boldsymbol{k}}{dt} = \hbar \boldsymbol{v}_n(\boldsymbol{k}) \cdot \left(-\frac{e}{\hbar c}\,\boldsymbol{v}_n(\boldsymbol{k}) \times \boldsymbol{B} \right) = 0 \quad . \tag{5.82}$$

Define $\boldsymbol{k}_\perp = \boldsymbol{k} - \widehat{\boldsymbol{B}}(\widehat{\boldsymbol{B}} \cdot \boldsymbol{k})$, the component of $\boldsymbol{k}$ along the direction $\widehat{\boldsymbol{B}}$. We then have $\dot{\boldsymbol{k}}_\perp = -\frac{e}{\hbar c}\,\boldsymbol{v}_n(\boldsymbol{k}) \times \boldsymbol{B}$ and $\frac{d}{dt}(\boldsymbol{k} \cdot \widehat{\boldsymbol{B}}) = 0$. Thus, the orbits $\boldsymbol{k}(t)$ lie in planes perpendicular to $\widehat{\boldsymbol{B}}$ (see the sketch in Fig. 5.11).

Consider now the differential $\boldsymbol{k}$-space area element $d^2k_\perp$ between transverse (to $\widehat{\boldsymbol{B}}$) slices of isoenergy surfaces at energies ε and $\varepsilon + d\varepsilon$. Clearly

$$d^2k_\perp = \frac{d\varepsilon\, d\ell(\varepsilon)}{\left|\partial\varepsilon/\partial \boldsymbol{k}_\perp\right|} \tag{5.83}$$

where $d\ell(\varepsilon)$ is the differential path length in the transverse plane along the surface of energy ε. Note that

$$\left|\frac{\partial\varepsilon}{\partial \boldsymbol{k}_\perp}\right| = \left|\hbar \boldsymbol{v}_\perp\right| = \hbar\left|\boldsymbol{v}_\perp \times \widehat{\boldsymbol{B}}\right| = \frac{\hbar^2 c}{eB}\left|\dot{\boldsymbol{k}}_\perp\right| \quad . \tag{5.84}$$

[8] In the three-dimensional case, m oscillates but usually does not change sign.

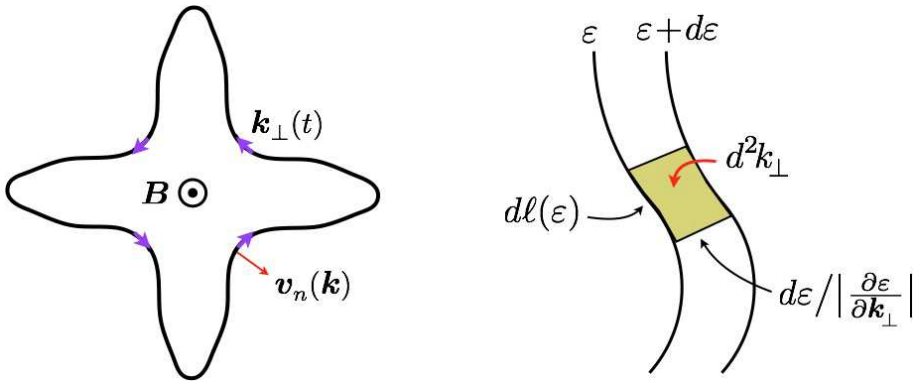


Figure 5.11: Left: Orbit of $\boldsymbol{k}(t)$ in the plane perpendicular to the field $\boldsymbol{B}$. Right: Geometry of Fermi surface orbits in the calculation of the de Haas-van Alphen effect.

Thus, the area enclosed by an orbit of energy ε and parallel wavevector component $k_\parallel = \boldsymbol{k}\cdot\widehat{\boldsymbol{B}}$ is

$$\begin{aligned} A(\varepsilon,k_\parallel) &= \int d^2k_\perp\,\Theta\big(\varepsilon-\varepsilon(\boldsymbol{k}_\perp,k_\parallel)\big) = \int\limits_{-\infty}^{\infty} d\varepsilon'\,\Theta(\varepsilon-\varepsilon')\oint\frac{d\ell(\varepsilon')}{\big|\partial\varepsilon'/\partial\boldsymbol{k}_\perp\big|} \\ &= \frac{eB}{\hbar^2c}\int\limits_{-\infty}^{\infty} d\varepsilon'\,\Theta(\varepsilon-\varepsilon')\oint\frac{d\ell(\varepsilon')}{\big|\dot{\boldsymbol{k}}_\perp\big|} = \frac{eB}{\hbar^2c}\int\limits_{-\infty}^{\varepsilon} d\varepsilon'\,T(\varepsilon',k_\parallel) \quad , \end{aligned} \tag{5.85}$$

where $T(\varepsilon,k_\parallel)$ is the period of the orbit. Note that we have assumed that the surface $\mathcal{S}_\varepsilon$ is closed, *i.e.* that there are no "open orbits" which run periodically across the Brillouin zone. We now have the result

$$\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon} = \frac{eB}{\hbar^2c}\,T(\varepsilon,k_\parallel) \quad . \tag{5.86}$$

For a free electron in a magnetic field, the orbital period is $2\pi/\omega_{\rm c}$. We accordingly define the *cyclotron mass* by the relation

$$T(\varepsilon,k_\parallel) = \frac{\hbar^2c}{eB}\,\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon} \equiv \frac{2\pi m_{\rm cyc}c}{eB} \qquad\Longrightarrow\qquad m_{\rm cyc} = \frac{\hbar^2}{2\pi}\,\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon} \quad . \tag{5.87}$$

Semiclassical quantization then yields the following relation for the energy level spacing:

$$\varepsilon_{n+1}(k_\parallel)-\varepsilon_n(k_\parallel) = \frac{2\pi\hbar}{T\big(\varepsilon_n(k_\parallel),k_\parallel\big)} = \frac{2\pi eB}{\hbar c}\Bigg/\left.\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_n(k_\parallel)} \quad . \tag{5.88}$$

Note that for free electrons,

$$A(\varepsilon,k_\parallel) = \pi\big(\boldsymbol{k}^2-k_\parallel^2\big) = \frac{2\pi m\varepsilon}{\hbar^2} - \pi k_\parallel^2 \quad , \tag{5.89}$$

and so

$$\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon}=\frac{2\pi m}{\hbar^2}\qquad\Longrightarrow\qquad\varepsilon_{n+1}(k_\parallel)-\varepsilon_n(k_\parallel)=\frac{\hbar eB}{mc}=\hbar\omega_{\rm c}\quad.\tag{5.90}$$

If the semiclassical orbit index n is large, we may approximate $\partial A/\partial\varepsilon$ by a ratio of differences, *viz.*

$$\left.\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_n(k_\parallel)}\simeq\frac{A\big(\varepsilon_{n+1}(k_\parallel),k_\parallel\big)-A\big(\varepsilon_n(k_\parallel),k_\parallel\big)}{\varepsilon_{n+1}(k_\parallel)-\varepsilon_n(k_\parallel)}\quad,\tag{5.91}$$

and invoking Eqn. 5.88 then gives

$$A\big(\varepsilon_{n+1}(k_\parallel),k_\parallel\big)-A\big(\varepsilon_n(k_\parallel),k_\parallel\big)=\frac{2\pi eB}{\hbar c}\quad.\tag{5.92}$$

We then conclude that the areas of the orbits in the plane transverse to $\widehat{\boldsymbol{B}}$ are quantized according to

$$A\big(\varepsilon_n(k_\parallel),k_\parallel\big)=(n+\alpha)\,\frac{2\pi eB}{\hbar c}\quad,\tag{5.93}$$

where α is a constant, a result first derived by Lars Onsager in 1952.

In the free particle model, the each dH-vA oscillation is associated with a Fermi level crossing by one of the Landau levels. Neglecting Zeeman splitting, the semiclassical density of states per unit volume is

$$\begin{aligned}g(\varepsilon)&=\frac{1}{2\pi\ell^2}\sum_n\int\limits_{-\infty}^{\infty}\frac{dk_\parallel}{2\pi}\,\delta\big(\varepsilon-\varepsilon_n(k_\parallel)\big)\\&=\frac{1}{2\pi\ell^2}\sum_n\int\limits_{-\infty}^{\infty}\frac{dk_\parallel}{2\pi}\,\frac{\delta\big(k_\parallel-\varepsilon_n^{-1}(\varepsilon)\big)}{\big|\partial\varepsilon_n/\partial k_\parallel\big|}\quad,\end{aligned}\tag{5.94}$$

where $\varepsilon_n^{-1}(\varepsilon)=k_\parallel$ when $\varepsilon_n(k_\parallel)=\varepsilon$, *i.e.* it is the inverse function. The DOS is peaked when the denominator vanishes, *i.e.* when $\partial\varepsilon_n/\partial k_\parallel=0$. This entails that the cross sectional Fermi surface area is at a maximum:

$$\frac{\partial}{\partial k_\parallel}A\big(\varepsilon_n(k_\parallel),k_\parallel\big)=\overbrace{\frac{\partial\varepsilon_n(k_\parallel)}{\partial k_\parallel}}^{=0}\cdot\left.\frac{\partial A(\varepsilon,k_\parallel)}{\partial\varepsilon}\right|_{\varepsilon=\varepsilon_n(k_\parallel)}+\left.\frac{\partial A(\varepsilon,k_\parallel)}{\partial k_\parallel}\right|_{\varepsilon=\varepsilon_n(k_\parallel)}\quad.\tag{5.95}$$

Thus, the DOS peaks when the Fermi energy lies on an *extremal orbit*, *i.e.* one which extremizes the cross-sectional Fermi surface area:

$$(n+\alpha)\,\frac{2\pi eB}{\hbar c}=S^*(\varepsilon_{\rm F})\qquad\Longrightarrow\qquad\Delta\left(\frac{1}{B}\right)=\frac{2\pi e}{\hbar c}\,\frac{1}{S^*(\varepsilon_{\rm F})}\quad,\tag{5.96}$$

where $S^*(\varepsilon)$ is a (possibly multi-valued) function giving the area(s) of the extremal orbits. Since

$$\frac{\hbar\omega_{\rm c}}{k_{\rm B}T} = 1.34 \times 10^{-4} \cdot \frac{B[{\rm T}]}{T[{\rm K}]} \quad , \tag{5.97}$$

(assuming $m = m_{\rm e}$), one needs high fields or low temperatures in order that the oscillations not be washed out by thermal fluctuations.

5.5 Simple Theory of Electron Transport in Metals

5.5.1 Drude model

Consider a particle of mass m^* and charge $(-e)$ moving in the presence of an electric field $\boldsymbol{E}$ and magnetic field $\boldsymbol{B}$. Newton's second law says that

$$\frac{d\boldsymbol{p}}{dt} = -e\boldsymbol{E} - \frac{e}{m^*c}\,\boldsymbol{p}\times\boldsymbol{B} - \frac{\boldsymbol{p}}{\tau} \quad , \tag{5.98}$$

where the last term on the RHS is a phenomenological dissipative (*i.e.* frictional) force. The constant τ, which has dimensions of time, is interpreted as the *momentum relaxation time* due to scattering off impurities, lattice excitations (*i.e.* phonons), or sample boundaries. Clearly when $\boldsymbol{E} = \boldsymbol{B} = 0$ we have $\boldsymbol{p}(t) = \boldsymbol{p}(0)\,\exp(-t/\tau)$, which says that $\boldsymbol{p}$ relaxes on a time scale τ.

When $\boldsymbol{E} \neq 0$ but $\boldsymbol{B} = 0$, we have $\dot{\boldsymbol{p}} = -e\boldsymbol{E} - \tau^{-1}\boldsymbol{p}$, and for time-independent $\boldsymbol{E}$ the steady state solution, valid at long times, is $\boldsymbol{p} = -e\tau\boldsymbol{E}$. is then

$$\boldsymbol{j} = -ne\boldsymbol{v} = -ne\,\frac{\boldsymbol{p}}{m^*} = \frac{ne^2\tau}{m^*}\,\boldsymbol{E} \quad . \tag{5.99}$$

Thus there is a linear relationship between the current density $\boldsymbol{j}$ and the applied field $\boldsymbol{E}$. One writes $\boldsymbol{j} = \sigma\boldsymbol{E}$, where σ is the *electrical conductivity*. The above theory says that $\sigma = ne^2\tau/m^*$, where n is the particle density.

We can extend our analysis to include time-dependent fields of the form $\boldsymbol{E}(t) = {\rm Re}\left[\,\widehat{\boldsymbol{E}}(\omega)\,e^{-i\omega t}\,\right]$. In steady state, $\boldsymbol{p}$ oscillates with the same frequency, and writing $\boldsymbol{p}(t) = {\rm Re}\left[\,\widehat{\boldsymbol{p}}(\omega)\,e^{-i\omega t}\,\right]$, we obtain the relation $(\tau^{-1} - i\omega)\,\widehat{\boldsymbol{p}}(\omega) = -e\widehat{\boldsymbol{E}}(\omega)$, and thus $\boldsymbol{j}(t) = {\rm Re}\left[\,\sigma(\omega)\,\widehat{\boldsymbol{E}}(\omega)\,e^{-i\omega t}\,\right]$, with

$$\sigma(\omega) = \frac{ne^2\tau}{m^*}\cdot\frac{1}{1-i\omega\tau} \quad . \tag{5.100}$$

The power density $\boldsymbol{j}(t)\cdot\boldsymbol{E}(t)$ then has terms which are constant, as well as terms oscillating with frequency 2ω. The average power dissipated is obtained by integrating over a period $\Delta t = 2\pi/\omega$, which eliminates the $e^{\pm 2i\omega t}$ terms, resulting in

$$\overline{\boldsymbol{j}(t)\cdot\boldsymbol{E}(t)} = {\rm Re}\,\sigma(\omega)\,\left|\widehat{\boldsymbol{E}}(\omega)\right|^2 \quad , \tag{5.101}$$

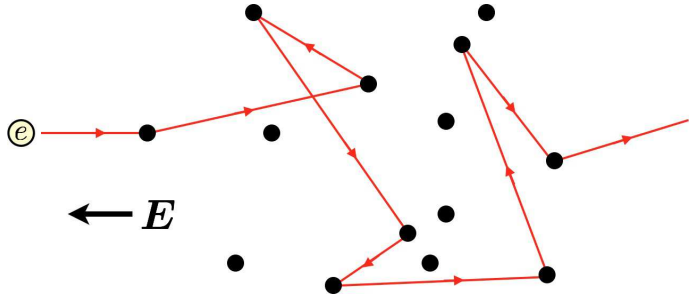


Figure 5.12: Scattering of an electron by impurities in the presence of a uniform electric field $\boldsymbol{E}$ is an example of a biased random walk.

where the bar denotes time averaging over the period $\Delta t = \pi/\omega$. So it is the real part of the conductivity which is responsible for power dissipation.

Another way to see it: write $\hat{\boldsymbol{j}}(\omega) = \sigma(\omega)\,\widehat{\boldsymbol{E}}(\omega)$, which is a complex vector quantity. If we separate the frequency-dependent conductivity $\sigma(\omega) = \sigma'(\omega) + i\sigma''(\omega)$ into its real and imaginary parts, we see that the $\sigma'(\omega)$ term leads to a current component which is in phase with the drive $\boldsymbol{E}(t)$, while the $\sigma''(\omega)$ term leads to a current component which is 90° out of phase with the drive $\boldsymbol{E}(t)$. The latter current leads to periodic fluctuations in the local energy density, but no net dissipation. The real and imaginary parts of $\sigma(\omega)$ are given by

$$\sigma'(\omega) = \frac{ne^2\tau}{m^*}\cdot\frac{1}{1+\omega^2\tau^2} \quad , \qquad \sigma''(\omega) = \frac{ne^2\tau}{m^*}\cdot\frac{\omega\tau}{1+\omega^2\tau^2} \quad . \tag{5.102}$$

When we try to apply the above physics to electrons in solids, we are confronted with several issues. One obvious question is: what do we mean by n? Filled bands carry no current, because the current density of the $n^{\rm th}$ filled band (accounting for spin),

$$j_n = -2e\int\limits_{\hat\Omega}\frac{d^3k}{(2\pi)^3}\,\overbrace{\frac{1}{\hbar}\frac{\partial E_n(\boldsymbol{k})}{\partial k^\alpha}}^{v_n(\boldsymbol{k})} = 0 \quad , \tag{5.103}$$

vanishes because $E_n(\boldsymbol{k})$ is periodic in the Brillouin zone, and the integral of the derivative of a periodic function over its period is zero. So the density n must only include contributions from partially filled bands. In fact, the situation is even more complicated because the scattering time can vary from band to band, may be energy-dependent, and there can even be *interband scattering* of electrons. Another question is how we account for scattering within the semiclassical model. We can't just add a term $-\boldsymbol{p}/\tau$ to the right hand side of the equation for $\hbar\dot{\boldsymbol{k}}$, because $\boldsymbol{p} = \hbar\boldsymbol{k}$ is not well-defined in a crystal. A more rigorous approach to transport is based on the Boltzmann equation, which describes how the distribution $f(\boldsymbol{r},\boldsymbol{k},t)$ of electron wave packets evolves and takes a steady state form.

The DC conductivity $\sigma = ne^2\tau/m^*$ is proportional to the number of carriers n. Another figure of merit is the *mobility* $\mu = e\tau/m^*$, which is independent of n. Note that the mobility is the ratio of the speed of the electron to the magnitude of the applied field: $\mu = v/E$. The conventional units of mobility are $[\mu] = \mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$. Mobility tells us mostly about the scattering time τ. For highly disordered systems, the scattering time τ is short and consequently μ is small. The electrons then move slowly as they execute a biased random walk scattering off impurities in the presence of an electric field (see Fig. 5.12. However, even low mobility systems may have high conductivity, owing to a large density n of conduction electrons. The highest purity semiconductors have mobilities on the order of $10^7\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$.

5.5.2 Magnetoresistance and magnetoconductance

Now let's introduce a uniform magnetic field $\boldsymbol{B}$. In component notation, Newton's second law gives

$$\left(\frac{1}{\tau}\,\delta_{\alpha\beta} + \frac{e}{m^*c}\,\varepsilon_{\alpha\beta\gamma}\,B^\gamma\right) p^\beta = -eE^\alpha \quad . \tag{5.104}$$

The current density is $\boldsymbol{j} = -ne\boldsymbol{v} = -ne\boldsymbol{p}/m^*$, hence $\boldsymbol{p} = -m^*\boldsymbol{j}/ne$, and we thus have

$$E^\alpha = \overbrace{\frac{1}{ne^2}\left(\frac{m^*}{\tau}\,\delta_{\alpha\beta} + \frac{e}{c}\,\varepsilon_{\alpha\beta\gamma}\,B^\gamma\right)}^{\rho_{\alpha\beta}} j^\beta \quad . \tag{5.105}$$

The *resistivity matrix* $\rho_{\alpha\beta}(B)$ defines the linear relationship between the electric field $\boldsymbol{E}$ and the current density $\boldsymbol{j}$. At finite frequency, it is easy to see that τ^{-1} must be replaced by $\tau^{-1} - i\omega$, hence, taking $\boldsymbol{B} = B\hat{\boldsymbol{z}}$, the $T = 0$ resistivity tensor is

$$\rho_{\alpha\beta}(\omega, B) = \frac{m^*}{ne^2\tau}\begin{pmatrix} 1 - i\omega\tau & \omega_{\rm c}\tau & 0 \\ -\omega_{\rm c}\tau & 1 - i\omega\tau & 0 \\ 0 & 0 & 1 - i\omega\tau \end{pmatrix} \quad , \tag{5.106}$$

with $\omega_{\rm c} = eB/m^*c$ the cyclotron frequency, as before. Note that the diagonal elements are independent of B, which says that the *magnetoresistance*

$$\Delta\rho_{xx}(B) = \rho_{xx}(B) - \rho_{xx}(0) \tag{5.107}$$

vanishes: $\Delta\rho_{xx}(B) = 0$.

The magnetoconductance, however, does not vanish! Recall that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad , \tag{5.108}$$

from which we have

$$\sigma_{\alpha\beta} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & 0 \\ \sigma_{yx} & \sigma_{yy} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix} \quad , \tag{5.109}$$

with

$$\begin{aligned} \sigma_{xx}(\omega, B) = \sigma_{yy}(\omega, B) &= \frac{ne^2\tau}{m^*} \cdot \frac{1 - i\omega\tau}{(1 - i\omega\tau)^2 + (\omega_{\rm c}\tau)^2} \\ \sigma_{yx}(\omega, B) = -\sigma_{xy}(\omega, B) &= \frac{ne^2\tau}{m^*} \cdot \frac{\omega_{\rm c}\tau}{(1 - i\omega\tau)^2 + (\omega_{\rm c}\tau)^2} \\ \sigma_{zz}(\omega, B) &= \frac{ne^2\tau}{m^*} \cdot \frac{1}{1 - i\omega\tau} \quad . \end{aligned} \tag{5.110}$$

Note that σ_{xx} is field-dependent, unlike ρ_{xx}.

Thus far we have assumed that the effective mass tensor $m^*_{\alpha\beta}$ is isotropic. In the general anisotropic case, $m^*_{\alpha\beta}$, which is a symmetric matrix, will have three orthogonal principal axes, which we denote as $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{z}}$. In this case, the resistivity tensor assumes the more general form

$$\rho_{\alpha\beta}(\omega, B) = \frac{1}{ne^2} \begin{pmatrix} (\tau^{-1} - i\omega)\, m_x^* & \pm eB_z/c & \mp eB_y/c \\ \mp eB_z/c & (\tau^{-1} - i\omega)\, m_y^* & \pm eB_x/c \\ \pm eB_y/c & \mp eB_x/c & (\tau^{-1} - i\omega)\, m_z^* \end{pmatrix} \quad , \tag{5.111}$$

where (m_x^*, m_y^*, m_z^*) are the three eigenvalues of $m^*_{\alpha\beta}$. The $\pm$ sign in the off-diagonal term distinguishes the case where the Fermi level is just above a quadratic minimum ($+$ sign), versus where it is just below quadratic maximum ($-$ sign). The latter case is described in terms of *holes* in a filled band, as opposed to *electrons* in an empty band. The effective mass tensors are then defined as

$$\big(m^*\big)^{-1}_{\alpha\beta} = \pm \frac{1}{\hbar^2} \frac{\partial^2 E_n(\boldsymbol{k})}{\partial k^\alpha \, \partial k^\beta} \quad , \tag{5.112}$$

where the top sign corresponds to electrons and the bottom sign to holes.

Note that the diagonal elements in Eqn. 5.111 are still independent of B and there is no magnetoresistance. Taking $\boldsymbol{B}$ along $\hat{\boldsymbol{z}}$, the corresponding elements of $\sigma_{\alpha\beta}$ are

$$\begin{aligned} \sigma_{xx}(\omega, B) &= \frac{ne^2\tau}{m_x^*} \cdot \frac{1 - i\omega\tau}{(1 - i\omega\tau)^2 + (\omega_{\rm c}\tau)^2} \\ \sigma_{yy}(\omega, B) &= \frac{ne^2\tau}{m_y^*} \cdot \frac{1 - i\omega\tau}{(1 - i\omega\tau)^2 + (\omega_{\rm c}\tau)^2} \\ \sigma_{yx}(\omega, B) &= \pm \frac{ne^2\tau}{m_\perp^*} \cdot \frac{\omega_{\rm c}\tau}{(1 - i\omega\tau)^2 + (\omega_{\rm c}\tau)^2} \\ \sigma_{zz}(\omega, B) &= \frac{ne^2\tau}{m_z^*} \cdot \frac{1}{1 - i\omega\tau} \quad , \end{aligned} \tag{5.113}$$

where $\omega_{\rm c} = eB/m_\perp^* c$ and $m_\perp^* = \sqrt{m_x^* m_y^*}$.

5.5.3 Hall effect in high fields

In the high field limit, we have that the resistivity and conductivity tensors are purely off-diagonal, with

$$\rho_{xy}(B) = \pm\frac{B}{nec} \quad , \quad \sigma_{xy}(B) = \mp\frac{nec}{B} \tag{5.114}$$

where the upper sign is again for conduction electrons, and the bottom sign for valence holes. Thus, the high field Hall effect may be used to determine the carrier concentration:

$$n = \pm \lim_{B\to\infty} \frac{B}{ec\rho_{xy}(B)} \quad . \tag{5.115}$$

5.5.4 Cyclotron resonance in semiconductors

A typical value for the effective mass in semiconductors is $m^* \sim 0.1\, m_{\rm e}$. From

$$\frac{e}{m_{\rm e}\, c} = 1.75 \times 10^7\,{\rm Hz/G} \quad , \tag{5.116}$$

we find that $eB/m^*c = 1.75 \times 10^{11}$ Hz in a field of $B = 1$ kG. In metals, the disorder is such that even at low temperatures $\omega_{\rm c}\tau$ typically is small. In semiconductors, however, the smallness of m^* and the relatively high purity (sometimes spectacularly so) mean that $\omega_{\rm c}\tau$ can get as large as 10^3 at modest fields. This allows for a measurement of the effective mass tensor using the technique of *cyclotron resonance*.

The absorption of electromagnetic radiation is proportional to the dissipative (*i.e.* real) part of the diagonal elements of $\sigma_{\alpha\beta}(\omega, B)$, which, again taking $\boldsymbol{B}$ along $\hat{\boldsymbol{z}}$, is given by

$$\sigma'_{xx}(\omega, B) = \frac{ne^2\tau}{m^*_x} \frac{1 + (\lambda^2+1)s^2}{1 + 2(\lambda^2+1)s^2 + (\lambda^2-1)^2 s^4} \quad , \tag{5.117}$$

where $\lambda = B/B_\omega$, with $B_\omega = m^*_\perp c\omega/e$, and $s = \omega\tau$. For fixed ω, the conductivity $\sigma'_{xx}(B)$ is then peaked at $B = B^*$. When $\omega\tau \gg 1$ and $\omega_{\rm c}\tau \gg 1$, B^* approaches B_ω, where $\sigma'_{xx}(\omega, B_\omega) = ne^2\tau/2m^*_x$. By measuring B_ω one can extract the quantity $m^*_\perp = eB_\omega/\omega c$. Varying the direction of the magnetic field, the entire effective mass tensor may be determined.

For finite $\omega\tau$, we can differentiate the above expression to obtain the location of the cyclotron resonance peak. One finds $B = (1+\alpha)^{1/2}\, B_\omega$, with

$$\begin{aligned}\alpha &= \frac{-(2s^2+1) + \sqrt{(2s^2+1)^2 - 1}}{s^2} \\ &= -\frac{1}{4s^4} + \frac{1}{8s^6} + \mathcal{O}(s^{-8}) \quad .\end{aligned}$$

As depicted in Fig. 5.13, the resonance peak shifts to the left of B_ω for finite values of $\omega\tau$. The peak collapses to $B = 0$ when $\omega\tau \leqslant 1/\sqrt{3} = 0.577$.

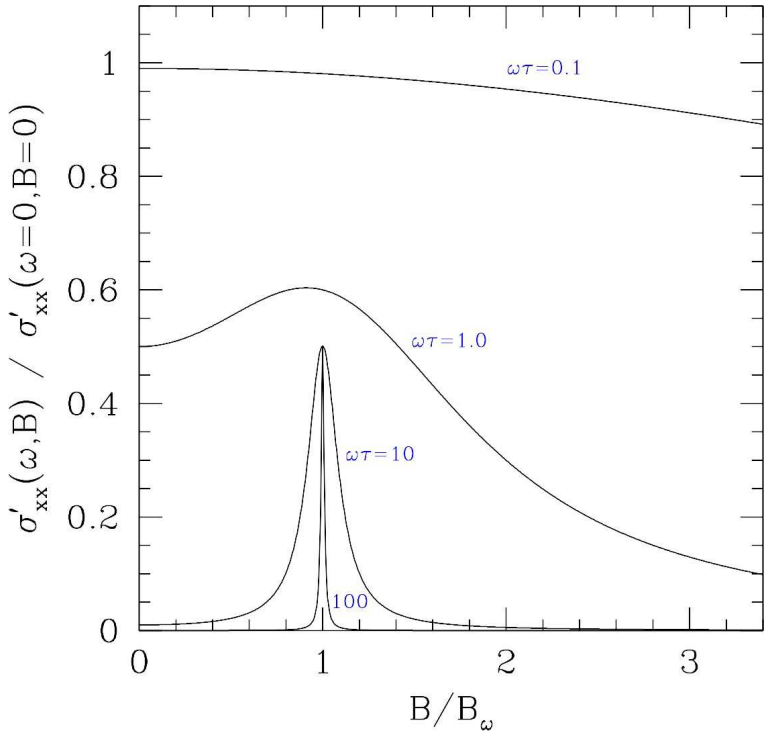


Figure 5.13: Theoretical cyclotron resonance peaks as a function of B/B_ω for different values of $\omega\tau$.

5.5.5 Magnetoresistance in a two band model

For a semiconductor with both electrons and holes present – a situation not uncommon to metals either (*e.g.* Aluminum) – each band contributes to the conductivity. The individual band conductivities are *additive* because the electron and hole conduction processes occur *in parallel*, *i.e.*

$$\sigma_{\alpha\beta}(\omega) = \sum_n \sigma^{(n)}_{\alpha\beta}(\omega) \quad , \tag{5.118}$$

where $\sigma^{(n)}_{\alpha\beta}$ is the conductivity tensor for band n, which may be computed in either the electron or hole picture (whichever is more convenient). We assume here that the two bands c and v may be treated independently, *i.e.* there is no interband scattering to account for.

The resistivity tensor of each band, $\rho^{(n)}_{\alpha\beta}$ exhibits no magnetoresistance, as we have found. However, if two bands are present, the total resistivity tensor ρ is obtained from $\rho^{-1} = \rho_{\mathrm{c}}^{-1} + \rho_{\mathrm{v}}^{-1}$, and

$$\rho = \left(\rho_{\mathrm{c}}^{-1} + \rho_{\mathrm{v}}^{-1}\right)^{-1} \tag{5.119}$$

will in general exhibit the phenomenon of magnetoresistance.

Explicitly, then, let us consider a model with isotropic and nondegenerate conduction band

minimum and valence band maximum. Taking $\boldsymbol{B} = B\hat{\boldsymbol{z}}$, we have

$$\rho_{\rm c} = \frac{(1-i\omega\tau_{\rm c})m_{\rm c}}{n_{\rm c}e^2\tau_{\rm c}}\,\mathbb{I} + \frac{B}{n_{\rm c}ec}\begin{pmatrix}0&1&0\\-1&0&0\\0&0&0\end{pmatrix} = \begin{pmatrix}\alpha_{\rm c}&\beta_{\rm c}&0\\-\beta_{\rm c}&\alpha_{\rm c}&0\\0&0&\alpha_{\rm c}\end{pmatrix}$$
$$\rho_{\rm v} = \frac{(1-i\omega\tau_{\rm v})m_{\rm v}}{n_{\rm v}e^2\tau_{\rm v}}\,\mathbb{I} - \frac{B}{n_{\rm v}ec}\begin{pmatrix}0&1&0\\-1&0&0\\0&0&0\end{pmatrix} = \begin{pmatrix}\alpha_{\rm v}&-\beta_{\rm v}&0\\\beta_{\rm v}&\alpha_{\rm v}&0\\0&0&\alpha_{\rm v}\end{pmatrix} \quad , \tag{5.120}$$

where

$$\alpha_{\rm c} = \frac{(1-i\omega\tau_{\rm c})m_{\rm c}}{n_{\rm c}e^2\tau_{\rm c}} \qquad \beta_{\rm c} = \frac{B}{n_{\rm c}ec}$$
$$\alpha_{\rm v} = \frac{(1-i\omega\tau_{\rm v})m_{\rm v}}{n_{\rm v}e^2\tau_{\rm v}} \qquad \beta_{\rm v} = \frac{B}{n_{\rm v}ec} \quad , \tag{5.121}$$

we obtain for the upper left 2×2 block of ρ:

$$\rho_\perp = \left[\left(\frac{\alpha_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}+\frac{\alpha_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2}\right)^2+\left(\frac{\beta_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}+\frac{\beta_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2}\right)^2\right]^{-1} \times\begin{pmatrix}\frac{\alpha_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}+\frac{\alpha_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2} & \frac{\beta_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}+\frac{\beta_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2}\\ -\frac{\beta_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}-\frac{\beta_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2} & \frac{\alpha_{\rm v}}{\alpha_{\rm v}^2+\beta_{\rm v}^2}+\frac{\alpha_{\rm c}}{\alpha_{\rm c}^2+\beta_{\rm c}^2}\end{pmatrix} \quad , \tag{5.122}$$

from which we compute the magnetoresistance,

$$\frac{\rho_{xx}(B)-\rho_{xx}(0)}{\rho_{xx}(0)} = \frac{\gamma_{\rm c}\,\gamma_{\rm v}\left(\frac{\gamma_{\rm c}}{n_{\rm c}ec}-\frac{\gamma_{\rm v}}{n_{\rm v}ec}\right)^2 B^2}{(\gamma_{\rm c}+\gamma_{\rm v})^2+(\gamma_{\rm c}\,\gamma_{\rm v})^2\left(\frac{1}{n_{\rm c}ec}+\frac{1}{n_{\rm v}ec}\right)^2 B^2} \quad , \tag{5.123}$$

where

$$\gamma_{\rm c} \equiv \alpha_{\rm c}^{-1} = \frac{n_{\rm c}e^2\tau_{\rm c}}{m_{\rm c}}\cdot\frac{1}{1-i\omega\tau_{\rm c}}$$

$$\gamma_{\rm v} \equiv \alpha_{\rm v}^{-1} = \frac{n_{\rm v}e^2\tau_{\rm v}}{m_{\rm v}}\cdot\frac{1}{1-i\omega\tau_{\rm v}} \quad .$$

Note that the magnetoresistance is *positive* within the two band model, and that it *saturates* in the high field limit:

$$\frac{\rho_{xx}(B\to\infty)-\rho_{xx}(0)}{\rho_{xx}(0)} = \frac{\gamma_{\rm c}\,\gamma_{\rm v}\left(\frac{\gamma_{\rm c}}{n_{\rm c}ec}-\frac{\gamma_{\rm v}}{n_{\rm v}ec}\right)^2}{(\gamma_{\rm c}\,\gamma_{\rm v})^2\left(\frac{1}{n_{\rm c}ec}+\frac{1}{n_{\rm v}ec}\right)^2} \quad . \tag{5.124}$$

The longitudinal resistivity is found to be

$$\rho_{zz} = (\gamma_{\rm c} + \gamma_{\rm v})^{-1} \quad , \tag{5.125}$$

and is independent of B.

In an intrinsic semiconductor, $n_{\rm c} = n_{\rm v} \propto \exp(-E_{\rm g}/2k_{\rm B}T)$, and $\Delta\rho_{xx}(B)/\rho_{xx}(0)$ is finite even as $T \to 0$. In the extrinsic (*i.e.* doped) case, one of the densities (say, $n_{\rm c}$ in a p-type material) vanishes much more rapidly than the other, and the magnetoresistance vanishes with the ratio $n_{\rm c}/n_{\rm v}$.

5.5.6 Optical reflectivity of metals and semiconductors

What happens when an electromagnetic wave is incident on a metal? Inside the metal we have Maxwell's equations:

$$\boldsymbol{\nabla} \times \boldsymbol{H} = \frac{4\pi}{c}\,\boldsymbol{j} + \frac{1}{c}\frac{\partial \boldsymbol{D}}{\partial t} \qquad \Longrightarrow \qquad i\boldsymbol{k} \times \boldsymbol{B} = \left(\frac{4\pi\sigma}{c} - \frac{i\omega}{c}\right)\boldsymbol{E} \tag{5.126}$$

and

$$\boldsymbol{\nabla} \times \boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t} \qquad \Longrightarrow \qquad i\boldsymbol{k} \times \boldsymbol{E} = \frac{i\omega}{c}\boldsymbol{B} \tag{5.127}$$

and

$$\boldsymbol{\nabla} \cdot \boldsymbol{E} = \boldsymbol{\nabla} \cdot \boldsymbol{B} = 0 \qquad \Longrightarrow \qquad i\boldsymbol{k} \cdot \boldsymbol{E} = i\boldsymbol{k} \cdot \boldsymbol{B} = 0 \quad , \tag{5.128}$$

where we've assumed $\mu = \epsilon = 1$ inside the metal, ignoring polarization due to virtual interband transitions (*i.e.* from core electrons). Hence,

$$\begin{aligned} \boldsymbol{k}^2 &= \frac{\omega^2}{c^2} + \frac{4\pi i\omega}{c^2}\,\sigma(\omega) \\ &= \frac{\omega^2}{c^2} + \frac{\omega_{\rm p}^2}{c^2}\,\frac{i\omega\tau}{1 - i\omega\tau} \equiv \epsilon(\omega)\,\frac{\omega^2}{c^2} \quad , \end{aligned} \tag{5.129}$$

where $\omega_{\rm p} = \sqrt{4\pi n e^2/m^*}$ is the *plasma frequency* for the conduction band. The dielectric function,

$$\epsilon(\omega) = 1 + \frac{4\pi i\sigma(\omega)}{\omega} = 1 + \frac{\omega_{\rm p}^2}{\omega^2}\,\frac{i\omega\tau}{1 - i\omega\tau} \tag{5.130}$$

determines the complex refractive index, $N(\omega) = \sqrt{\epsilon(\omega)}$, leading to the electromagnetic dispersion relation $k = N(\omega)\,\omega/c$.

Consider a wave normally incident upon a metallic surface normal to $\hat{\boldsymbol{z}}$. In the vacuum ($z < 0$), we write

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r}, t) &= E_1\,\hat{\boldsymbol{x}}\,e^{i\omega z/c}e^{-i\omega t} + E_2\,\hat{\boldsymbol{x}}\,e^{-i\omega z/c}e^{-i\omega t} \\ \boldsymbol{B}(\boldsymbol{r}, t) &= \frac{c}{i\omega}\,\boldsymbol{\nabla} \times \boldsymbol{E} = E_1\,\hat{\boldsymbol{y}}\,e^{i\omega z/c}e^{-i\omega t} - E_2\,\hat{\boldsymbol{y}}\,e^{-i\omega z/c}e^{-i\omega t} \end{aligned} \tag{5.131}$$

while in the metal ($z > 0$),

$$\begin{aligned} \boldsymbol{E}(\boldsymbol{r},t) &= E_3\,\hat{\boldsymbol{x}}\,e^{iN\omega z/c}e^{-i\omega t} \\ \boldsymbol{B}(\boldsymbol{r},t) &= \frac{c}{i\omega}\,\boldsymbol{\nabla}\times\boldsymbol{E} = NE_3\,\hat{\boldsymbol{y}}\,e^{iN\omega z/c}e^{-i\omega t} \quad . \end{aligned} \tag{5.132}$$

Continuity of $\boldsymbol{E}\times\hat{\boldsymbol{n}}$ gives $E_1 + E_2 = E_3$. Continuity of $\boldsymbol{H}\times\hat{\boldsymbol{n}}$ gives $E_1 - E_2 = NE_3$. Thus,

$$\frac{E_2}{E_1} = \frac{1-N}{1+N} \qquad , \qquad \frac{E_3}{E_1} = \frac{2}{1+N} \tag{5.133}$$

and the reflection and transmission coefficients are

$$\begin{aligned} R(\omega) &= \left|\frac{E_2}{E_1}\right|^2 = \left|\frac{1-N(\omega)}{1+N(\omega)}\right|^2 \\ T(\omega) &= \left|\frac{E_3}{E_1}\right|^2 = \frac{4}{\left|1+N(\omega)\right|^2} \quad . \end{aligned} \tag{5.134}$$

We've now solved the electromagnetic boundary value problem.

Typical values – For a metal with $n = 10^{22}\,\mathrm{cm}^3$ and $m^* = m_{\rm e}$, the plasma frequency is $\omega_{\rm p} = 5.7\times10^{15}\,\mathrm{s}^{-1}$. The scattering time varies considerably as a function of temperature. In high purity copper at $T = 4\,\mathrm{K}$, $\tau \approx 2\times10^{-9}\,\mathrm{s}$ and $\omega_{\rm p}\tau \approx 10^7$. At $T = 300\,\mathrm{K}$, $\tau \approx 2\times10^{-14}\,\mathrm{s}$ and $\omega_{\rm p}\tau \approx 100$. In either case, $\omega_{\rm p}\tau \gg 1$. There are then three regimes to consider:

Low frequencies : $\omega\tau \ll 1 \ll \omega_{\rm p}\tau$

We may approximate $1 - i\omega\tau \approx 1$, hence

$$\begin{aligned} N^2(\omega) &= 1 + \frac{i\,\omega_{\rm p}^2\tau}{\omega(1-i\omega\tau)} \approx \frac{i\,\omega_{\rm p}^2\tau}{\omega} \\ N(\omega) &\approx \frac{1+i}{\sqrt{2}}\left(\frac{\omega_{\rm p}^2\tau}{\omega}\right)^{1/2} \quad\Longrightarrow\quad R \approx 1 - \frac{2\sqrt{2\omega\tau}}{\omega_{\rm p}\tau} \quad . \end{aligned} \tag{5.135}$$

Hence $R \approx 1$ and the metal reflects.

Intermediate frequencies : $1 \ll \omega\tau \ll \omega_{\rm p}\tau$

In this regime,

$$N^2(\omega) \approx 1 - \frac{\omega_{\rm p}^2}{\omega^2} + \frac{i\,\omega_{\rm p}^2}{\omega^3\tau} \tag{5.136}$$

which is almost purely real and negative. Hence N is almost purely imaginary and $R \approx 1$. (To lowest nontrivial order, $R = 1 - 2/\omega_{\rm p}\tau$.) Still high reflectivity.

High frequencies : $1 \ll \omega_{\rm p}\tau \ll \omega\tau$

Here we have

$$N^2(\omega) \approx 1 - \frac{\omega_{\rm p}^2}{\omega^2} \quad \Longrightarrow \quad R = \frac{\omega_{\rm p}}{2\omega} \tag{5.137}$$

and $R \ll 1$ – the metal is transparent at frequencies large compared to $\omega_{\rm p}$.

Optical reflectivity of semiconductors

In our analysis of the electrodynamics of metals, we assumed that the dielectric constant due to all the filled bands was simply $\epsilon = 1$. This is not quite right. We should instead have written

$$\begin{aligned} \boldsymbol{k}^2 &= \epsilon_\infty \, \frac{\omega^2}{c^2} + \frac{4\pi i\omega\sigma(\omega)}{c^2} \\ \epsilon(\omega) &= \epsilon_\infty \left\{ 1 + \frac{\omega_{\rm p}^2}{\omega^2} \frac{i\omega\tau}{1 - i\omega\tau} \right\} \quad , \end{aligned} \tag{5.138}$$

where ϵ_∞ is the dielectric constant due to virtual transitions to fully occupied (*i.e.* core) and fully unoccupied bands, at a frequency small compared to the interband frequency. The plasma frequency is now defined as

$$\omega_{\rm p} = \left(\frac{4\pi n e^2}{m^* \, \epsilon_\infty} \right)^{1/2} \tag{5.139}$$

where n is the conduction electron density. Note that $\epsilon(\omega \to \infty) = \epsilon_\infty$, although again this is only true for ω smaller than the gap to neighboring bands. It turns out that for insulators one can write

$$\epsilon_\infty \simeq 1 + \frac{\omega_{\rm pv}^2}{\omega_{\rm g}^2} \tag{5.140}$$

where $\omega_{\rm pv} = \sqrt{4\pi n_{\rm v} e^2/m_{\rm e}}$, with $n_{\rm v}$ the number density of valence electrons, and $\omega_{\rm g}$ is the energy gap between valence and conduction bands. In semiconductors such as Si and Ge, $\omega_{\rm g} \sim 4\,$eV, while $\omega_{\rm pv} \sim 16\,$eV, hence $\epsilon_\infty \sim 17$, which is in rough agreement with the experimental values of ~ 12 for Si and ~ 16 for Ge. In metals, the band gaps generally are considerably larger.

There are some important differences to consider in comparing semiconductors and metals:

- The carrier density n typically is much smaller in semiconductors than in metals, ranging from $n \sim 10^{16}\,{\rm cm}^{-3}$ in intrinsic (*i.e.* undoped, thermally excited at room temperature) materials to $n \sim 10^{19}\,{\rm cm}^{-3}$ in doped materials.

- $\epsilon_\infty \approx 10 - 20$ and $m^*/m_{\rm e} \approx 0.1$. The product $\epsilon_\infty m^*$ thus differs only slightly from its free electron value.

Since $n_{\rm semi} \lesssim 10^{-4}\, n_{\rm metal}$, one has

$$\omega_{\rm p}^{\rm semi} \approx 10^{-2}\, \omega_{\rm p}^{\rm metal} \approx 10^{-14}\,{\rm s} \quad . \tag{5.141}$$

In high purity semiconductors the mobility $\mu = e\tau/m^* \gtrsim 10^5\,{\rm cm}^2/{\rm vs}$ the low temperature scattering time is typically $\tau \approx 10^{-11}\,{\rm s}$. Thus, for $\omega \gtrsim 3 \times 10^{15}\,{\rm s}^{-1}$ in the optical range, we have $\omega\tau \gg \omega_{\rm p}\tau \gg 1$, in which case $N(\omega) \approx \sqrt{\epsilon_\infty}$ and the reflectivity is

$$R = \left| \frac{1 - \sqrt{\epsilon_\infty}}{1 + \sqrt{\epsilon_\infty}} \right|^2 \quad . \tag{5.142}$$

Taking $\epsilon_\infty = 10$, one obtains $R = 0.27$, which is high enough so that polished Si wafers appear shiny.

5.5.7 Theory for Bloch wavepackets

But then how do we implement the semiclassical equations of motion for Bloch wavepackets,

$$\begin{aligned} \frac{d\boldsymbol{r}}{dt} &= \boldsymbol{v}_n(\boldsymbol{k}) - \frac{d\boldsymbol{k}}{dt} \times \boldsymbol{\Omega}_n(\boldsymbol{k}) \\ \hbar\,\frac{d\boldsymbol{k}}{dt} &= -e\boldsymbol{E}(\boldsymbol{r},t) - \frac{e}{c}\,\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B}(\boldsymbol{r},t) \quad ? \end{aligned} \tag{5.143}$$

In particular, how do we account for scattering within the semiclassical model? In what follows, we shall assume that the topological density $\boldsymbol{\Omega}_n(\boldsymbol{k}) = 0$.

A more rigorous approach to this issue is based on the Boltzmann equation, which describes how the distribution $f(\boldsymbol{r},\boldsymbol{k},t)$ of electron wave packets evolves and takes a steady state form. Here we will opt for a more callow treatment which yields equivalent results. The most naïve generalization of the semiclassical equations would involve adding the 'scattering' term $-\boldsymbol{p}/\tau$ to the right hand side of the equation for $\dot{\boldsymbol{p}} = \hbar\dot{\boldsymbol{k}}$, *i.e.*

$$M(\boldsymbol{k})\,\frac{d\boldsymbol{v}(\boldsymbol{k})}{dt} = -e\boldsymbol{E}(\boldsymbol{r},t) - \frac{e}{c}\,\frac{d\boldsymbol{r}}{dt} \times \boldsymbol{B}(\boldsymbol{r},t) - \frac{1}{\tau}\,M(\boldsymbol{k})\,\boldsymbol{v}(\boldsymbol{k}) \quad , \tag{5.144}$$

where

$$M_{\alpha\beta}(\boldsymbol{k}) = \hbar\,\frac{\partial k^\alpha}{\partial v^\beta} \qquad \Longleftrightarrow \qquad M^{-1}_{\alpha\beta}(\boldsymbol{k}) = \frac{1}{\hbar^2}\,\frac{\partial^2 E(\boldsymbol{k})}{\partial k^\alpha\,\partial k^\beta} \quad . \tag{5.145}$$

However, while $\dot{\boldsymbol{k}}$ is well-defined, $\boldsymbol{k}$ itself, and hence $\boldsymbol{p} = \hbar\boldsymbol{k}$, is not, because it is defined only modulo a reciprocal lattice vector.

5.6 Boltzmann Equation in Solids

5.6.1 Semiclassical dynamics and distribution functions

The semiclassical dynamics of a wavepacket in a solid are described by the equations[9]

$$\frac{d\boldsymbol{r}}{dt} = \frac{1}{\hbar}\frac{\partial\varepsilon_n(\boldsymbol{k})}{\partial\boldsymbol{k}} - \frac{d\boldsymbol{k}}{dt}\times\boldsymbol{\Omega}_n(\boldsymbol{k}) \tag{5.146}$$

$$\frac{d\boldsymbol{k}}{dt} = -\frac{e}{\hbar}\boldsymbol{E}(\boldsymbol{r},t) - \frac{e}{\hbar c}\frac{d\boldsymbol{r}}{dt}\times\boldsymbol{B}(\boldsymbol{r},t) - \frac{e}{2\hbar mc}\boldsymbol{\nabla}(\boldsymbol{\sigma}\cdot\boldsymbol{B}) \quad . \quad . \tag{5.147}$$

Here n is the band index and $\varepsilon_n(\boldsymbol{k})$ is the dispersion relation for band n. The Zeeman contribution to the Hamiltonian is $H_{\mathsf{Z}} = (e\hbar/2mc)\boldsymbol{\sigma}\cdot\boldsymbol{B}$, and we will typically choose the internal $\hat{\boldsymbol{z}}$ axis as the spin quantization axis, in which case $H_{\mathsf{Z}} = (e\hbar/2mc)\sigma B^z$. The wavevector is $\boldsymbol{k}$ ($\hbar\boldsymbol{k}$ is the 'crystal momentum'), and $\varepsilon_n(\boldsymbol{k})$ is periodic under $\boldsymbol{k}\to\boldsymbol{k}+\boldsymbol{G}$, where $\boldsymbol{G}$ is any reciprocal lattice vector. The second term on the RHS of Eqn. 5.146 is the so-called Karplus-Luttinger term, defined by

$$\mathcal{A}_n^\mu(\boldsymbol{k}) = i\left\langle\, u_n(\boldsymbol{k}) \,\middle|\, \frac{\partial}{\partial k^\mu} \,\middle|\, u_n(\boldsymbol{k})\,\right\rangle \tag{5.148}$$

$$\Omega_n^\mu(\boldsymbol{k}) = \epsilon^{\mu\nu\lambda}\,\frac{\partial\mathcal{A}_n^\lambda(\boldsymbol{k})}{\partial k^\nu} \quad , \tag{5.149}$$

arising from the Berry phases generated by the one-particle Bloch cell functions $|u_n(\boldsymbol{k})\rangle$. These formulae are valid only at sufficiently weak fields. They neglect, for example, Zener tunneling processes in which an electron may change its band index as it traverses the Brillouin zone. We assume $\boldsymbol{\Omega}_n(\boldsymbol{k}) = 0$ in our discussion, *i.e.* we assume the Bloch bands are non topological. Finally, we neglect the orbital magnetization of the Bloch wavepacket and contributions from the spin-orbit interaction. When the orbital moment of the Bloch electrons is included, we must substitute

$$\varepsilon_n(\boldsymbol{k}) \to \varepsilon_n(\boldsymbol{k}) - \boldsymbol{M}_n(\boldsymbol{k})\cdot\boldsymbol{B}(\boldsymbol{r},t) \tag{5.150}$$

where

$$M_n^\mu(\boldsymbol{k}) = e\,\epsilon^{\mu\nu\lambda}\,\mathsf{Im}\left\langle\frac{\partial u_n}{\partial k^\nu}\,\middle|\,\varepsilon_n(\boldsymbol{k}) - H_0(\boldsymbol{k})\,\middle|\,\frac{\partial u_n}{\partial k^\lambda}\right\rangle \quad , \tag{5.151}$$

where $\hat{H}_0(\boldsymbol{k}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,\hat{H}_0\,e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}$ and $\hat{H}_0 = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r})$ is the one-electron Hamiltonian in the crystalline potential $V(\boldsymbol{r}) = V(\boldsymbol{r}+\boldsymbol{R})$, where $\boldsymbol{R}$ is any direct lattice vector. Note $\hat{H}_0(\boldsymbol{k})\,|u_n(\boldsymbol{k})\rangle = \varepsilon_n(\boldsymbol{k})\,|u_n(\boldsymbol{k})\rangle$ and that $u_n(\boldsymbol{k},\boldsymbol{r}+\boldsymbol{R}) = u_n(\boldsymbol{k},\boldsymbol{r})$ is periodic in the direct lattice.

[9] See G. Sundaram and Q. Niu, *Phys. Rev. B* **59**, 14915 (1999).

We are of course interested in more than just a single electron, hence to that end let us consider the distribution function $f_n(\boldsymbol{r}, \boldsymbol{k}, t)$, defined such that[10]

$$f_{n\sigma}(\boldsymbol{r}, \boldsymbol{k}, t)\,\frac{d^3r\,d^3k}{(2\pi)^3} \equiv \begin{array}{c}\text{\# of electrons of spin } \sigma \text{ in band } n \text{ with positions within} \\ d^3r \text{ of } \boldsymbol{r} \text{ and wavevectors within } d^3k \text{ of } \boldsymbol{k} \text{ at time } t.\end{array} \tag{5.152}$$

Note that the distribution function is dimensionless. By performing integrals over the distribution function, we can obtain various physical quantities. For example, the current density at $\boldsymbol{r}$ is given by

$$\boldsymbol{j}(\boldsymbol{r}, t) = -e \sum_{n,\sigma} \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, f_{n\sigma}(\boldsymbol{r}, \boldsymbol{k}, t)\, \boldsymbol{v}_n(\boldsymbol{k}) \quad . \tag{5.153}$$

The symbol $\hat{\Omega}$ in the above formula is to remind us that the wavevector integral is performed only over the first Brillouin zone.

We now ask how the distribution functions $f_{n\sigma}(\boldsymbol{r}, \boldsymbol{k}, t)$ evolve in time. To simplify matters, we will consider a single band and drop the indices n and σ. It is clear that in the absence of collisions, the distribution function must satisfy the continuity equation,

$$\frac{\partial f}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{u} f) = 0 \quad . \tag{5.154}$$

This is just the condition of number conservation for electrons. Take care to note that $\boldsymbol{\nabla}$ and $\boldsymbol{u}$ are *six*-dimensional *phase space* vectors:

$$\begin{aligned} \boldsymbol{u} &= (\ \dot{x}\ ,\ \dot{y}\ ,\ \dot{z}\ ,\ \dot{k}_x\ ,\ \dot{k}_y\ ,\ \dot{k}_z\ ) \\ \boldsymbol{\nabla} &= \left(\frac{\partial}{\partial x}\ , \frac{\partial}{\partial y}\ , \frac{\partial}{\partial z}\ , \frac{\partial}{\partial k_x}\ , \frac{\partial}{\partial k_y}\ , \frac{\partial}{\partial k_z} \right) \quad . \end{aligned} \tag{5.155}$$

Now note that as a consequence of the dynamics (5.146,5.147) that, provided $\boldsymbol{\Omega}_n(\boldsymbol{k}) = 0$, we have $\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0$, *i.e.* phase space flow is *incompressible*, provided that $\varepsilon(\boldsymbol{k})$ is a function of $\boldsymbol{k}$ alone, and not of $\boldsymbol{r}$. Thus, in the absence of collisions, we have

$$\frac{\partial f}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla} f = 0 \quad . \tag{5.156}$$

The differential operator $D_t \equiv \partial_t + \boldsymbol{u} \cdot \boldsymbol{\nabla}$ is sometimes called the 'convective derivative'.

When $\boldsymbol{\Omega}_n(\boldsymbol{k}) \neq 0$, we found in §4.5.2 that $\boldsymbol{\nabla} \cdot \boldsymbol{u} = -d \ln D_n / dt$, with

$$D_n(\boldsymbol{r}, \boldsymbol{k}, t) = 1 + \frac{e}{\hbar c}\, \boldsymbol{B}(\boldsymbol{r}, t) \cdot \boldsymbol{\Omega}(\boldsymbol{k}) \quad . \tag{5.157}$$

[10] We will assume three space dimensions. The discussion may be generalized to quasi-two dimensional and quasi-one dimensional systems as well.

In this case, we must redefine the phase space measure as

$$d\mu = \frac{d^3r\,d^3k}{(2\pi)^3} \quad \longrightarrow \quad d\tilde{\mu} \equiv D_n(\boldsymbol{r},\boldsymbol{k},t)\,\frac{d^3r\,d^3k}{(2\pi)^3} \quad . \tag{5.158}$$

This means that the expectation of any local observable $\mathcal{O}$ is given by

$$\begin{aligned}\langle\mathcal{O}\rangle(\boldsymbol{r}',t) &= \sum_n \int\limits_{\hat{\Omega}} d\tilde{\mu}\, f_n(\boldsymbol{r},\boldsymbol{k},t)\,\langle\, u_{n\boldsymbol{k}}\,|\,\mathcal{O}\,|\,u_{n\boldsymbol{k}}\,\rangle\,\delta(\boldsymbol{r}-\boldsymbol{r}') \\ &= \sum_n \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\,D_n(\boldsymbol{r},'\boldsymbol{k},t)\,f_n(\boldsymbol{r}',\boldsymbol{k},t)\,\langle\, u_{n\boldsymbol{k}}\,|\,\mathcal{O}\,|\,u_{n\boldsymbol{k}}\,\rangle \quad ,\end{aligned} \tag{5.159}$$

Thus, for example

$$\begin{aligned}\boldsymbol{j}(\boldsymbol{r},t) &= \sum_n \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\;D_n(\boldsymbol{r},\boldsymbol{k},t)\,f_n(\boldsymbol{r},\boldsymbol{k},t)\,(-e\dot{\boldsymbol{r}}) \\ &= -e\int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\left\{\boldsymbol{v}_n + \frac{e}{\hbar c}\,(\boldsymbol{v}_n\cdot\boldsymbol{\Omega}_n)\,\boldsymbol{B} + \frac{e}{\hbar}\,\boldsymbol{E}\times\boldsymbol{\Omega}_n\right\} f_n(\boldsymbol{r},\boldsymbol{k},t) \quad .\end{aligned} \tag{5.160}$$

Throughout the rest of this chapter, we will assume $\boldsymbol{\Omega}_n(\boldsymbol{k}) = 0$. Here we have absorbed the spin polarization index σ into the band index, so there are twice as many n values as before. This notation is more appropriate when spin-orbit interaction terms are present, which can lead to cell functions $|u_{n\boldsymbol{k}}\rangle$ which have internal spin space structure.

Next we must consider the effect of *collisions*, which are not accounted for by the semiclassical dynamics. In a collision process, an electron with wavevector $\boldsymbol{k}$ and one with wavevector $\boldsymbol{k}'$ can instantaneously convert into a pair with wavevectors $\boldsymbol{k}+\boldsymbol{q}$ and $\boldsymbol{k}'-\boldsymbol{q}$ (modulo a reciprocal lattice vector $\boldsymbol{G}$), where $\boldsymbol{q}$ is the wavevector transfer. Note that the total wavevector is preserved (mod $\boldsymbol{G}$). This means that $D_t f \neq 0$. Rather, we should write

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{r}}\cdot\frac{\partial f}{\partial \boldsymbol{r}} + \dot{\boldsymbol{k}}\cdot\frac{\partial f}{\partial \boldsymbol{k}} = \left(\frac{\partial f}{\partial t}\right)_{\rm coll} \equiv \mathcal{I}_{\boldsymbol{k}}[f] \tag{5.161}$$

where the right side is known as the *collision integral*. The collision integral is in general a *function* of $\boldsymbol{r}$, $\boldsymbol{k}$, and $\boldsymbol{t}$ and a *functional* of the distribution f. As the $\boldsymbol{k}$-dependence is the most important for our concerns, we will write $\mathcal{I}_{\boldsymbol{k}}$ in order to make this dependence explicit. Some examples should help clarify the situation.

First, let's consider a very simple model of the collision integral,

$$\mathcal{I}_{\boldsymbol{k}}[f] = -\frac{f(\boldsymbol{r},\boldsymbol{k},t) - f^0(\boldsymbol{r},\boldsymbol{k})}{\tau(\varepsilon(\boldsymbol{k}))} \quad . \tag{5.162}$$

This model is known as the *relaxation time approximation*. Here, $f^0(\boldsymbol{r},\boldsymbol{k})$ is a static distribution function which describes a *local equilibrium* at $\boldsymbol{r}$. The quantity $\tau(\varepsilon(\boldsymbol{k}))$ is the *relaxation time*, which may be energy-dependent. Note that the collision integral indeed depends on the variables $(\boldsymbol{r},\boldsymbol{k},t)$, and has a particularly simple functional dependence on the distribution f.

A more sophisticated model might invoke Fermi's golden rule, Consider elastic scattering from a static potential $\mathcal{U}(\boldsymbol{r})$ which induces transitions between different momentum states. We can then write

$$\begin{aligned}\mathcal{I}_{\boldsymbol{k}}[f] &= \frac{2\pi}{\hbar}\sum_{\boldsymbol{k}'\in\hat{\Omega}} |\langle\,\boldsymbol{k}'\,|\,\mathcal{U}\,|\,\boldsymbol{k}\,\rangle|^2\,(f_{\boldsymbol{k}'}-f_{\boldsymbol{k}})\,\delta(\varepsilon_{\boldsymbol{k}}-\varepsilon_{\boldsymbol{k}'}) \\ &= \frac{2\pi}{\hbar V}\int\limits_{\hat{\Omega}}\frac{d^3k'}{(2\pi)^3}\,|\hat{\mathcal{U}}(\boldsymbol{k}-\boldsymbol{k}')|^2\,(f_{\boldsymbol{k}'}-f_{\boldsymbol{k}})\,\delta(\varepsilon_{\boldsymbol{k}}-\varepsilon_{\boldsymbol{k}'}) \quad ,\end{aligned} \tag{5.163}$$

where we abbreviate $f_{\boldsymbol{k}}\equiv f(\boldsymbol{r},\boldsymbol{k},t)$. In deriving the last line we've used plane wave wavefunctions[11] $\psi_{\boldsymbol{k}}(\boldsymbol{r})=\exp(i\boldsymbol{k}\cdot\boldsymbol{r})/\sqrt{V}$, as well as the result

$$\sum_{\boldsymbol{k}\in\hat{\Omega}} A(\boldsymbol{k}) = V\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,A(\boldsymbol{k}) \tag{5.164}$$

for smooth functions $A(\boldsymbol{k})$. Note the factor of V^{-1} in front of the integral in Eqn. 5.163. What this tells us is that for a bounded localized potential $\mathcal{U}(\boldsymbol{r})$, the contribution to the collision integral is inversely proportional to the size of the system. This makes sense because the number of electrons scales as V but the potential is only appreciable over a region of volume $\propto V^0$. Later on, we shall consider a finite density of scatterers, writing $\mathcal{U}(\boldsymbol{r})=\sum_{i=1}^{N_{\rm imp}}U(\boldsymbol{r}-\boldsymbol{R}_i)$, where the impurity density $n_{\rm imp}=N_{\rm imp}/V$ is finite, scaling as V^0. In this case $\hat{\mathcal{U}}(\boldsymbol{k}-\boldsymbol{k}')$ apparently scales as V, which would mean $\mathcal{I}_{\boldsymbol{k}}\{f\}$ scales as V, which is unphysical. As we shall see, the random positioning of the impurities means that the $\mathcal{O}(V^2)$ contribution to $|\hat{\mathcal{U}}(\boldsymbol{k}-\boldsymbol{k}')|^2$ is *incoherent* and averages out to zero. The coherent piece scales as V, canceling the V in the denominator of Eqn. 5.163, resulting in a finite value for the collision integral in the thermodynamic limit (*i.e.* neither infinite nor infinitesimal).

Later on we will discuss electron-phonon scattering, which is *inelastic*. An electron with wavevector $\boldsymbol{k}'$ can scatter into a state with wavevector $\boldsymbol{k}=\boldsymbol{k}'-\boldsymbol{q}\bmod\boldsymbol{G}$ by absorption of a phonon of wavevector $\boldsymbol{q}$ or emission of a phonon of wavevector $-\boldsymbol{q}$. Similarly, an electron of wavevector $\boldsymbol{k}$ can scatter into the state $\boldsymbol{k}'$ by emission of a phonon of wavevector $-\boldsymbol{q}$ or absorption of a phonon of wavevector $\boldsymbol{q}$. The matrix element for these processes depends on $\boldsymbol{k}$, $\boldsymbol{k}'$, and the polarization index of the phonon. Overall, energy is conserved. These considerations lead us

[11] Rather than plane waves, we should use Bloch waves $\psi_{n\boldsymbol{k}}(\boldsymbol{r})=\exp(i\boldsymbol{k}\cdot\boldsymbol{r})\,u_{n\boldsymbol{k}}(\boldsymbol{r})$, where cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ satisfies $u_{n\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R})=u_{n\boldsymbol{k}}(\boldsymbol{r})$, where $\boldsymbol{R}$ is any direct lattice vector. Plane waves do not contain the cell functions, although they do exhibit Bloch periodicity $\psi_{n\boldsymbol{k}}(\boldsymbol{r}+\boldsymbol{R})=\exp(i\boldsymbol{k}\cdot\boldsymbol{R})\,\psi_{n\boldsymbol{k}}(\boldsymbol{r})$.

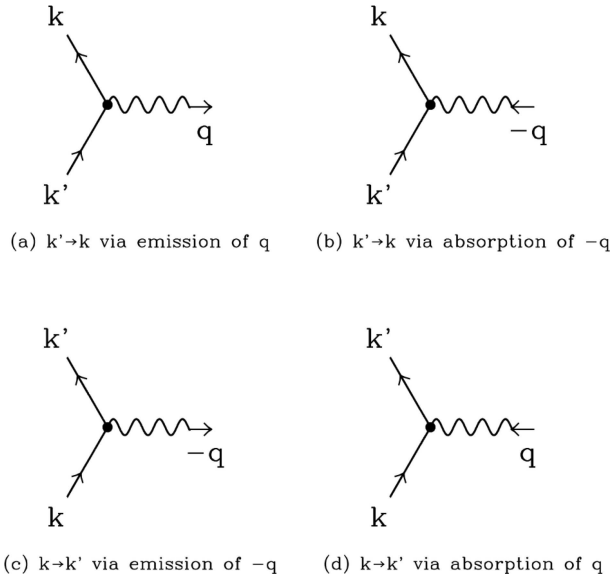


Figure 5.14: Electron-phonon vertices.

to the following collision integral:

$$\mathcal{I}_{\boldsymbol{k}}[f,n] = \frac{2\pi}{\hbar V}\sum_{\boldsymbol{k}',\lambda} |g_\lambda(\boldsymbol{k},\boldsymbol{k}')|^2 \Big\{ (1-f_{\boldsymbol{k}})\, f_{\boldsymbol{k}'}\, (1+n_{\boldsymbol{q},\lambda})\, \delta(\varepsilon_{\boldsymbol{k}} + \hbar\omega_{\boldsymbol{q}\lambda} - \varepsilon_{\boldsymbol{k}'}) \tag{5.165}$$
$$+ (1-f_{\boldsymbol{k}})\, f_{\boldsymbol{k}'}\, n_{-\boldsymbol{q}\lambda}\, \delta(\varepsilon_{\boldsymbol{k}} - \hbar\omega_{-\boldsymbol{q}\lambda} - \varepsilon_{\boldsymbol{k}'}) - f_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}'})\,(1+n_{-\boldsymbol{q}\lambda})\,\delta(\varepsilon_{\boldsymbol{k}} - \hbar\omega_{-\boldsymbol{q}\lambda} - \varepsilon_{\boldsymbol{k}'})$$
$$- f_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}'})\,n_{\boldsymbol{q}\lambda}\,\delta(\varepsilon_{\boldsymbol{k}} + \hbar\omega_{\boldsymbol{q}\lambda} - \varepsilon_{\boldsymbol{k}'}) \Big\}\, \delta_{\boldsymbol{q},\boldsymbol{k}'-\boldsymbol{k} \bmod \boldsymbol{G}} \quad ,$$

which is a functional of both the electron distribution $f_{\boldsymbol{k}}$ as well as the phonon distribution $n_{\boldsymbol{q}\lambda}$. The four terms inside the curly brackets correspond, respectively, to cases (a) through (d) in Fig. 5.14.

Collisional invariants

While collisions will violate crystal momentum conservation, they do not violate conservation of particle number. Hence we should have[12]

$$\int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \mathcal{I}_{\boldsymbol{k}}[f] = 0 \quad . \tag{5.166}$$

[12]If collisions are purely local, then $\int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \mathcal{I}_{\boldsymbol{k}}[f] = 0$ at every point $\boldsymbol{r}$ in space.

The total particle number,

$$N = \int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \; f(\boldsymbol{r}, \boldsymbol{k}, t) \tag{5.167}$$

is a *collisional invariant* - a quantity which is preserved in the collision process. Other collisional invariants include energy (when all sources are accounted for), spin (total spin), and crystal momentum (if there is no breaking of lattice translation symmetry)[13]. Consider a function $F(\boldsymbol{r}, \boldsymbol{k})$ of position and wavevector. Its average value is

$$\bar{F}(t) = \int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, F(\boldsymbol{r}, \boldsymbol{k}) \, f(\boldsymbol{r}, \boldsymbol{k}, t) \quad . \tag{5.168}$$

Taking the time derivative,

$$\begin{aligned} \frac{d\bar{F}}{dt} = \frac{\partial \bar{F}}{\partial t} &= \int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, F(\boldsymbol{r}, \boldsymbol{k}) \left\{ -\frac{\partial}{\partial \boldsymbol{r}} \cdot (\dot{\boldsymbol{r}} f) - \frac{\partial}{\partial \boldsymbol{k}} \cdot (\dot{\boldsymbol{k}} f) + \mathcal{I}_{\boldsymbol{k}}[f] \right\} \\ &= \int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \left\{ \left[\frac{\partial F}{\partial \boldsymbol{r}} \cdot \frac{d\boldsymbol{r}}{dt} + \frac{\partial F}{\partial \boldsymbol{k}} \cdot \frac{d\boldsymbol{k}}{dt} \right] f + F \, \mathcal{I}_{\boldsymbol{k}}[f] \right\} \quad . \end{aligned} \tag{5.169}$$

Hence, if F is preserved by the dynamics between collisions, then

$$\frac{d\bar{F}}{dt} = \int d^3r \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, F \, \mathcal{I}_{\boldsymbol{k}}[f] \quad , \tag{5.170}$$

which says that $\bar{F}(t)$ changes only as a result of collisions. If F is a collisional invariant, then $\dot{\bar{F}} = 0$. This is the case when $F = 1$, in which case $\bar{F}$ is the total number of particles, or when $F = \varepsilon(\boldsymbol{k})$, in which case $\bar{F}$ is the total energy.

5.6.2 Local equilibrium

The equilibrium Fermi distribution,

$$f^0(\boldsymbol{k}) = \left\{ \exp\left(\frac{\varepsilon(\boldsymbol{k}) - \mu}{k_{\rm B} T} \right) + 1 \right\}^{-1} \tag{5.171}$$

is a space-independent and time-independent solution to the Boltzmann equation. Since collisions act *locally* in space, they act on short time scales to establish a *local equilibrium* described

[13] Note that the relaxation time approximation violates all such conservation laws. Within the relaxation time approximation, there are no collisional invariants.

by a distribution function

$$f^0(\boldsymbol{r},\boldsymbol{k},t) = \left\{\exp\left(\frac{\varepsilon(\boldsymbol{k})-\mu(\boldsymbol{r},t)}{k_{\rm B}T(\boldsymbol{r},t)}\right)+1\right\}^{-1} \quad . \tag{5.172}$$

This is, however, not a solution to the full Boltzmann equation due to the 'streaming terms' $\dot{\boldsymbol{r}}\cdot\partial_{\boldsymbol{r}} + \dot{\boldsymbol{k}}\cdot\partial_{\boldsymbol{k}}$ in the convective derivative. These, though, act on longer time scales than those responsible for the establishment of local equilibrium. To obtain a solution, we write

$$f(\boldsymbol{r},\boldsymbol{k},t) = f^0(\boldsymbol{r},\boldsymbol{k},t) + \delta f(\boldsymbol{r},\boldsymbol{k},t) \tag{5.173}$$

and solve for the deviation $\delta f(\boldsymbol{r},\boldsymbol{k},t)$. We will assume $\mu = \mu(\boldsymbol{r})$ and $T = T(\boldsymbol{r})$ are time-independent. We first compute the differential of f^0,

$$\begin{aligned} df^0 &= k_{\rm B}T\,\frac{\partial f^0}{\partial\varepsilon}\,d\left(\frac{\varepsilon-\mu}{k_{\rm B}T}\right) \\ &= k_{\rm B}T\,\frac{\partial f^0}{\partial\varepsilon}\left\{-\frac{d\mu}{k_{\rm B}T} - \frac{(\varepsilon-\mu)\,dT}{k_{\rm B}T^2} + \frac{d\varepsilon}{k_{\rm B}T}\right\} \\ &= -\frac{\partial f^0}{\partial\varepsilon}\left\{\frac{\partial\mu}{\partial\boldsymbol{r}}\cdot d\boldsymbol{r} + \frac{\varepsilon-\mu}{T}\,\frac{\partial T}{\partial\boldsymbol{r}}\cdot d\boldsymbol{r} - \frac{\partial\varepsilon}{\partial\boldsymbol{k}}\cdot d\boldsymbol{k}\right\} \quad , \end{aligned} \tag{5.174}$$

from which we read off

$$\begin{aligned} \frac{\partial f^0}{\partial\boldsymbol{r}} &= \left\{\frac{\partial\mu}{\partial\boldsymbol{r}} + \frac{\varepsilon-\mu}{T}\,\frac{\partial T}{\partial\boldsymbol{r}}\right\}\left(-\frac{\partial f^0}{\partial\varepsilon}\right) \\ \frac{\partial f^0}{\partial\boldsymbol{k}} &= \hbar\boldsymbol{v}\,\frac{\partial f^0}{\partial\varepsilon} \quad . \end{aligned} \tag{5.175}$$

We thereby obtain

$$\frac{\partial\delta f}{\partial t} + \boldsymbol{v}\cdot\boldsymbol{\nabla}\,\delta f - \frac{e}{\hbar}\left[\boldsymbol{E}+\frac{1}{c}\boldsymbol{v}\times\boldsymbol{B}\right]\cdot\frac{\partial\,\delta f}{\partial\boldsymbol{k}} + \boldsymbol{v}\cdot\left[e\,\boldsymbol{\mathcal{E}} + \frac{\varepsilon-\mu}{T}\boldsymbol{\nabla}T\right]\left(-\frac{\partial f^0}{\partial\varepsilon}\right) = \mathcal{I}_{\boldsymbol{k}}\big[f^0+\delta f\big] \quad , \tag{5.176}$$

where $\boldsymbol{\mathcal{E}} = -\boldsymbol{\nabla}(\phi-\mu/e)$ is the gradient of the 'electrochemical potential'; we'll henceforth refer to $\boldsymbol{\mathcal{E}}$ as the electric field. Eqn (5.176) is a nonlinear integrodifferential equation in δf, with the nonlinearity coming from the collision integral. (In some cases, such as impurity scattering, the collision integral may be a linear functional.) We will solve a *linearized* version of this equation, assuming the system is always close to a state of local equilibrium.

Note that the inhomogeneous term in (5.176) involves the electric field and the temperature gradient $\boldsymbol{\nabla}T$. This means that δf is proportional to these quantities, and if they are small then δf is small. The gradient of δf is then of second order in smallness, since the external fields

$\phi - \mu/e$ and T are assumed to be slowly varying in space. To lowest order in smallness, then, we obtain the following *linearized Boltzmann equation*:

$$\frac{\partial \delta f}{\partial t} - \frac{e}{\hbar c}\,\boldsymbol{v}\times\boldsymbol{B}\cdot\frac{\partial\,\delta f}{\partial \boldsymbol{k}} + \boldsymbol{v}\cdot\left[e\,\boldsymbol{\mathcal{E}} + \frac{\varepsilon-\mu}{T}\boldsymbol{\nabla}T\right]\left(-\frac{\partial f^0}{\partial\varepsilon}\right) = \mathcal{L}\,\delta f \quad , \tag{5.177}$$

where $\mathcal{L}\,\delta f$ is the *linearized collision integral*; $\mathcal{L}$ is a linear operator acting on δf (we suppress denoting the $\boldsymbol{k}$ dependence of $\mathcal{L}$). Note that we have not assumed that $\boldsymbol{B}$ is small. Indeed later on we will derive expressions for high B transport coefficients.

Note also that we also have dropped the term

$$\frac{\partial f^0}{\partial t} = -\frac{\partial f^0}{\partial\varepsilon}\left\{\frac{\partial\mu}{\partial t} + \frac{\varepsilon-\mu}{T}\,\frac{\partial T}{\partial t}\right\} \tag{5.178}$$

from the LHS of the linearized Boltzmann equation. This is because we assume that the spatially uniform components of $\mu(\boldsymbol{r},t)$ and $T(\boldsymbol{r},t)$ are time-independent, which means that the nonzero contributions to $\partial\mu/\partial t$ and $\partial T/\partial t$ involve at least one space derivative as well as one time derivative, and are thus doubly small and therefore negligible.

5.7 Conductivity of Normal Metals

5.7.1 Relaxation time approximation

Consider a normal metal in the presence of an electric field $\boldsymbol{\mathcal{E}}$. We'll assume $\boldsymbol{B} = 0$, $\boldsymbol{\nabla}T = 0$, and also that $\boldsymbol{\mathcal{E}}$ is spatially uniform as well. This in turn guarantees that δf itself is spatially uniform. The Boltzmann equation then reduces to

$$\frac{\partial\,\delta f}{\partial t} - \frac{\partial f^0}{\partial\varepsilon}\,e\boldsymbol{v}\cdot\boldsymbol{\mathcal{E}} = \mathcal{I}_{\boldsymbol{k}}\big[f^0+\delta f\big] \quad . \tag{5.179}$$

We'll solve this by adopting the relaxation time approximation for $\mathcal{I}_{\boldsymbol{k}}[f]$:

$$\mathcal{I}_{\boldsymbol{k}}[f] = -\frac{f-f^0}{\tau} = -\frac{\delta f}{\tau} \quad , \tag{5.180}$$

where τ, which may be $\boldsymbol{k}$-dependent, is the *relaxation time*. In the absence of any fields or temperature and electrochemical potential gradients, the Boltzmann equation becomes $\dot{\delta f} = -\delta f/\tau$, with the solution $\delta f(t) = \delta f(0)\exp(-t/\tau)$. The distribution thereby relaxes to the equilibrium one on the scale of τ. In fact, this result is *wrong*, because the total particle number is a collisional invariant. Electrons can's simply disappear! Rather, the local number density must relax to the equilibrium value via the slower mechanism of *diffusion*. While the fact that collisional invariants are not respected is a defect of the relaxation time approximation, this won't

much affect the validity of our conclusions regarding various transport coefficients, such as the electrical conductivity.

Writing $\boldsymbol{\mathcal{E}}(t) = \boldsymbol{\mathcal{E}}\, e^{-i\omega t}$, we solve

$$\frac{\partial\, \delta f(\boldsymbol{k},t)}{\partial t} - e\, \boldsymbol{v}(\boldsymbol{k}) \cdot \boldsymbol{\mathcal{E}}\, e^{-i\omega t}\, \frac{\partial f^0}{\partial \varepsilon} = -\frac{\delta f(\boldsymbol{k},t)}{\tau(\varepsilon(\boldsymbol{k}))} \tag{5.181}$$

and obtain

$$\delta f(\boldsymbol{k},t) = \frac{e\, \boldsymbol{\mathcal{E}} \cdot \boldsymbol{v}(\boldsymbol{k})\, \tau(\varepsilon(\boldsymbol{k}))}{1 - i\omega\tau(\varepsilon(\boldsymbol{k}))}\, \frac{\partial f^0}{\partial \varepsilon}\, e^{-i\omega t} \quad . \tag{5.182}$$

The equilibrium distribution $f^0(\boldsymbol{k})$ results in zero current, since $f^0(-\boldsymbol{k}) = f^0(\boldsymbol{k})$. Thus, the current density is given by the expression

$$\begin{aligned} j^\alpha(\boldsymbol{r},t) &= -2e \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \delta f\, v^\alpha \\ &= 2e^2\, \mathcal{E}^\beta\, e^{-i\omega t} \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \frac{\tau(\varepsilon(\boldsymbol{k}))\, v^\alpha(\boldsymbol{k})\, v^\beta(\boldsymbol{k})}{1 - i\omega\tau(\varepsilon(\boldsymbol{k}))} \left(-\frac{\partial f^0}{\partial \varepsilon}\right) \quad . \end{aligned} \tag{5.183}$$

In the above calculation, the factor of two arises from summing over spin polarizations. The conductivity tensor is defined by the linear relation $j^\alpha(\omega) = \sigma_{\alpha\beta}(\omega)\, \mathcal{E}^\beta(\omega)$. We have thus derived an expression for the conductivity tensor,

$$\sigma_{\alpha\beta}(\omega) = 2e^2 \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\, \frac{\tau(\varepsilon(\boldsymbol{k}))\, v^\alpha(\boldsymbol{k})\, v^\beta(\boldsymbol{k})}{1 - i\omega\tau(\varepsilon(\boldsymbol{k}))} \left(-\frac{\partial f^0}{\partial \varepsilon}\right) \quad . \tag{5.184}$$

Note that the conductivity is a property of the *Fermi surface*. For $k_{\rm B}T \ll \varepsilon_{\rm F}$, we have $-\partial f^0/\partial\varepsilon \approx \delta(\varepsilon_{\rm F} - \varepsilon(\boldsymbol{k}))$ and the above integral is over the Fermi surface alone. Explicitly, we change variables to energy ε and coordinates along a constant energy surface, writing

$$d^3k = \frac{d\varepsilon\, dS_\varepsilon}{|\partial\varepsilon/\partial\boldsymbol{k}|} = \frac{d\varepsilon\, dS_\varepsilon}{\hbar|\boldsymbol{v}|} \quad , \tag{5.185}$$

where dS_ε is the differential area on the constant energy surface $\varepsilon(\boldsymbol{k}) = \varepsilon$, and $\boldsymbol{v}(\boldsymbol{k}) = \hbar^{-1}\boldsymbol{\nabla}_{\boldsymbol{k}}\, \varepsilon(\boldsymbol{k})$ is the velocity. For $T \ll T_{\rm F}$, then,

$$\sigma_{\alpha\beta}(\omega) = \frac{e^2}{4\pi^3\hbar}\, \frac{\tau(\varepsilon_{\rm F})}{1 - i\omega\tau(\varepsilon_{\rm F})} \int dS_{\rm F}\, \frac{v^\alpha(\boldsymbol{k})\, v^\beta(\boldsymbol{k})}{|\boldsymbol{v}(\boldsymbol{k})|} \quad . \tag{5.186}$$

For free electrons in a parabolic band, we write $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m^*$, so $v^\alpha(\boldsymbol{k}) = \hbar k^\alpha/m^*$. To further simplify matters, let us assume that τ is constant, or at least very slowly varying in the vicinity of the Fermi surface. We find

$$\sigma_{\alpha\beta}(\omega) = \delta_{\alpha\beta}\, \frac{2}{3m^*}\, \frac{e^2\tau}{1 - i\omega\tau} \int\limits_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, \varepsilon \left(-\frac{\partial f^0}{\partial \varepsilon}\right) \quad , \tag{5.187}$$

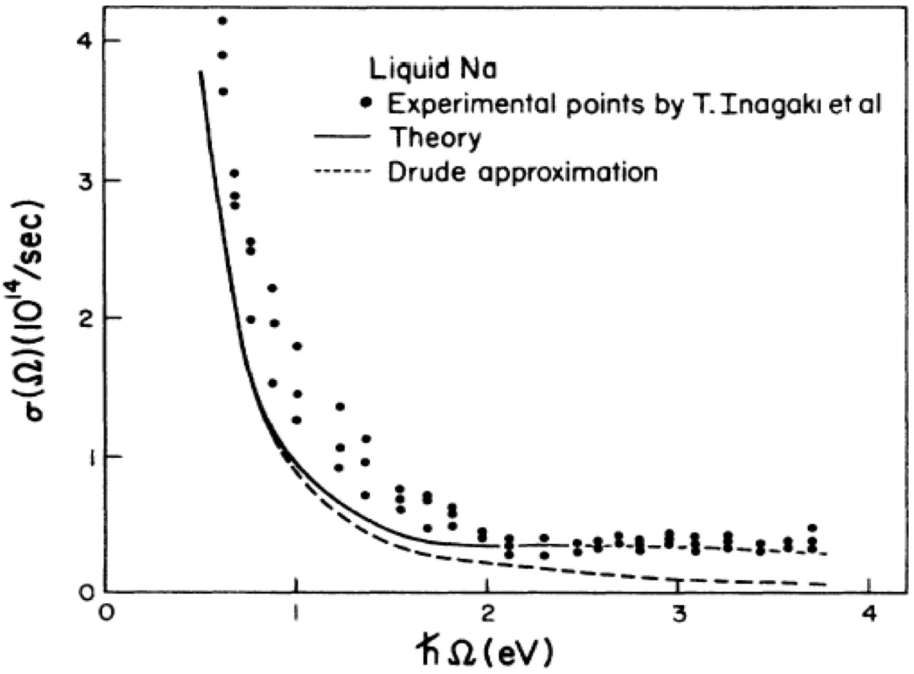


Figure 5.15: Frequency-dependent conductivity of liquid sodium by T. Inagaki *et al*, *Phys. Rev. B* **13**, 5610 (1976).

where $g(\varepsilon)$ is the *density of states*,

$$g(\varepsilon) = 2 \int_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, \delta\left(\varepsilon - \varepsilon(\boldsymbol{k})\right) \quad . \tag{5.188}$$

The (three-dimensional) parabolic band density of states is found to be

$$g(\varepsilon) = \frac{(2m^*)^{3/2}}{2\pi^2\hbar^3} \sqrt{\varepsilon} \, \Theta(\varepsilon) \quad , \tag{5.189}$$

where $\Theta(x)$ is the step function. In fact, integrating (5.187) by parts, we only need to know about the $\sqrt{\varepsilon}$ dependence in $g(\varepsilon)$, and not the details of its prefactor:

$$\int d\varepsilon \, g(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon}\right) = \int d\varepsilon \, f^0(\varepsilon) \frac{\partial}{\partial \varepsilon} \left(\varepsilon \, g(\varepsilon)\right) = \tfrac{3}{2} \int d\varepsilon \, g(\varepsilon) \, f^0(\varepsilon) = \tfrac{3}{2} n \quad , \tag{5.190}$$

where $n = N/V$ is the electron number density for the conduction band. The final result for the conductivity tensor is

$$\sigma_{\alpha\beta}(\omega) = \frac{ne^2\tau}{m^*} \frac{\delta_{\alpha\beta}}{1 - i\omega\tau} \quad . \tag{5.191}$$

We have recovered the the Drude theory of electrical conduction in metals.

5.7.2 Optical conductivity and the Fermi surface

At high frequencies, when $\omega\tau \gg 1$, our expression for the conductivity, Eqn. (5.184), yields

$$\sigma(\omega) = \frac{ie^2}{12\pi^3\hbar\omega}\int d\varepsilon\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\big|\boldsymbol{v}(\boldsymbol{k})\big| \quad , \tag{5.192}$$

where we have presumed sufficient crystalline symmetry to guarantee that $\sigma_{\alpha\beta} = \sigma\,\delta_{\alpha\beta}$ is diagonal. In the isotropic case, and at temperatures low compared with $T_{\rm F}$, the integral over the Fermi surface gives $4\pi k_{\rm F}^2\, v_{\rm F} = 12\pi^3\hbar n/m^*$, whence $\sigma = ine^2/m^*\omega$, which is the large frequency limit of our previous result. For a general Fermi surface, we can define

$$\sigma(\omega \gg \tau^{-1}) \equiv \frac{ine^2}{m_{\rm opt}\omega} \tag{5.193}$$

where the *optical mass* $m_{\rm opt}$ is given by

$$\frac{1}{m_{\rm opt}} = \frac{1}{12\pi^3\hbar n}\int d\varepsilon\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,\big|\boldsymbol{v}(\boldsymbol{k})\big| \quad . \tag{5.194}$$

Note that at high frequencies $\sigma(\omega)$ is purely imaginary. What does this mean? If

$$\boldsymbol{\mathcal{E}}(t) = \boldsymbol{\mathcal{E}}\cos(\omega t) = \tfrac{1}{2}\,\boldsymbol{\mathcal{E}}\left(e^{-i\omega t} + e^{+i\omega t}\right) \tag{5.195}$$

then

$$\boldsymbol{j}(t) = \tfrac{1}{2}\,\boldsymbol{\mathcal{E}}\,\left\{\sigma(\omega)\,e^{-i\omega t} + \sigma(-\omega)\,e^{+i\omega t}\right\} = \frac{ne^2}{m_{\rm opt}\omega}\,\boldsymbol{\mathcal{E}}\,\sin(\omega t) \quad , \tag{5.196}$$

where we have invoked $\sigma(-\omega) = \sigma^*(\omega)$. The current is therefore 90° out of phase with the voltage, and the average over a cycle $\langle \boldsymbol{j}(t)\cdot\boldsymbol{\mathcal{E}}(t)\rangle = 0$. Recall that we found metals to be transparent for $\omega \gg \omega_{\rm p} \gg \tau^{-1}$.

At zero temperature, the optical mass is given by

$$\frac{1}{m_{\rm opt}} = \frac{1}{12\pi^3\hbar n}\int dS_{\rm F}\,\big|\boldsymbol{v}(\boldsymbol{k})\big| \quad . \tag{5.197}$$

The density of states, $g(\varepsilon_{\rm F})$, is

$$g(\varepsilon_{\rm F}) = \frac{1}{4\pi^3\hbar}\int dS_{\rm F}\,\big|\boldsymbol{v}(\boldsymbol{k})\big|^{-1} \quad , \tag{5.198}$$

from which one can define the thermodynamic effective mass $m^*_{\rm th}$, appealing to the low temperature form of the specific heat,

$$c_V = \frac{\pi^2}{3}\,k_{\rm B}^2\,T\,g(\varepsilon_{\rm F}) \equiv \frac{m^*_{\rm th}}{m_{\rm e}}\,c_V^0 \quad , \tag{5.199}$$

Metal	$m^*_{\rm opt}/m_{\rm e}$ thy	expt	$m^*_{\rm th}/m_{\rm e}$ thy	expt
Li	1.45	1.57	1.64	2.23
Na	1.00	1.13	1.00	1.27
K	1.02	1.16	1.07	1.26
Rb	1.08	1.16	1.18	1.36
Cs	1.29	1.19	1.75	1.79
Cu	-	-	1.46	1.38
Ag	-	-	1.00	1.00
Au	-	-	1.09	1.08

Table 5.1: Optical and thermodynamic effective masses of monovalent metals. (Taken from Smith and Jensen).

where

$$c_V^0 \equiv \frac{m_{\rm e}\, k_{\rm B}^2\, T}{3\hbar^2}\,(3\pi^2 n)^{1/3} \tag{5.200}$$

is the specific heat for a free electron gas of density n. Thus,

$$m^*_{\rm th} = \frac{\hbar}{4\pi(3\pi^2 n)^{1/3}} \int dS_{\rm F}\,\big|\boldsymbol{v}(\boldsymbol{k})\big|^{-1} \tag{5.201}$$

5.8 Calculation of the Scattering Time

5.8.1 Potential scattering and Fermi's golden rule

Let us go beyond the relaxation time approximation and calculate the scattering time τ from first principles. We will concern ourselves with scattering of electrons from crystalline impurities. We begin with Fermi's Golden Rule[14],

$$\mathcal{I}_{\boldsymbol{k}}[f] = \frac{2\pi}{\hbar}\sum_{\boldsymbol{k}'}\big|\big\langle\,\boldsymbol{k}'\,\big|\,\mathcal{U}\,\big|\,\boldsymbol{k}\,\big\rangle\big|^2\,(f_{\boldsymbol{k}'}-f_{\boldsymbol{k}})\,\delta(\varepsilon(\boldsymbol{k})-\varepsilon(\boldsymbol{k}'))\quad , \tag{5.202}$$

where $\mathcal{U}(\boldsymbol{r})$ is a sum over individual impurity ion potentials, $\mathcal{U}(\boldsymbol{r}) = \sum_{j=1}^{N_{\rm imp}} U(\boldsymbol{r}-\boldsymbol{R}_j)$. Thus,

$$\big|\big\langle\,\boldsymbol{k}'\,\big|\,\mathcal{U}\,\big|\,\boldsymbol{k}\,\big\rangle\big|^2 = V^{-2}\,|\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2\cdot\left|\sum_{j=1}^{N_{\rm imp}} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}_j}\right|^2\quad , \tag{5.203}$$

[14] We'll treat the scattering of each spin species separately. We assume no spin-flip scattering takes place.

where V is the volume of the solid and

$$\hat{U}(\boldsymbol{q}) = \int d^3r\, U(\boldsymbol{r})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} \tag{5.204}$$

is the Fourier transform of the impurity potential. Note that we are assuming a single species of impurities; the method can be generalized to account for different impurity species.

To make progress, we assume the impurity positions are random and uncorrelated, and we average over them. Using

$$\overline{\left|\sum_{j=1}^{N_{\rm imp}} e^{i\boldsymbol{q}\cdot\boldsymbol{R}_j}\right|^2} = N_{\rm imp} + N_{\rm imp}(N_{\rm imp}-1)\,\delta_{\boldsymbol{q},0} \quad , \tag{5.205}$$

we obtain

$$\overline{\left|\langle\, \boldsymbol{k}' \,|\, \mathcal{U} \,|\, \boldsymbol{k}\,\rangle\right|^2} = \frac{N_{\rm imp}}{V^2}\,|\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2 + \frac{N_{\rm imp}(N_{\rm imp}-1)}{V^2}\,|\hat{U}(0)|^2\,\delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \tag{5.206}$$

EXERCISE: Verify Eqn. (5.205).

We will neglect the second term in Eqn. 5.206 arising from the spatial average ($\boldsymbol{q} = 0$ Fourier component) of the potential. As we will see, in the end it will cancel out. Writing $f = f^0 + \delta f$, we have

$$\mathcal{I}_{\boldsymbol{k}}[f] = \frac{2\pi n_{\rm imp}}{\hbar}\int\limits_{\hat{\Omega}} \frac{d^3k'}{(2\pi)^3}\,|\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2\,\delta\left(\frac{\hbar^2\boldsymbol{k}^2}{2m^*} - \frac{\hbar^2\boldsymbol{k}'^2}{2m^*}\right)(\delta f_{\boldsymbol{k}'} - \delta f_{\boldsymbol{k}}) \quad , \tag{5.207}$$

where $n_{\rm imp} = N_{\rm imp}/V$ is the number density of impurities. Note that we are assuming a parabolic band. We next make the *Ansatz*

$$\delta f_{\boldsymbol{k}} = \tau(\varepsilon(\boldsymbol{k}))\, e\,\boldsymbol{\mathcal{E}}\cdot\boldsymbol{v}(\boldsymbol{k})\,\frac{\partial f^0}{\partial\varepsilon}\bigg|_{\varepsilon(\boldsymbol{k})} \tag{5.208}$$

and solve for $\tau(\varepsilon(\boldsymbol{k}))$. The (time-independent) Boltzmann equation is

$$\begin{aligned} -e\,\boldsymbol{\mathcal{E}}\cdot\boldsymbol{v}(\boldsymbol{k})\,\frac{\partial f^0}{\partial\varepsilon} = \frac{2\pi}{\hbar}\,n_{\rm imp}\,e\,\boldsymbol{\mathcal{E}}\cdot\int\limits_{\hat{\Omega}}\frac{d^3k'}{(2\pi)^3}\,|\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2\,\delta\left(\frac{\hbar^2\boldsymbol{k}^2}{2m^*} - \frac{\hbar^2\boldsymbol{k}'^2}{2m^*}\right) \\ \times\left(\tau(\varepsilon(\boldsymbol{k}'))\,\boldsymbol{v}(\boldsymbol{k}')\,\frac{\partial f^0}{\partial\varepsilon}\bigg|_{\varepsilon(\boldsymbol{k}')} - \tau(\varepsilon(\boldsymbol{k}))\,\boldsymbol{v}(\boldsymbol{k})\,\frac{\partial f^0}{\partial\varepsilon}\bigg|_{\varepsilon(\boldsymbol{k})}\right) \quad . \end{aligned} \tag{5.209}$$

Due to the isotropy of the problem, we must have $\tau(\varepsilon(\boldsymbol{k}))$ is a function only of the magnitude of $\boldsymbol{k}$. We then obtain[15]

$$\frac{\hbar\boldsymbol{k}}{m^*} = \frac{n_{\rm imp}}{4\pi^2\hbar}\,\tau(\varepsilon(\boldsymbol{k}))\int\limits_0^\infty dk'\,k'^2\int d\hat{\boldsymbol{k}}'\,|\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2\,\frac{\delta(k-k')}{\hbar^2k/m^*}\,\frac{\hbar}{m^*}(\boldsymbol{k}-\boldsymbol{k}') \quad , \tag{5.210}$$

[15]We assume that the Fermi surface is contained within the first Brillouin zone.

whence

$$\frac{1}{\tau(\varepsilon_{\rm F})} = \frac{m^*\,k_{\rm F}\,n_{\rm imp}}{4\pi^2\hbar^3}\int d\hat{\boldsymbol{k}}'\,|U(k_{\rm F}\hat{\boldsymbol{k}} - k_{\rm F}\hat{\boldsymbol{k}}')|^2\,(1-\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}') \quad . \tag{5.211}$$

If the impurity potential $U(\boldsymbol{r})$ itself is isotropic, then its Fourier transform $\hat{U}(\boldsymbol{q})$ is a function of $q^2 = 4k_{\rm F}^2\sin^2\frac{1}{2}\vartheta$ where $\cos\vartheta = \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'$ and $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k}$ is the transfer wavevector. Recalling the Born approximation for differential scattering cross section,

$$\sigma(\vartheta) = \left(\frac{m^*}{2\pi\hbar^2}\right)^2 |\hat{U}(\boldsymbol{k}-\boldsymbol{k}')|^2 \quad , \tag{5.212}$$

we may finally write

$$\frac{1}{\tau(\varepsilon_{\rm F})} = 2\pi n_{\rm imp} v_{\rm F}\int_0^\pi d\vartheta\,\sigma_{\rm F}(\vartheta)\,(1-\cos\vartheta)\,\sin\vartheta \quad , \tag{5.213}$$

where $v_{\rm F} = \hbar k_{\rm F}/m^*$ is the Fermi velocity[16]. The *mean free path* is defined by $\ell = v_{\rm F}\tau$.

Notice the factor $(1-\cos\vartheta)$ in the integrand of (5.213). This tells us that forward scattering ($\vartheta = 0$) doesn't contribute to the scattering rate, which justifies our neglect of the second term in Eqn. (5.206). Why should τ be utterly insensitive to forward scattering? Because $\tau(\varepsilon_{\rm F})$ is the *transport lifetime*, and forward scattering does not degrade the current. Therefore, $\sigma(\vartheta = 0)$ does not contribute to the 'transport scattering rate' $\tau^{-1}(\varepsilon_{\rm F})$. Oftentimes one sees reference in the literature to a 'single particle lifetime' as well, which is given by the same expression but without this factor:

$$\left\{\begin{matrix}\tau_{\rm sp}^{-1}\\ \tau_{\rm tr}^{-1}\end{matrix}\right\} = 2\pi n_{\rm imp} v_{\rm F}\int_0^\pi d\vartheta\,\sigma_{\rm F}(\vartheta)\left\{\begin{matrix}1\\ (1-\cos\vartheta)\end{matrix}\right\}\sin\vartheta \quad . \tag{5.214}$$

Note that $\tau_{\rm sp} = (n_{\rm imp}\,v_{\rm F}\,\sigma_{\rm F,tot})^{-1}$, where $\sigma_{\rm F,tot}$ is the total scattering cross section at energy $\varepsilon_{\rm F}$, a formula familiar from elementary kinetic theory.

The Boltzmann equation defines an infinite hierarchy of lifetimes classified by the angular momentum scattering channel. To derive this hierarchy, one can examine the linearized time-*dependent* Boltzmann equation with $\boldsymbol{\mathcal{E}} = 0$,

$$\frac{\partial\,\delta f_{\boldsymbol{k}}}{\partial t} = n_{\rm imp}\,v_{\rm F}\int d\hat{\boldsymbol{k}}'\,\sigma(\vartheta_{\boldsymbol{k}\boldsymbol{k}'})\,(\delta f_{\boldsymbol{k}'} - \delta f_{\boldsymbol{k}}) \quad , \tag{5.215}$$

where $v = \hbar k/m^*$ is the velocity, and where the kernel is $\vartheta_{\boldsymbol{k}\boldsymbol{k}'} = \cos^{-1}(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}')$. We now expand in spherical harmonics, writing

$$\sigma(\vartheta_{\boldsymbol{k}\boldsymbol{k}'}) \equiv \sigma_{\rm tot}\sum_{L,M}\nu_L\,Y_{LM}(\hat{\boldsymbol{k}})\,Y_{LM}^*(\hat{\boldsymbol{k}}') \quad , \tag{5.216}$$

[16]The subscript on $\sigma_{\rm F}(\vartheta)$ is to remind us that the cross section depends on $k_{\rm F}$ as well as ϑ.

where as before

$$\sigma_{\rm tot} = 2\pi \int\limits_0^\pi d\vartheta\ \sin\vartheta\, \sigma(\vartheta) \quad , \tag{5.217}$$

which fixes $\nu_{L=0} = 1$. Expanding

$$\delta f_{\boldsymbol{k}}(t) = \sum_{L,M} A_{LM}(t)\, Y_{LM}(\hat{\boldsymbol{k}}) \quad , \tag{5.218}$$

the linearized Boltzmann equation simplifies to

$$\frac{\partial A_{LM}}{\partial t} + (1-\nu_L)\, n_{\rm imp} v_{\rm F}\, \sigma_{\rm tot}\, A_{LM} = 0 \quad , \tag{5.219}$$

whence one obtains a hierarchy of relaxation rates,

$$\tau_L^{-1} = (1-\nu_L)\, n_{\rm imp} v_{\rm F}\, \sigma_{\rm tot} \quad , \tag{5.220}$$

which depend only on the total angular momentum quantum number L. These rates describe the relaxation of nonuniform distributions when $\delta f_{\boldsymbol{k}}(t=0)$ is proportional to some spherical harmonic $Y_{LM}(\boldsymbol{k})$. Note that $\tau_{L=0}^{-1} = 0$, which reflects the fact that the total particle number is a collisional invariant. The single particle lifetime is identified as

$$\tau_{\rm sp} \equiv \tau_{L\to\infty} = \big(n_{\rm imp} v_{\rm F}\, \sigma_{\rm tot}\big)^{-1} \quad , \tag{5.221}$$

corresponding to a point distortion of the uniform distribution. The transport lifetime is then $\tau_{\rm tr} = \tau_{L=1}$.

5.8.2 Screening and the transport lifetime

For a Coulomb impurity, with $U(\boldsymbol{r}) = -Ze^2/r$ we have $\hat{U}(\boldsymbol{q}) = -4\pi Ze^2/q^2$. Consequently,

$$\sigma_{\rm F}(\vartheta) = \left(\frac{Ze^2}{4\varepsilon_{\rm F} \sin^2\frac{1}{2}\vartheta}\right)^2 \quad , \tag{5.222}$$

and there is a strong divergence as $\vartheta \to 0$, with $\sigma_{\rm F}(\vartheta) \propto \vartheta^{-4}$. The transport lifetime diverges logarithmically! What went wrong?

What went wrong is that we have failed to account for *screening*. Free charges will rearrange themselves so as to screen an impurity potential. At long range, the effective (screened) potential decays exponentially, rather than as $1/r$. The screened potential is of the Yukawa form, and its increase at low q is cut off on the scale of the inverse screening length λ^{-1}. There are two types of screening to consider:

- Thomas-Fermi Screening : This is the typical screening mechanism in metals. A weak local electrostatic potential $\phi(\boldsymbol{r})$ will induce a change in the local electronic density according to $\delta n(\boldsymbol{r}) = e\phi(\boldsymbol{r})\, g(\varepsilon_{\rm F})$, where $g(\varepsilon_{\rm F})$ is the density of states at the Fermi level. This charge imbalance is again related to $\phi(\boldsymbol{r})$ through the Poisson equation. The result is a self-consistent equation for $\phi(\boldsymbol{r})$,

$$\begin{aligned}\nabla^2\phi &= 4\pi e\,\delta n \\ &= 4\pi e^2 g(\varepsilon_{\rm F})\,\phi \equiv \lambda_{\rm TF}^{-2}\,\phi \quad .\end{aligned} \tag{5.223}$$

  The Thomas-Fermi screening length is $\lambda_{\rm TF} = \big(4\pi e^2 g(\varepsilon_{\rm F})\big)^{-1/2}$.

- Debye-Hückel Screening : This mechanism is typical of ionic solutions, although it may also be of relevance in solids with ultra-low Fermi energies. From classical statistical mechanics, the local variation in electron number density induced by a potential $\phi(\boldsymbol{r})$ is

$$\delta n(\boldsymbol{r}) = n\, e^{e\phi(\boldsymbol{r})/k_{\rm B}T} - n \approx \frac{ne\phi(\boldsymbol{r})}{k_{\rm B}T} \quad , \tag{5.224}$$

  where we assume the potential is weak on the scale of $k_{\rm B}T/e$. Poisson's equation now gives us

$$\nabla^2\phi = 4\pi e\,\delta n = \frac{4\pi ne^2}{k_{\rm B}T}\,\phi \equiv \lambda_{\rm DH}^{-2}\,\phi \quad . \tag{5.225}$$

A screened test charge Ze at the origin obeys

$$\nabla^2\phi = \lambda^{-2}\,\phi - 4\pi Ze\,\delta(\boldsymbol{r}) \quad , \tag{5.226}$$

the solution of which is

$$U(\boldsymbol{r}) = -e\phi(\boldsymbol{r}) = -\frac{Ze^2}{r}\,e^{-r/\lambda} \quad \Longrightarrow \quad \hat{U}(\boldsymbol{q}) = -\frac{4\pi Ze^2}{\boldsymbol{q}^2 + \lambda^{-2}} \quad . \tag{5.227}$$

The differential scattering cross section is now

$$\sigma_{\rm F}(\vartheta) = \left(\frac{Ze^2}{4\varepsilon_{\rm F}}\cdot\frac{1}{\sin^2\frac{1}{2}\vartheta + (2k_{\rm F}\lambda)^{-2}}\right)^2 \tag{5.228}$$

and the divergence at small angle is cut off. The transport lifetime for screened Coulomb scattering is therefore given by

$$\begin{aligned}\frac{1}{\tau(\varepsilon_{\rm F})} &= 2\pi n_{\rm imp} v_{\rm F}\left(\frac{Ze^2}{4\varepsilon_{\rm F}}\right)^2 \int\limits_0^\pi d\vartheta\,\sin\vartheta\,(1-\cos\vartheta)\,\left(\frac{1}{\sin^2\frac{1}{2}\vartheta + (2k_{\rm F}\lambda)^{-2}}\right)^2 \\ &= 2\pi n_{\rm imp} v_{\rm F}\left(\frac{Ze^2}{2\varepsilon_{\rm F}}\right)^2 \left\{\ln(1+\pi\zeta) - \frac{\pi\zeta}{1+\pi\zeta}\right\} \quad ,\end{aligned} \tag{5.229}$$

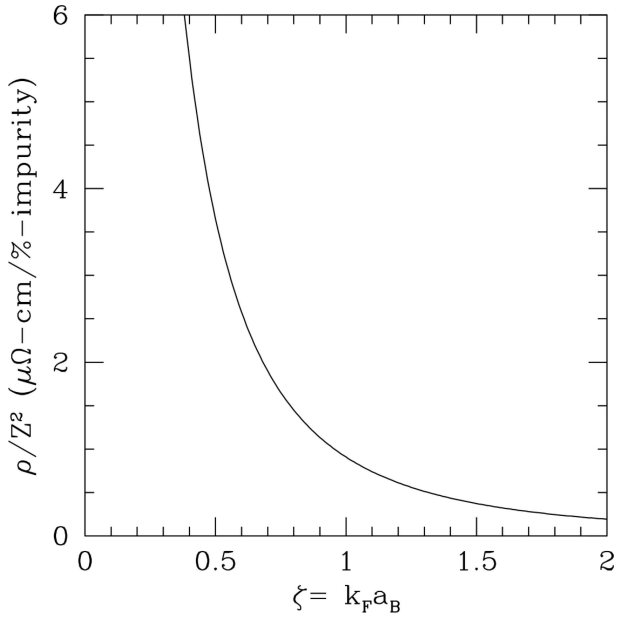


Figure 5.16: Residual resistivity per percent impurity.

with $\zeta = \frac{4}{\pi} k_{\rm F}^2 \lambda^2$, In the case of Thomas-Fermi screening, from $g(\varepsilon_{\rm F}) = m^* k_{\rm F}/\pi^2\hbar^2$, and we have

$$\zeta = \frac{4}{\pi} k_{\rm F}^2 \lambda^2 = \frac{\hbar^2 k_{\rm F}}{m^* e^2} = k_{\rm F} a_{\rm B}^* \quad . \tag{5.230}$$

Here $a_{\rm B}^* = \epsilon_\infty \, \hbar^2/m^* e^2$ is the effective Bohr radius (restoring the ϵ_∞ factor). The resistivity is therefore given by

$$\rho = \frac{m^*}{n e^2 \tau} = \frac{h}{e^2} \, Z^2 \, a_{\rm B}^* \, (n_{\rm imp}/n) \, F(k_{\rm F} \, a_{\rm B}^*) \quad , \tag{5.231}$$

where

$$F(\zeta) = \frac{1}{\zeta^3} \left\{ \ln(1+\pi\zeta) - \frac{\pi\zeta}{1+\pi\zeta} \right\} \quad . \tag{5.232}$$

With $h/e^2 = 25,813\,\Omega$ and $a_{\rm B}^* \approx a_{\rm B} = 0.529\,\text{Å}$, we have

$$\rho = 1.37 \times 10^{-4}\,\Omega \cdot \text{cm} \; \times Z^2 \, (n_{\rm imp}/n) \, F(k_{\rm F} a_{\rm B}) \quad . \tag{5.233}$$

In Tab. 5.2, we show the observed residual resistivity per percent impurity for various ions in copper at low temperatures.

Impurity Ion	$\Delta\rho$ per % ($\mu\Omega$-cm)	Impurity Ion	$\Delta\rho$ per % ($\mu\Omega$-cm)
Be	0.64	Si	3.2
Mg	0.60	Ge	3.7
B	1.4	Sn	2.8
Al	1.2	As	6.5
In	1.2	Sb	5.4

Table 5.2: Observed residual resistivity of copper per percent impurity. (From Smith and Jensen.)

5.9 Dynamics of Holes

5.9.1 Properties of holes

Since filled bands carry no current, we have that the current density from band n is

$$\boldsymbol{j}_n(\boldsymbol{r},t) = -2e\int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\; f_n(\boldsymbol{r},\boldsymbol{k},t)\,\boldsymbol{v}_n(\boldsymbol{k}) = +2e\int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\; \bar{f}_n(\boldsymbol{r},\boldsymbol{k},t)\,\boldsymbol{v}_n(\boldsymbol{k}) \quad , \tag{5.234}$$

where $\bar{f} \equiv 1 - f$. Thus, we can regard the current to be carried by fictitious particles of charge $+e$ with a distribution $\bar{f}(\boldsymbol{r},\boldsymbol{k},t)$. These fictitious particles are called *holes*.

The Four Laws of Holes

1. Under the influence of an applied electromagnetic field, the unoccupied levels of a band evolve as if they were occupied by real electrons of charge $-e$. That is, whether or not a state is occupied is irrelevant to the time evolution of that state, which is described by the semiclassical dynamics of eqs. (5.146, 5.147).

2. The current density due to a hole of wavevector $\boldsymbol{k}$ is $+e\,\boldsymbol{v}_n(\boldsymbol{k})/V$.

3. The crystal momentum of a hole of wavevector $\boldsymbol{k}$ is $\boldsymbol{P} = -\hbar\boldsymbol{k}$.

4. Any band can be described in terms of electrons or in terms of holes, but not both simultaneously. A "mixed" description is redundant at best, wrong at worst, and confusing always. However, it is often convenient to treat some bands within the electron picture and others within the hole picture.

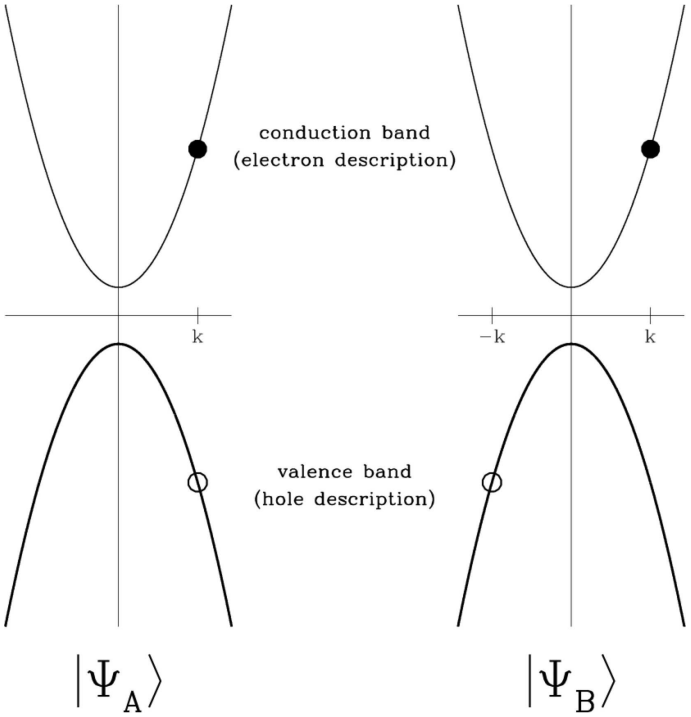


Figure 5.17: Two states: $\big|\,\Psi_{\rm A}\,\big\rangle = e_k^\dagger\,h_k^\dagger\big|\,0\,\big\rangle$ and $\big|\,\Psi_{\rm B}\,\big\rangle = e_k^\dagger\,h_{-k}^\dagger\big|\,0\,\big\rangle$. Which state carries more current? What is the crystal momentum of each state?

It is instructive to consider the exercise of Fig. 5.17. The two states to be analyzed are

$$\begin{aligned}\big|\,\Psi_{\rm A}\,\big\rangle &= \psi_{{\rm c},k}^\dagger\,\psi_{{\rm v},k}\big|\,\Psi_0\,\big\rangle = e_k^\dagger\,h_k^\dagger\big|\,0\,\big\rangle \\ \big|\,\Psi_{\rm B}\,\big\rangle &= \psi_{{\rm c},k}^\dagger\,\psi_{{\rm v},-k}\big|\,\Psi_0\,\big\rangle = e_k^\dagger\,h_{-k}^\dagger\big|\,0\,\big\rangle \quad ,\end{aligned} \tag{5.235}$$

where $e_k^\dagger \equiv \psi_{{\rm c},k}^\dagger$ is the creation operator for electrons in the conduction band, and $h_k^\dagger \equiv \psi_{{\rm v},k}$ is the creation operator for holes (and hence the destruction operator for electrons) in the valence band. The state $\big|\,\Psi_0\,\big\rangle$ has all states below the top of the valence band filled, and all states above the bottom of the conduction band empty. The state $\big|\,0\,\big\rangle$ is the same state, but represented now as a *vacuum* for conduction electrons and valence *holes*. The current density in each state is given by $j = e(v_{\rm h} - v_{\rm e})/V$, where V is the volume (*i.e.* length) of the system. The dispersions are taken to be $\varepsilon_{\rm c,v}(k) = \pm\frac{1}{2}E_{\rm g} \pm \hbar^2k^2/2m^*$, where $E_{\rm g}$ is the energy gap.

- State $\big|\,\Psi_{\rm A}\,\big\rangle$:
  The electron velocity is $v_{\rm e} = \hbar k/m^*$; the hole velocity is $v_{\rm h} = -\hbar k/m^*$. The total current density is $j = x - 2e\hbar k/m^*V$ and the total crystal momentum is $P = p_{\rm e} + p_{\rm h} = \hbar k - \hbar k = 0$.
- State $\big|\,\Psi_{\rm B}\,\big\rangle$:
  The electron velocity is $v_{\rm e} = \hbar k/m^*$; the hole velocity is $v_{\rm h} = -\hbar(-k)/m^*$. The total current density is $j = 0$, and the total crystal momentum is $P = p_{\rm e} + p_{\rm h} = \hbar k - \hbar(-k) = 2\hbar k$.

Consider next the dynamics of electrons near the bottom of the conduction band and holes near the top of the valence band. (We'll assume a 'direct gap', *i.e.* the conduction band minimum is located directly above the valence band maximum, which we take to be at the Brillouin zone center $\boldsymbol{k} = 0$, otherwise known as the Γ point.) Expanding the dispersions about their extrema to second order,

$$\begin{aligned}\varepsilon_{\rm v}(\boldsymbol{k}) &= \varepsilon_0^{\rm v} - \tfrac{1}{2}\hbar^2 (m^{\rm v})^{-1}_{\alpha\beta}\, k^\alpha\, k^\beta \\ \varepsilon_{\rm c}(\boldsymbol{k}) &= \varepsilon_0^{\rm c} + \tfrac{1}{2}\hbar^2 (m^{\rm c})^{-1}_{\alpha\beta}\, k^\alpha\, k^\beta \quad .\end{aligned} \tag{5.236}$$

The velocity is

$$v^\alpha(\boldsymbol{k}) = \frac{1}{\hbar}\frac{\partial\varepsilon}{\partial k^\alpha} = \pm\hbar\, m^{-1}_{\alpha\beta}\, k^\beta \quad , \tag{5.237}$$

where the $+$ sign is used in conjunction with $m^{\rm c}$ and the $-$ sign with $m^{\rm v}$. We compute the acceleration $\boldsymbol{a} = \ddot{\boldsymbol{r}}$ via the chain rule,

$$\begin{aligned}a^\alpha &= \frac{\partial v^\alpha}{\partial k^\beta}\cdot\frac{dk^\beta}{dt} \\ &= \mp e\, m^{-1}_{\alpha\beta}\left[E^\beta + \frac{1}{c}\,(\boldsymbol{v}\times\boldsymbol{B})^\beta\right] \\ F^\alpha &= m_{\alpha\beta}\, a^\beta = \mp e\left[E^\beta + \frac{1}{c}\,(\boldsymbol{v}\times\boldsymbol{B})^\beta\right] \quad .\end{aligned} \tag{5.238}$$

Thus, the hole wavepacket accelerates as if it has charge $+e$ but a *positive* effective mass. Note that we have above presumed a *direct band gap*, *i.e.* that the conduction band minimum lies directly above the valence band maximum, at the same value of $\boldsymbol{k}$ (typically the Γ point in the Brillouin zone). However, many materials have an *indirect band gap* in which case

$$\begin{aligned}\varepsilon_{\rm v}(\boldsymbol{k}) &= \varepsilon_0^{\rm v} - \tfrac{1}{2}\hbar^2 (m^{\rm v})^{-1}_{\alpha\beta}\,(k^\alpha - K^\alpha_{\rm v})\,(k^\beta - K^\beta_{\rm v}) \\ v^\alpha_{\rm v}(\boldsymbol{k}) &= -(m^{\rm v})^{-1}_{\alpha\beta}\,(k^\beta - K^\beta_{\rm v})\end{aligned} \tag{5.239}$$

and

$$\begin{aligned}\varepsilon_{\rm c}(\boldsymbol{k}) &= \varepsilon_0^{\rm c} + \tfrac{1}{2}\hbar^2 (m^{\rm c})^{-1}_{\alpha\beta}\,(k^\alpha - K^\alpha_{\rm c})\,(k^\beta - K^\beta_{\rm c}) \\ v^\alpha_{\rm c}(\boldsymbol{k}) &= +(m^{\rm c})^{-1}_{\alpha\beta}\,(k^\beta - K^\beta_{\rm c}) \quad .\end{aligned} \tag{5.240}$$

5.9.2 Boltzmann equation for holes

Finally, what form does the Boltzmann equation take for holes? Starting with the Boltzmann equation for electrons,

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{r}}\cdot\frac{\partial f}{\partial \boldsymbol{r}} + \dot{\boldsymbol{k}}\cdot\frac{\partial f}{\partial \boldsymbol{k}} = \mathcal{I}_{\boldsymbol{k}}[f] \quad , \tag{5.241}$$

we recast this in terms of the hole distribution $\bar{f} = 1 - f$, and obtain

$$\frac{\partial \bar{f}}{\partial t} + \dot{\boldsymbol{r}} \cdot \frac{\partial \bar{f}}{\partial \boldsymbol{r}} + \dot{\boldsymbol{k}} \cdot \frac{\partial \bar{f}}{\partial \boldsymbol{k}} = -\mathcal{I}_{\boldsymbol{k}}\big[1 - \bar{f}\,\big] \quad . \tag{5.242}$$

This then is the Boltzmann equation for the hole distribution $\bar{f}$. Recall that we can expand the collision integral functional as

$$\mathcal{I}_{\boldsymbol{k}}\big[f^0 + \delta f\big] = \mathcal{L}\,\delta f + \ldots \tag{5.243}$$

where $\mathcal{L}$ is a linear operator, and the higher order terms are formally of order $(\delta f)^2$. Note that the zeroth order term $\mathcal{I}_{\boldsymbol{k}}\big[f^0\big]$ vanishes due to the fact that f^0 represents a local equilibrium. Thus, after writing $\bar{f} = \bar{f}^0 + \delta\bar{f}$

$$-\mathcal{I}_{\boldsymbol{k}}\big[1 - \bar{f}\,\big] = -\mathcal{I}_{\boldsymbol{k}}\big[1 - \bar{f}^0 - \delta\bar{f}\,\big] = \mathcal{L}\,\delta\bar{f} + \ldots \tag{5.244}$$

and the linearized collisionless Boltzmann equation for holes is

$$\frac{\partial \delta\bar{f}}{\partial t} - \frac{e}{\hbar c}\,\boldsymbol{v} \times \boldsymbol{B} \cdot \frac{\partial\,\delta\bar{f}}{\partial \boldsymbol{k}} - \boldsymbol{v} \cdot \left[e\boldsymbol{\mathcal{E}} + \frac{\varepsilon - \mu}{T}\boldsymbol{\nabla} T\right] \frac{\partial \bar{f}^0}{\partial \varepsilon} = \mathcal{L}\,\delta\bar{f} \quad , \tag{5.245}$$

which is of precisely the same form as the electron case in Eqn. (5.177). Note that the local equilibrium distribution for holes is given by

$$\bar{f}^0(\boldsymbol{r}, \boldsymbol{k}, t) = \left\{\exp\left(\frac{\mu(\boldsymbol{r}, t) - \varepsilon(\boldsymbol{k})}{k_{\rm B} T(\boldsymbol{r}, t)}\right) + 1\right\}^{-1} \quad . \tag{5.246}$$

5.10 Magnetoresistance and Hall Effect

5.10.1 Boltzmann theory for $\rho_{\alpha\beta}(\omega, \boldsymbol{B})$

In the presence of an external magnetic field $\boldsymbol{B}$, the linearized Boltzmann equation takes the form[17]

$$\frac{\partial \delta f}{\partial t} - e\,\boldsymbol{v} \cdot \boldsymbol{\mathcal{E}}\,\frac{\partial f^0}{\partial \varepsilon} - \frac{e}{\hbar c}\,\boldsymbol{v} \times \boldsymbol{B} \cdot \frac{\partial \delta f}{\partial \boldsymbol{k}} = \mathcal{L}\,\delta f \quad . \tag{5.247}$$

We will obtain an explicit solution within the relaxation time approximation $\mathcal{L}\,\delta f = -\delta f/\tau$ and the effective mass approximation,

$$\varepsilon(\boldsymbol{k}) = \pm\tfrac{1}{2}\,\hbar^2\, m^{-1}_{\alpha\beta}\, k^\alpha\, k^\beta \quad \Longrightarrow \quad v^\alpha = \pm\,\hbar\, m^{-1}_{\alpha\beta}\, k^\beta \quad , \tag{5.248}$$

where the top sign applies for electrons and the bottom sign for holes. With $\boldsymbol{\mathcal{E}}(t) = \boldsymbol{\mathcal{E}}\, e^{-i\omega t}$, we try a solution of the form

$$\delta f(\boldsymbol{k}, t) = \boldsymbol{k} \cdot \boldsymbol{A}(\varepsilon)\, e^{-i\omega t} \equiv \delta f(\boldsymbol{k})\, e^{-i\omega t} \tag{5.249}$$

[17] For holes, we replace $f^0 \to \bar{f}^0$ and $\delta f \to \delta\bar{f}$.

where $\boldsymbol{A}(\varepsilon)$ is a vector function of ε to be determined. Each component A_α is a function of $\boldsymbol{k}$ through its dependence on $\varepsilon = \varepsilon(\boldsymbol{k})$. We now have

$$(\tau^{-1} - i\omega)\, k^\mu A^\mu - \frac{e}{\hbar c}\, \epsilon_{\alpha\beta\gamma}\, v^\alpha\, B^\beta\, \frac{\partial}{\partial k^\gamma}\, (k^\mu A^\mu) = e\, \boldsymbol{v} \cdot \boldsymbol{\mathcal{E}}\ \frac{\partial f^0}{\partial \varepsilon} \quad , \tag{5.250}$$

where $\epsilon_{\alpha\beta\gamma}$ is the Levi-Civita tensor. Note that

$$\begin{aligned} \epsilon_{\alpha\beta\gamma}\, v^\alpha B^\beta\, \frac{\partial}{\partial k^\gamma}\, (k^\mu A^\mu) &= \epsilon_{\alpha\beta\gamma}\, v^\alpha B^\beta \left(A^\gamma + k^\mu\, \frac{\partial A^\mu}{\partial k^\gamma} \right) \\ &= \epsilon_{\alpha\beta\gamma}\, v^\alpha B^\beta \left(A^\gamma + \hbar\, k^\mu\, v^\gamma\, \frac{\partial A^\mu}{\partial \varepsilon} \right) = \epsilon_{\alpha\beta\gamma}\, v^\alpha B^\beta A^\gamma \quad , \end{aligned} \tag{5.251}$$

owing to the asymmetry of the Levi-Civita tensor: $\epsilon_{\alpha\beta\gamma}\, v^\alpha\, v^\gamma = 0$. We now invoke the identity $\hbar\, k^\alpha = \pm m_{\alpha\beta}\, v^\beta$ and match the coefficients of v^α in each term of the Boltzmann equation. This yields,

$$\left[(\tau^{-1} - i\omega)\, m_{\alpha\beta} \pm \frac{e}{c}\, \epsilon_{\alpha\beta\gamma}\, B^\gamma \right] A^\beta = \pm\, \hbar\, e\, \frac{\partial f^0}{\partial \varepsilon}\, \mathcal{E}^\alpha \quad . \tag{5.252}$$

Defining

$$\Gamma_{\alpha\beta} \equiv (\tau^{-1} - i\omega)\, m_{\alpha\beta} \pm \frac{e}{c}\, \epsilon_{\alpha\beta\gamma}\, B^\gamma \quad , \tag{5.253}$$

we obtain the solution

$$\delta f = \pm e\, v^\alpha\, m_{\alpha\beta}\, \Gamma^{-1}_{\beta\gamma}\, \mathcal{E}^\gamma\, \frac{\partial f^0}{\partial \varepsilon} \quad . \tag{5.254}$$

From this, we can compute the current density and the conductivity tensor. The electrical current density is

$$j^\alpha = \mp 2e \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\ v^\alpha\, \delta f = +2e^2\, \mathcal{E}^\gamma \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\ v^\alpha\, v^\nu\, m_{\nu\beta}\, \Gamma^{-1}_{\beta\gamma}(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \quad , \tag{5.255}$$

where we allow for an energy-dependent relaxation time $\tau(\varepsilon)$. Note that $\Gamma_{\alpha\beta}(\varepsilon)$ is energy-dependent due to its dependence on τ. The conductivity is then

$$\begin{aligned} \sigma_{\alpha\beta}(\omega, \boldsymbol{B}) &= 2\hbar^2 e^2\, m^{-1}_{\alpha\mu} \left\{ \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3}\ k^\mu\, k^\nu \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \Gamma^{-1}_{\nu\beta}(\varepsilon) \right\} \\ &= \tfrac{2}{3}\, e^2 \int\limits_{-\infty}^{\infty} d\varepsilon\, \varepsilon\, g(\varepsilon)\, \Gamma^{-1}_{\alpha\beta}(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \quad , \end{aligned} \tag{5.256}$$

where the chemical potential is measured with respect to the band edge. Thus,

$$\sigma_{\alpha\beta}(\omega, \boldsymbol{B}) = ne^2\, \langle \Gamma^{-1}_{\alpha\beta} \rangle \quad , \tag{5.257}$$

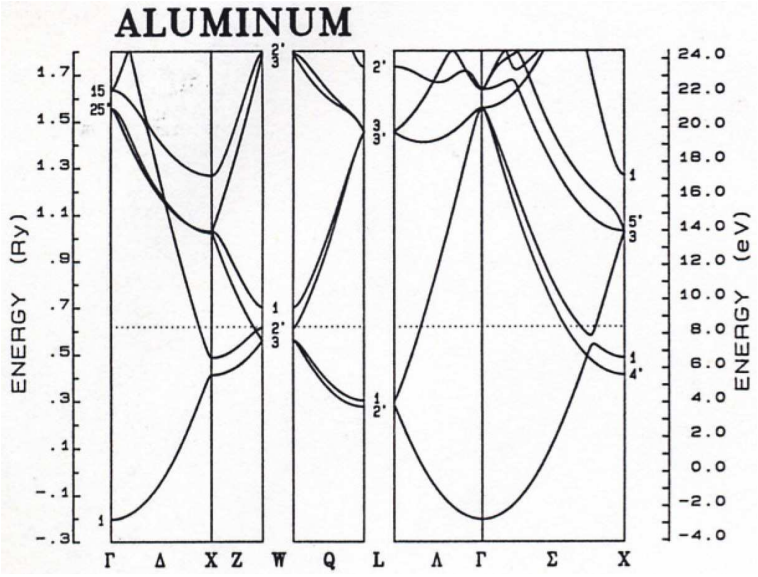


Figure 5.18: Energy bands in aluminum.

where averages denoted by angular brackets are defined by

$$\langle \Gamma^{-1}_{\alpha\beta} \rangle \equiv \frac{\int\limits_{-\infty}^{\infty} d\varepsilon \, \varepsilon \, g(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \Gamma^{-1}_{\alpha\beta}(\varepsilon)}{\int\limits_{-\infty}^{\infty} d\varepsilon \, \varepsilon \, g(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right)} \quad . \tag{5.258}$$

The quantity n is the carrier density,

$$n = \int\limits_{-\infty}^{\infty} d\varepsilon \, g(\varepsilon) \times \begin{cases} f^0(\varepsilon) & \text{(electrons)} \\ \big[1 - f^0(\varepsilon)\big] & \text{(holes)} \end{cases} \tag{5.259}$$

EXERCISE: Verify Eqn. (5.256).

For the sake of simplicity, let us assume an energy-independent scattering time, or that the temperature is sufficiently low that only $\tau(\varepsilon_{\rm F})$ matters, and we denote this scattering time simply by τ. Putting this all together, then, we obtain

$$\begin{aligned} \sigma_{\alpha\beta} &= ne^2 \, \Gamma^{-1}_{\alpha\beta} \\ \rho_{\alpha\beta} &= \frac{1}{ne^2} \Gamma_{\alpha\beta} = \frac{1}{ne^2} \left[(\tau^{-1} - i\omega) m_{\alpha\beta} \pm \frac{e}{c} \, \epsilon_{\alpha\beta\gamma} \, B^\gamma \right] \quad . \end{aligned} \tag{5.260}$$

We thereby recover the results of §5.5.2.

5.10.2 Hall effect in high fields

In the high field limit, one may neglect the collision integral entirely, and write (at $\omega = 0$)

$$-e\,\boldsymbol{v}\cdot\boldsymbol{\mathcal{E}}\,\frac{\partial f^0}{\partial\varepsilon} - \frac{e}{\hbar c}\,\boldsymbol{v}\times\boldsymbol{B}\cdot\frac{\partial\delta f}{d\boldsymbol{k}} = 0 \quad . \tag{5.261}$$

We'll consider the case of electrons, and take $\boldsymbol{\mathcal{E}} = \mathcal{E}\hat{\boldsymbol{y}}$ and $\boldsymbol{B} = B\hat{\boldsymbol{z}}$, in which case the solution is

$$\delta f = \frac{\hbar c\mathcal{E}}{B}\,k_x\,\frac{\partial f^0}{\partial\varepsilon} \quad . \tag{5.262}$$

Note that k_x is not a smooth single-valued function over the Brillouin-zone due to Bloch periodicity. This treatment, then, will make sense only if the derivative $\partial f^0/\partial\varepsilon$ confines $\boldsymbol{k}$ to a closed orbit within the first Brillouin zone. In this case, we have

$$j_x = 2ec\,\frac{\mathcal{E}}{B}\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,k_x\,\frac{\partial\varepsilon}{\partial k_x}\,\frac{\partial f^0}{\partial\varepsilon} = 2ec\,\frac{\mathcal{E}}{B}\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,k_x\,\frac{\partial f^0}{\partial k_x} \quad . \tag{5.263}$$

Now we may integrate by parts, if we assume that f^0 vanishes on the boundary of the Brillouin zone. We obtain

$$j_x = -\frac{2ec\mathcal{E}}{B}\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\;f^0 = -\frac{nec}{B}\,\mathcal{E} \quad . \tag{5.264}$$

We conclude that

$$\sigma_{xy} = -\sigma_{yx} = -\frac{nec}{B} \quad , \tag{5.265}$$

independent of the details of the band structure. "Open orbits" – trajectories along Fermi surfaces which cross Brillouin zone boundaries and return in another zone – pose a subtler problem, and generally lead to a finite, *non-saturating* magnetoresistance. For holes, we have $\bar{f}^0 = 1 - f^0$ and

$$j_x = -\frac{2ec\mathcal{E}}{B}\int\limits_{\hat{\Omega}}\frac{d^3k}{(2\pi)^3}\,k_x\,\frac{\partial\bar{f}^0}{\partial k_x} = +\frac{nec}{B}\,\mathcal{E} \tag{5.266}$$

and $\sigma_{xy} = +nec/B$, where n is the hole density.

We define the *Hall coefficient* $R_{\rm H} = -\rho_{xy}/B$ and the *Hall number*

$$z_{\rm H} \equiv -\frac{1}{n_{\rm ion}ecR_{\rm H}} \quad , \tag{5.267}$$

where $n_{\rm ion}$ is the ion density. For high fields, the off-diagonal elements of both $\rho_{\alpha\beta}$ and $\sigma_{\alpha\beta}$ are negligible, and $\rho_{xy} \approx -1/\sigma_{xy}$. Hence $R_{\rm H} \approx \mp 1/nec$, and $z_{\rm H} \approx \pm n/n_{\rm ion}$. The high field Hall coefficient is used to determine both the carrier density as well as the sign of the charge carriers; $z_{\rm H}$ is a measure of valency.

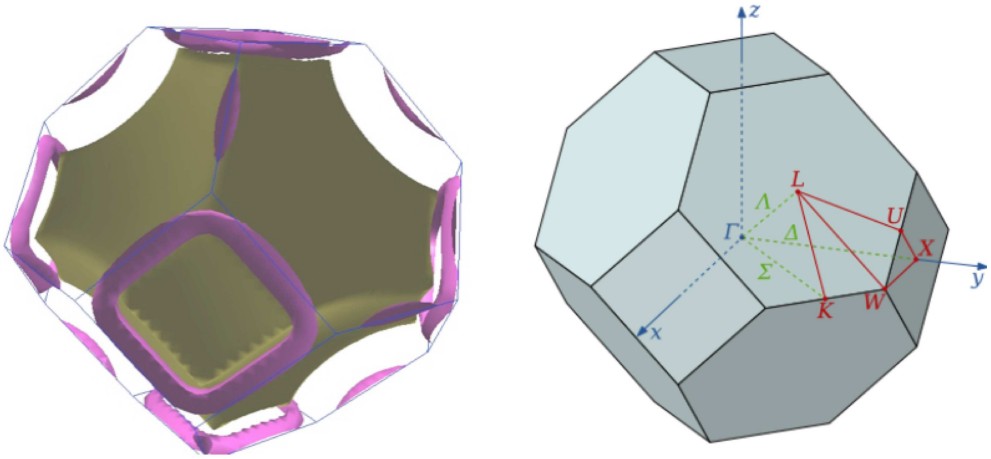


Figure 5.19: Fermi surfaces for electron (pink) and hole (gold) bands in Aluminum.

In Al, the high field Hall coefficient saturates at $z_{\rm H} = -1$. Why is $z_{\rm H}$ negative? As it turns out, aluminum has both electron and hole bands. Its valence is 3; two electrons go into a filled band, leaving one valence electron to split between the electron and hole bands. Thus $n = 3n_{\rm ion}$ The Hall conductivity is $\sigma_{xy} = (n_{\rm h} - n_{\rm e})\, ec/B$. The difference $n_{\rm h} - n_{\rm e}$ is determined by the following argument. The *electron* density in the hole band is $n'_{\rm e} = 2n_{ion} - n_{\rm h}$, *i.e.* the total density of levels in the band (two states per unit cell) minus the number of empty levels in which there are holes. Thus, $n_{\rm h} - n_{\rm e} = 2n_{\rm ion} - (n_{\rm e} + n'_{\rm e}) = n_{\rm ion}$, where we've invoked $n_{\rm e} + n'_{\rm e} = n_{\rm ion}$, since precisely one electron from each ion is shared between the two partially filled bands. Thus, $\sigma_{xy} = n_{\rm ion} ec/B = nec/3B$ and $z_{\rm H} = -1$. At lower fields, $z_{\rm H} = +3$ is observed, which is what one would expect from the free electron model. Interband scattering, which is suppressed at high fields, leads to this result.

5.11 Thermal Transport

5.11.1 Boltzmann theory

Consider a small region of solid with a fixed volume ΔV. The first law of thermodynamics applied to this region gives $T\Delta S = \Delta E - \mu \Delta N$. Dividing by ΔV gives

$$dq \equiv T\, ds = d\varepsilon - \mu\, dn \quad , \tag{5.268}$$

where s is the entropy density, ε is energy density, and n the number density. This can be directly recast as the following relation among current densities:

$$\boldsymbol{j}_q = T\boldsymbol{j}_s = \boldsymbol{j}_\varepsilon - \mu\, \boldsymbol{j}_n \quad , \tag{5.269}$$

where $\boldsymbol{j}_n = \boldsymbol{j}/(-e)$ is the number current density, $\boldsymbol{j}_\varepsilon$ is the energy current density,

$$\boldsymbol{j}_\varepsilon = 2 \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, \varepsilon \, \boldsymbol{v} \, \delta f \quad , \tag{5.270}$$

and $\boldsymbol{j}_s$ is the entropy current density. Accordingly, the thermal (heat) current density $\boldsymbol{j_q}$ is defined as

$$\boldsymbol{j}_q \equiv T \boldsymbol{j}_s = \boldsymbol{j}_\varepsilon + \frac{\mu}{e} \boldsymbol{j} = 2 \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, (\varepsilon - \mu) \, \boldsymbol{v} \, \delta f \quad . \tag{5.271}$$

In the presence of a time-independent temperature gradient and electric field, linearized Boltzmann equation in the relaxation time approximation has the solution

$$\delta f = -\tau(\varepsilon) \, \boldsymbol{v} \cdot \left(e\boldsymbol{\mathcal{E}} + \frac{\varepsilon - \mu}{T} \boldsymbol{\nabla} T \right) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \quad . \tag{5.272}$$

We now consider both the electrical current $\boldsymbol{j}$ as well as the thermal current density $\boldsymbol{j}_q$. One readily obtains

$$\begin{aligned} \boldsymbol{j} &= -2e \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, \boldsymbol{v} \, \delta f \equiv L_{11} \, \boldsymbol{\mathcal{E}} - L_{12} \, \boldsymbol{\nabla} T \\ \boldsymbol{j}_q &= 2 \int\limits_{\hat{\Omega}} \frac{d^3k}{(2\pi)^3} \, (\varepsilon - \mu) \, \boldsymbol{v} \, \delta f \equiv L_{21} \, \boldsymbol{\mathcal{E}} - L_{22} \, \boldsymbol{\nabla} T \end{aligned} \tag{5.273}$$

where the *transport coefficients* L^{11} *etc.* are matrices:

$$\begin{aligned} L_{11}^{\alpha\beta} &= \frac{e^2}{4\pi^3\hbar} \int\limits_{-\infty}^{\infty} d\varepsilon \, \tau(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \int dS_\varepsilon \, \frac{v^\alpha \, v^\beta}{|\boldsymbol{v}|} \\ L_{21}^{\alpha\beta} = T L_{12}^{\alpha\beta} &= -\frac{e}{4\pi^3\hbar} \int\limits_{-\infty}^{\infty} d\varepsilon \, \tau(\varepsilon) \, (\varepsilon - \mu) \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \int dS_\varepsilon \, \frac{v^\alpha \, v^\beta}{|\boldsymbol{v}|} \\ L_{22}^{\alpha\beta} &= \frac{1}{4\pi^3\hbar \, T} \int\limits_{-\infty}^{\infty} d\varepsilon \, \tau(\varepsilon) \, (\varepsilon - \mu)^2 \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \int dS_\varepsilon \, \frac{v^\alpha \, v^\beta}{|\boldsymbol{v}|} \quad . \end{aligned} \tag{5.274}$$

If we define the hierarchy of integral expressions

$$\mathcal{J}_n^{\alpha\beta} \equiv \frac{1}{4\pi^3\hbar} \int\limits_{-\infty}^{\infty} d\varepsilon \, \tau(\varepsilon) \, (\varepsilon - \mu)^n \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \int dS_\varepsilon \, \frac{v^\alpha \, v^\beta}{|\boldsymbol{v}|} \tag{5.275}$$

then we may write

$$L_{11}^{\alpha\beta} = e^2 \mathcal{J}_0^{\alpha\beta} \qquad L_{21}^{\alpha\beta} = T L_{12}^{\alpha\beta} = -e\,\mathcal{J}_1^{\alpha\beta} \qquad L_{22}^{\alpha\beta} = \frac{1}{T}\,\mathcal{J}_2^{\alpha\beta} \quad . \tag{5.276}$$

The linear relations in Eqn. (5.273) may be recast in the following form:

$$\begin{aligned} \boldsymbol{\mathcal{E}} &= \rho\,\boldsymbol{j} + Q\,\boldsymbol{\nabla} T \\ \boldsymbol{j}_q &= \sqcap\,\boldsymbol{j} - \kappa\,\boldsymbol{\nabla} T \quad , \end{aligned} \tag{5.277}$$

where the matrices ρ, Q, $\sqcap$, and κ are given by

$$\begin{aligned} \rho &= L_{11}^{-1} & Q &= L_{11}^{-1}\,L_{12} \\ \sqcap &= L_{21}\,L_{11}^{-1} & \kappa &= L_{22} - L_{21}\,L_{11}^{-1}\,L_{12} \quad , \end{aligned} \tag{5.278}$$

or, in terms of the $\mathcal{J}_n$,

$$\begin{aligned} \rho &= \frac{1}{e^2}\,\mathcal{J}_0^{-1} & Q &= -\frac{1}{e\,T}\,\mathcal{J}_0^{-1}\,\mathcal{J}_1 \\ \sqcap &= -\frac{1}{e}\,\mathcal{J}_1\,\mathcal{J}_0^{-1} & \kappa &= \frac{1}{T}\left(\mathcal{J}_2 - \mathcal{J}_1\,\mathcal{J}_0^{-1}\,\mathcal{J}_1\right) \quad , \end{aligned} \tag{5.279}$$

The names and physical interpretation of these four transport coefficients is as follows:

- ρ is the *resistivity*: $\boldsymbol{\mathcal{E}} = \rho\boldsymbol{j}$ under the condition of zero thermal gradient (*i.e.* $\boldsymbol{\nabla} T = 0$).
- Q is the *thermopower*: $\boldsymbol{\mathcal{E}} = Q\,\boldsymbol{\nabla} T$ under the condition of zero electrical current (*i.e.* $\boldsymbol{j} = 0$). Q is also called the *Seebeck coefficient*.
- $\sqcap$ is the *Peltier coefficient*: $\boldsymbol{j}_q = \sqcap \boldsymbol{j}$ when $\boldsymbol{\nabla} T = 0$.
- κ is the *thermal conductivity*: $\boldsymbol{j}_q = -\kappa\,\boldsymbol{\nabla} T$ when $\boldsymbol{j} = 0$.

One practical way to measure the thermopower is to form a junction between two dissimilar metals, A and B. The junction is held at temperature T_1 and the other ends of the metals are held at temperature T_0. One then measures a voltage difference between the free ends of the metals – this is known as the Seebeck effect. Integrating the electric field from the free end of A to the free end of B gives

$$V_{\rm A} - V_{\rm B} = -\int_{\rm A}^{\rm B} \boldsymbol{\mathcal{E}} \cdot d\boldsymbol{l} = (Q_{\rm B} - Q_{\rm A})(T_1 - T_0) \quad . \tag{5.280}$$

What one measures here is really the difference in thermopowers of the two metals. For an absolute measurement of $Q_{\rm A}$, replace B by a superconductor ($Q = 0$ for a superconductor). A device which converts a temperature gradient into an emf is known as a *thermocouple*.

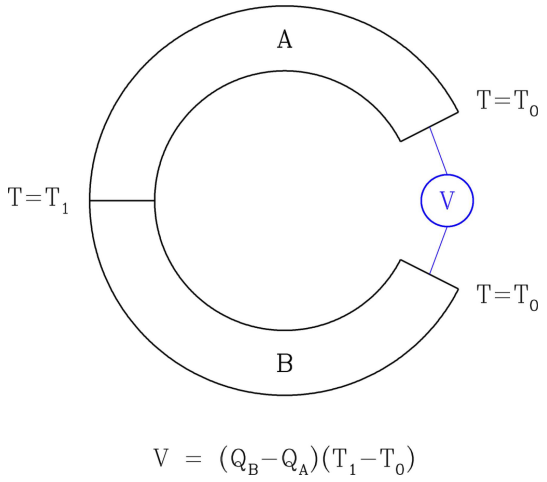


Figure 5.20: A thermocouple is a junction formed of two dissimilar metals. With no electrical current passing, an electric field is generated in the presence of a temperature gradient, resulting in a voltage $V = V_{\rm A} - V_{\rm B}$.

The Peltier effect has practical applications in refrigeration technology. Suppose an electrical current I is passed through a junction between two dissimilar metals, A and B. Due to the difference in Peltier coefficients, there will be a net heat current into the junction of $W = (\Pi_{\rm A} - \Pi_{\rm B})\, I$. Note that this is proportional to I, rather than the familiar I^2 result from Joule heating. The sign of W depends on the direction of the current. If a second junction is added, to make an ABA configuration, then heat absorbed at the first junction will be liberated at the second[18].

5.11.2 The heat equation

We begin with the continuity equations for charge density ρ and energy density ε:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j} = 0 \quad , \qquad \frac{\partial \varepsilon}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{j}_\varepsilon = \boldsymbol{j} \cdot \boldsymbol{E} \quad , \tag{5.281}$$

where $\boldsymbol{E}$ is the electric field[19]. Now we invoke local thermodynamic equilibrium and write

$$\frac{\partial \varepsilon}{\partial t} = \frac{\partial \varepsilon}{\partial n}\frac{\partial n}{\partial t} + \frac{\partial \varepsilon}{\partial T}\frac{\partial T}{\partial t} = -\frac{\mu}{e}\frac{\partial \rho}{\partial t} + c_V \frac{\partial T}{\partial t} \quad , \tag{5.282}$$

[18]To create a refrigerator, stick the cold junction inside a thermally insulated box and the hot junction outside the box.

[19]Note that it is $\boldsymbol{E} \cdot \boldsymbol{j}$ and *not* $\boldsymbol{\mathcal{E}} \cdot \boldsymbol{j}$ which is the source term in the energy continuity equation.

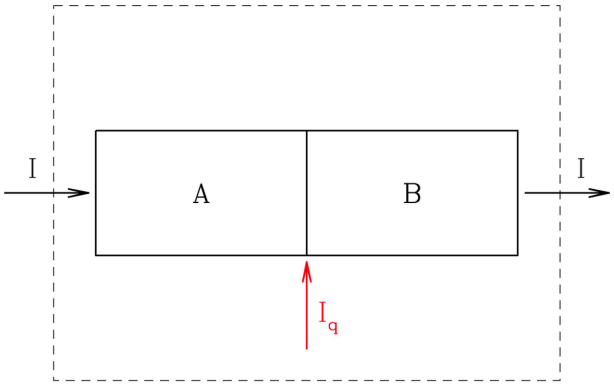


Figure 5.21: A sketch of a Peltier effect refrigerator. An electrical current I is passed through a junction between two dissimilar metals. If the dotted line represents the boundary of a thermally well-insulated body, then the body cools when $\Pi_{\rm B} > \Pi_{\rm A}$, in order to maintain a heat current balance at the junction.

where n is the electron *number density* ($n = -\rho/e$) and c_V is the specific heat. We may now write

$$\begin{aligned} c_V \frac{\partial T}{\partial t} &= \frac{\partial \varepsilon}{\partial t} + \frac{\mu}{e}\frac{\partial \rho}{\partial t} \\ &= \boldsymbol{j}\cdot\boldsymbol{E} - \boldsymbol{\nabla}\cdot\boldsymbol{j}_\varepsilon - \frac{\mu}{e}\boldsymbol{\nabla}\cdot\boldsymbol{j} = \boldsymbol{j}\cdot\boldsymbol{\mathcal{E}} - \boldsymbol{\nabla}\cdot\boldsymbol{j}_q \quad . \end{aligned} \tag{5.283}$$

Invoking $\boldsymbol{j}_q = \Pi\,\boldsymbol{j} - \kappa\,\boldsymbol{\nabla}T$, we see that if there is no electrical current ($\boldsymbol{j} = 0$), we obtain the *heat equation*

$$c_V \frac{\partial T}{\partial t} = \kappa_{\alpha\beta}\,\frac{\partial^2 T}{\partial x^\alpha\,\partial x^\beta} \quad . \tag{5.284}$$

This results in a time scale τ_T for temperature diffusion $\tau_T = \mathcal{C}L^2 c_V/\kappa$, where L is a typical length scale and $\mathcal{C}$ is a numerical constant. For a cube of size L subjected to a sudden external temperature change, L is the side length and $\mathcal{C} = 1/3\pi^2$ (solve by separation of variables).

5.11.3 Calculation of transport coefficients

We will henceforth assume that sufficient crystalline symmetry exists (*e.g.* cubic symmetry) to render all the transport coefficients multiples of the identity matrix. Under such conditions, we may write $\mathcal{J}_n^{\alpha\beta} = \mathcal{J}_n\,\delta_{\alpha\beta}$ with

$$\mathcal{J}_n = \frac{1}{12\pi^3\hbar}\int\limits_{-\infty}^{\infty} d\varepsilon\,\tau(\varepsilon)\,(\varepsilon-\mu)^n\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\int dS_\varepsilon\,|\boldsymbol{v}| \quad . \tag{5.285}$$

The low-temperature behavior is extracted using the Sommerfeld expansion (see §5.3.2),

$$\mathcal{I} \equiv \int\limits_{-\infty}^{\infty} d\varepsilon\, H(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon}\right) = \pi \mathcal{D} \csc(\pi \mathcal{D})\, H(\varepsilon)\,\Big|_{\varepsilon=\mu} = H(\mu) + \frac{\pi^2}{6}(k_{\rm B}T)^2\, H''(\mu) + \ldots \tag{5.286}$$

where $\mathcal{D} \equiv k_{\rm B}T\,\frac{\partial}{\partial \varepsilon}$ is a dimensionless differential operator.[20]

Let us now perform some explicit calculations in the case of a parabolic band with an energy-independent scattering time τ. In this case, one readily finds

$$\mathcal{J}_n = \frac{\sigma_0}{e^2}\,\varepsilon_{\rm F}^{-3/2}\,\pi\mathcal{D}\csc\pi\mathcal{D}\;\varepsilon^{3/2}\,(\varepsilon-\mu)^n\Big|_{\varepsilon=\mu} \quad , \tag{5.287}$$

where $\sigma_0 = ne^2\tau/m^*$. Note that

$$n = \frac{1}{3\pi^2}\left(\frac{2m^*\varepsilon_{\rm F}}{\hbar^2}\right)^{3/2} \tag{5.288}$$

and that $\varepsilon_{\rm F}$ and μ are related by

$$\varepsilon_{\rm F}^{3/2} = \pi\mathcal{D}\csc\pi\mathcal{D}\;\varepsilon^{3/2}\,\Big|_{\varepsilon=\mu} \quad . \tag{5.289}$$

Thus,

$$\mathcal{J}_0 = \frac{\sigma_0}{e^2} \quad , \quad \mathcal{J}_1 = \frac{\sigma_0}{e^2}\frac{\pi^2}{2}\frac{(k_{\rm B}T)^2}{\varepsilon_{\rm F}} + \ldots \quad , \quad \mathcal{J}_2 = \frac{\sigma_0}{e^2}\frac{\pi^2}{3}(k_{\rm B}T)^2 + \ldots \quad , \tag{5.290}$$

from which we obtain the low-T results $\rho = \sigma_0^{-1}$,

$$Q = -\frac{\pi^2}{2}\,\frac{k_{\rm B}^2 T}{e\,\varepsilon_{\rm F}} \quad , \quad \kappa = \frac{\pi^2}{3}\,\frac{n\tau}{m^*}\,k_{\rm B}^2 T \quad , \tag{5.291}$$

and of course $\Pi = TQ$. The predicted universal ratio

$$\frac{\kappa}{\sigma T} = \frac{\pi^2}{3}\,(k_{\rm B}/e)^2 = 2.45\times 10^{-8}\ {\rm V^2\,K^{-2}} \quad , \tag{5.292}$$

is known as the *Wiedemann-Franz law*. Note also that our result for the thermopower is unambiguously negative. In actuality, several nearly free electron metals have positive low-temperature thermopowers (Cs and Li, for example). What went wrong? We have neglected electron-phonon scattering!

[20] Remember that physically the fixed quantities are temperature and total carrier number density (or charge density, in the case of electron and hole bands), and *not* temperature and chemical potential. An equation of state relating n, μ, and T is then inverted to obtain $\mu(n,T)$, so that all results ultimately may be expressed in terms of n and T.

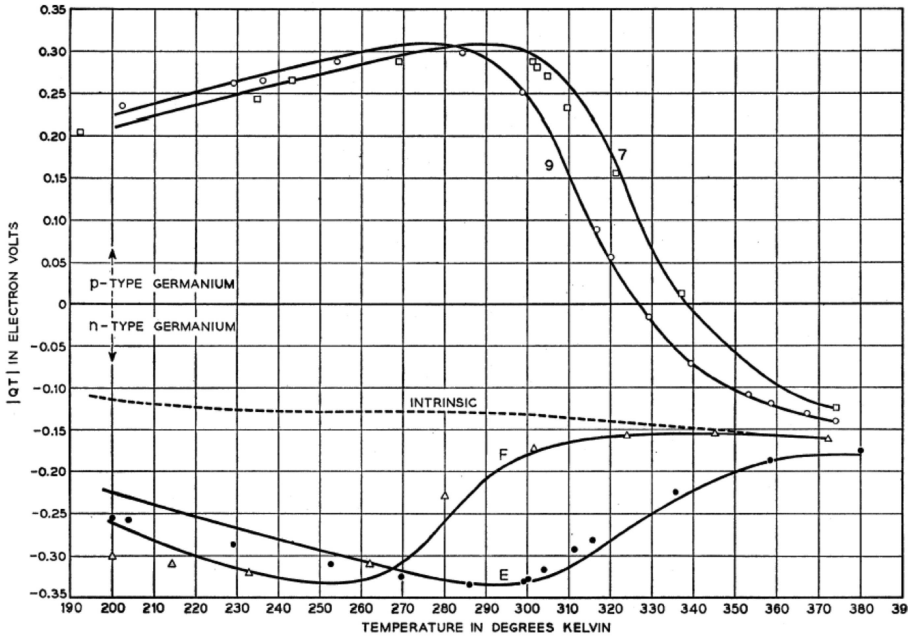


Figure 5.22: QT product for p-type and n-type Ge, from T. H. Geballe and J. W. Hull, *Phys. Rev.* **94**, 1134 (1954). Samples 7, 9, E, and F are distinguished by different doping properties, or by their resistivities at $T = 300\,\mathrm{K}$: $21.5\,\Omega$-cm (7), $34.5\,\Omega$-cm (9), $18.5\,\Omega$-cm (E), and $46.0\,\Omega$-cm (F).

5.11.4 Onsager relations

Transport phenomena are described in general by a set of linear relations,

$$J_i = L_{ik}\,F_k \quad , \tag{5.293}$$

where the $\{F_k\}$ are *generalized forces* and the $\{J_i\}$ are *generalized currents*. Moreover, to each force F_i corresponds a unique conjugate current J_i, such that the rate of internal entropy production is

$$\dot{S} = \sum_i F_i\,J_i \quad \Longrightarrow \quad F_i = \frac{\partial \dot{S}}{\partial J_i} \quad . \tag{5.294}$$

The Onsager relations (also known as *Onsager reciprocity*) states that

$$L_{ik}(\boldsymbol{B}) = \eta_i\,\eta_k\,L_{ki}(-\boldsymbol{B}) \quad , \tag{5.295}$$

where η_i describes the parity of J_i under time reversal:

$$\mathcal{T}\,J_i = \eta_i\,J_i \quad . \tag{5.296}$$

We shall not prove the Onsager relations.

The Onsager relations have some remarkable consequences. For example, they require, for $\boldsymbol{B} = 0$, that the thermal conductivity tensor κ_{ij} of any crystal must be symmetric, independent

of the crystal structure. In general,this result does not follow from considerations of crystalline symmetry. It also requires that for every 'off-diagonal' transport phenomenon, *e.g.* the Seebeck effect, there exists a distinct corresponding phenomenon, *e.g.* the Peltier effect.

For the transport coefficients studied, Onsager reciprocity means that in the presence of an external magnetic field,

$$\begin{aligned} \rho_{\alpha\beta}(\boldsymbol{B}) &= \rho_{\beta\alpha}(-\boldsymbol{B}) \\ \kappa_{\alpha\beta}(\boldsymbol{B}) &= \kappa_{\beta\alpha}(-\boldsymbol{B}) \\ \sqcap_{\alpha\beta}(\boldsymbol{B}) &= T\,Q_{\beta\alpha}(-\boldsymbol{B}) \quad . \end{aligned} \tag{5.297}$$

Let's consider an isotropic system in a weak magnetic field, and expand the transport coefficients to first order in $\boldsymbol{B}$:

$$\begin{aligned} \rho_{\alpha\beta}(\boldsymbol{B}) &= \rho\,\delta_{\alpha\beta} + \nu\,\epsilon_{\alpha\beta\gamma}\,B^\gamma \\ \kappa_{\alpha\beta}(\boldsymbol{B}) &= \kappa\,\delta_{\alpha\beta} + \varpi\,\epsilon_{\alpha\beta\gamma}\,B^\gamma \\ Q_{\alpha\beta}(\boldsymbol{B}) &= Q\,\delta_{\alpha\beta} + \zeta\,\epsilon_{\alpha\beta\gamma}\,B^\gamma \\ \sqcap_{\alpha\beta}(\boldsymbol{B}) &= \sqcap\,\delta_{\alpha\beta} + \theta\,\epsilon_{\alpha\beta\gamma}B^\gamma \quad . \end{aligned} \tag{5.298}$$

Onsager reciprocity requires $\sqcap = T\,Q$ and $\theta = T\,\zeta$. We can now write

$$\begin{aligned} \boldsymbol{\mathcal{E}} &= \rho\,\boldsymbol{j} + \nu\,\boldsymbol{j}\times\boldsymbol{B} + Q\,\boldsymbol{\nabla}T + \zeta\,\boldsymbol{\nabla}T\times\boldsymbol{B} \\ \boldsymbol{j}_q &= \sqcap\,\boldsymbol{j} + \theta\,\boldsymbol{j}\times\boldsymbol{B} - \kappa\,\boldsymbol{\nabla}T - \varpi\,\boldsymbol{\nabla}T\times\boldsymbol{B} \quad . \end{aligned} \tag{5.299}$$

There are several new phenomena lurking!

- Hall Effect ($\frac{\partial T}{\partial x} = \frac{\partial T}{\partial y} = j_y = 0$)
  An electrical current $\boldsymbol{j} = j_x\,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z\,\hat{\boldsymbol{z}}$ yield an electric field $\boldsymbol{\mathcal{E}}$. The Hall coefficient is $R_{\rm H} = \mathcal{E}_y/j_x\,B_z = -\nu$.

- Ettingshausen Effect ($\frac{\partial T}{\partial x} = j_y = j_{q,y} = 0$)
  An electrical current $\boldsymbol{j} = j_x\,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z\,\hat{\boldsymbol{z}}$ yield a temperature gradient $\frac{\partial T}{\partial y}$. The Ettingshausen coefficient is $P = \frac{\partial T}{\partial y}\big/ j_x\,B_z = -\theta/\kappa$.

- Nernst Effect ($j_x = j_y = \frac{\partial T}{\partial y} = 0$)
  A temperature gradient $\boldsymbol{\nabla}T = \frac{\partial T}{\partial x}\,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z\,\hat{\boldsymbol{z}}$ yield an electric field $\boldsymbol{\mathcal{E}}$. The Nernst coefficient is $\Lambda = \mathcal{E}_y\big/\frac{\partial T}{\partial x}\,B_z = -\zeta$.

- Righi-Leduc Effect ($j_x = j_y = \mathcal{E}_y = 0$)
  A temperature gradient $\boldsymbol{\nabla}T = \frac{\partial T}{\partial x}\,\hat{\boldsymbol{x}}$ and a field $\boldsymbol{B} = B_z\,\hat{\boldsymbol{z}}$ yield an orthogonal temperature gradient $\frac{\partial T}{\partial y}$. The Righi-Leduc coefficient is $\mathcal{L} = \frac{\partial T}{\partial y}\big/\frac{\partial T}{\partial x}B_z = \zeta/Q$.

5.12 Electron-Phonon Scattering

5.12.1 Introductory remarks

We begin our discussion by recalling some elementary facts about phonons in solids:

- In a crystal with r atoms per unit cell, there are $3(r-1)$ optical modes and 3 acoustic modes, the latter guaranteed by the breaking of the three generators of space translations. We write the phonon dispersion as $\omega = \omega_\lambda(\boldsymbol{q})$, where $\lambda \in \{1, \ldots, 3r\}$ labels the phonon branch, and $\boldsymbol{q} \in \hat{\Omega}$. If j labels an acoustic mode, $\omega_j(\boldsymbol{q}) = c_j(\hat{\boldsymbol{q}})\, q$ as $\boldsymbol{q} \to 0$.
- Phonons are bosonic particles with zero chemical potential. The equilibrium phonon distribution is
$$n^0_{\boldsymbol{q}\lambda} = \frac{1}{\exp(\hbar\omega_\lambda(\boldsymbol{q})/k_{\rm B}T) - 1} \quad . \tag{5.300}$$
- The maximum phonon frequency is roughly given by the Debye frequency $\omega_{\rm D}$. The Debye temperature $\Theta_{\rm D} = \hbar\omega_{\rm D} \sim 100\,{\rm K}$ – $1000\,{\rm K}$ in most solids.

At high temperatures, equipartition gives $\langle(\delta\boldsymbol{R}_i)^2\rangle \propto k_{\rm B}T$, hence the effective scattering cross-section $\sigma_{\rm tot}$ increases as T, and $\tau \gtrsim 1/n_{\rm ion}v_{\rm F}\sigma_{\rm tot} \propto T^{-1}$. From $\rho = m^*/ne^2\tau$, then, we deduce that the high temperature resistivity should be linear in temperature due to phonon scattering: $\rho(T) \propto T$. Of course, when the mean free path $\ell = v_{\rm F}\tau$ becomes as small as the Fermi wavelength $\lambda_{\rm F}$, the entire notion of coherent quasiparticle transport becomes problematic, and rather than continuing to grow we expect that the resistivity should saturate: $\rho(T \to \infty) \approx h/k_{\rm F}e^2$, known as the *Ioffe-Regel limit*. For $k_{\rm F} = 10^8\,{\rm cm}^{-1}$, this takes the value $260\,\mu\Omega\,{\rm cm}$.

5.12.2 Electron-phonon interaction

Let $\boldsymbol{R}_i = \boldsymbol{R}^0_i + \delta\boldsymbol{R}_i$ denote the position of the $i^{\rm th}$ ion, and let $U(\boldsymbol{r}) = -Ze^2\exp(-r/\lambda_{\rm TF})/r$ be the electron-ion interaction. Expanding in terms of the ionic displacements $\delta\boldsymbol{R}_i$,

$$\mathcal{H}_{\rm el-ion} = \sum_i U(\boldsymbol{r} - \boldsymbol{R}^0_i) - \sum_i \delta\boldsymbol{R}_i \cdot \boldsymbol{\nabla}\, U(\boldsymbol{r} - \boldsymbol{R}^0_i) \quad , \tag{5.301}$$

where i runs from 1 to $N_{\rm ion}$[21]. The deviation $\delta\boldsymbol{R}_i$ may be expanded in terms of the vibrational normal modes of the lattice, *i.e.* the phonons, as

$$\delta R^\alpha_i = \frac{1}{\sqrt{N_{\rm ion}}} \sum_{\boldsymbol{q}\lambda} \left(\frac{\hbar}{2\,\omega_\lambda(\boldsymbol{q})}\right)^{1/2} \hat{\rm e}^\alpha_\lambda(\boldsymbol{q})\, e^{i\boldsymbol{q}\cdot\boldsymbol{R}^0_i}\, (a_{\boldsymbol{q}\lambda} + a^\dagger_{-\boldsymbol{q}\lambda}) \quad . \tag{5.302}$$

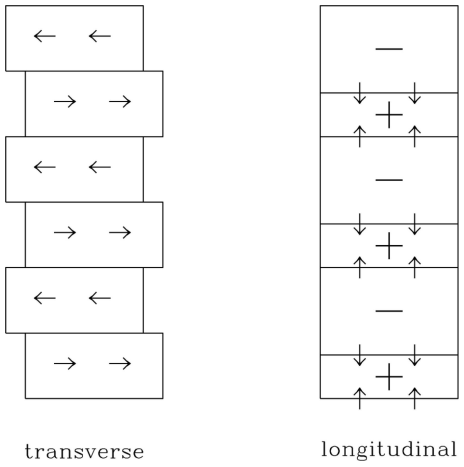


Figure 5.23: Transverse and longitudinal phonon polarizations. Transverse phonons do not result in charge accumulation. Longitudinal phonons create local charge buildup and therefore couple to electronic excitations via the Coulomb interaction.

The phonon polarization vectors satisfy $\hat{\mathbf{e}}_\lambda(\boldsymbol{q}) = \hat{\mathbf{e}}^*_\lambda(-\boldsymbol{q})$ as well as the generalized orthonormality relations

$$\begin{aligned}\sum_\alpha \hat{\mathrm{e}}^\alpha_\lambda(\boldsymbol{q})\,\hat{\mathrm{e}}^\alpha_{\lambda'}(-\boldsymbol{q}) &= M^{-1}\,\delta_{\lambda\lambda'} \\ \sum_\lambda \hat{e}^\alpha_\lambda(\boldsymbol{q})\,\hat{e}^\beta_\lambda(-\boldsymbol{q}) &= M^{-1}\delta_{\alpha\beta} \quad ,\end{aligned} \tag{5.303}$$

where M is the ionic mass. The number of unit cells in the crystal is $N_{\rm ion} = V/\Omega$, where Ω is the Wigner-Seitz cell volume. Again, we approximate Bloch states by plane waves $\psi_{\boldsymbol{k}}(\boldsymbol{r}) = \exp(i\boldsymbol{k}\cdot\boldsymbol{r})/\sqrt{V}$, in which case

$$\left\langle\, \boldsymbol{k}' \,\middle|\, \boldsymbol{\nabla}\, U(\boldsymbol{r} - \boldsymbol{R}^0_i) \,\middle|\, \boldsymbol{k} \,\right\rangle = -\frac{i}{V}\, e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{R}^0_i}\, \frac{4\pi Z e^2\,(\boldsymbol{k}-\boldsymbol{k}')}{(\boldsymbol{k}-\boldsymbol{k}')^2 + \lambda_{\rm TF}^{-2}} \quad . \tag{5.304}$$

The sum over lattice sites gives

$$\sum_{i=1}^{N_{\rm ion}} e^{i(\boldsymbol{k}-\boldsymbol{k}'+\boldsymbol{q})\cdot\boldsymbol{R}^0_i} = N_{\rm ion}\,\delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}\ \mathrm{mod}\ \boldsymbol{G}} \quad , \tag{5.305}$$

so that

$$\mathcal{H}_{\rm el-ph} = \frac{1}{\sqrt{V}} \sum_{\substack{\boldsymbol{k}\boldsymbol{k}'\sigma \\ \boldsymbol{q}\lambda\boldsymbol{G}}} g_\lambda(\boldsymbol{k},\boldsymbol{k}')\,(a^\dagger_{\boldsymbol{q}\lambda} + a_{-\boldsymbol{q}\lambda})\,\psi^\dagger_{\boldsymbol{k}\sigma}\,\psi_{\boldsymbol{k}'\sigma}\,\delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}+\boldsymbol{G}} \quad , \tag{5.306}$$

[21] We assume a Bravais lattice, for simplicity.

Metal	Θ_s	Θ_D	λ_{el-ph}	Metal	Θ_s	Θ_D	λ_{el-ph}
Na	220	150	0.47	Au	310	170	0.08
K	150	100	0.25	Be	1940	1000	0.59
Cu	490	315	0.16	Al	910	394	0.90
Ag	340	215	0.12	In	300	129	1.05

Table 5.3: Electron-phonon interaction parameters for some metals. Temperatures are in Kelvins.

with

$$g_\lambda(\boldsymbol{k}, \boldsymbol{k}+\boldsymbol{q}+\boldsymbol{G}) = -i\left(\frac{\hbar}{2\,\Omega\,\omega_\lambda(\boldsymbol{q})}\right)^{1/2} \frac{4\pi Z e^2}{(\boldsymbol{q}+\boldsymbol{G})^2 + \lambda_{TF}^{-2}}\,(\boldsymbol{q}+\boldsymbol{G})\cdot\hat{\mathbf{e}}^*_\lambda(\boldsymbol{q}) \quad . \tag{5.307}$$

In an isotropic solid[22] ('jellium'), the phonon polarization at wavevector $\boldsymbol{q}$ either is parallel to $\boldsymbol{q}$ (longitudinal waves), or perpendicular to $\boldsymbol{q}$ (transverse waves). We see that only longitudinal waves couple to the electrons. This is because transverse waves do not result in any local accumulation of charge density, and it is to the charge density that electrons couple, via the Coulomb interaction.

Restricting our attention to the longitudinal phonon, we have $\hat{\mathbf{e}}_L(\boldsymbol{q}) = \hat{\boldsymbol{q}}/\sqrt{M}$ and hence, for small $\boldsymbol{q} = \boldsymbol{k}' - \boldsymbol{k}$,

$$g_L(\boldsymbol{k}, \boldsymbol{k}+\boldsymbol{q}) = -i\left(\frac{\hbar}{2M\Omega}\right)^{1/2} \frac{4\pi Z e^2}{q^2 + \lambda_{TF}^{-2}}\, c_L^{-1/2}\, q^{1/2} \quad , \tag{5.308}$$

where c_L is the longitudinal phonon velocity. Thus, for small $\boldsymbol{q}$ we that the electron-longitudinal phonon coupling $g_L(\boldsymbol{k}, \boldsymbol{k}+\boldsymbol{q}) \equiv g_{\boldsymbol{q}}$ satisfies

$$|g_{\boldsymbol{q}}|^2 = \lambda_{el-ph} \cdot \frac{\hbar c_L q}{g(\varepsilon_F)} \quad , \tag{5.309}$$

where $g(\varepsilon_F)$ is the electronic density of states, and where the dimensionless *electron-phonon coupling constant* is

$$\lambda_{el-ph} = \frac{Z^2}{2Mc_L^2\,\Omega g(\varepsilon_F)} = \frac{2Z}{3}\,\frac{m^*}{M}\left(\frac{\varepsilon_F}{k_B\Theta_s}\right)^2 \quad , \tag{5.310}$$

with $\Theta_s \equiv \hbar c_L k_F/k_B$. Table 5.3 lists Θ_s, the Debye temperature Θ_D, and the electron-phonon coupling λ_{el-ph} for various metals.

EXERCISE: Derive Eqn. (12.35).

[22] The jellium model ignores $\boldsymbol{G} \neq 0$ *Umklapp* processes.

5.12.3 Boltzmann equation for electron-phonon scattering

Earlier we had quoted the result for the electron-phonon collision integral,

$$\begin{aligned}\mathcal{I}_{\boldsymbol{k}}[f,n] = \frac{2\pi}{\hbar V}\sum_{\boldsymbol{k}',\lambda}|g_\lambda(\boldsymbol{k},\boldsymbol{k}')|^2\Big\{&(1-f_{\boldsymbol{k}})\,f_{\boldsymbol{k}'}\,(1+n_{\boldsymbol{q},\lambda})\,\delta(\varepsilon_{\boldsymbol{k}}+\hbar\omega_{\boldsymbol{q}\lambda}-\varepsilon_{\boldsymbol{k}'}) \\ &+(1-f_{\boldsymbol{k}})\,f_{\boldsymbol{k}'}\,n_{-\boldsymbol{q}\lambda}\,\delta(\varepsilon_{\boldsymbol{k}}-\hbar\omega_{-\boldsymbol{q}\lambda}-\varepsilon_{\boldsymbol{k}'})-f_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}'})\,(1+n_{-\boldsymbol{q}\lambda})\,\delta(\varepsilon_{\boldsymbol{k}}-\hbar\omega_{-\boldsymbol{q}\lambda}-\varepsilon_{\boldsymbol{k}'}) \\ &-f_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}'})\,n_{\boldsymbol{q}\lambda}\,\delta(\varepsilon_{\boldsymbol{k}}+\hbar\omega_{\boldsymbol{q}\lambda}-\varepsilon_{\boldsymbol{k}'})\Big\}\,\delta_{\boldsymbol{q},\boldsymbol{k}'-\boldsymbol{k}\bmod\boldsymbol{G}}\quad .\end{aligned} \tag{5.311}$$

The four terms inside the curly brackets correspond, respectively, to cases (a) through (d) in Fig. 5.14. The $(1+n)$ factors in the phonon emission terms arise from both spontaneous as well as stimulated emission processes. There is no spontaneous absorption.

EXERCISE: Verify that in equilibrium $\mathcal{I}_{\boldsymbol{k}}\{f^0,n^0\}=0$.

In principle we should also write down a Boltzmann equation for the phonon distribution $n_{\boldsymbol{q}\lambda}$ and solve the two coupled sets of equations. The electronic contribution to the phonon collision integral is written as $\mathcal{J}_{\boldsymbol{q}\lambda}[f,n]$, with

$$\begin{aligned}\mathcal{J}_{\boldsymbol{q}\lambda}[f,n] \equiv \left(\frac{\partial n_{\boldsymbol{q}\lambda}}{\partial t}\right)_{\rm coll} = \frac{4\pi}{\hbar V}\left|g_{\boldsymbol{q}\lambda}\right|^2\sum_{\boldsymbol{k}\in\hat\Omega}\Big\{&(1+n_{\boldsymbol{q}\lambda})\,f_{\boldsymbol{k}+\boldsymbol{q}}\,(1-f_{\boldsymbol{k}}) \\ &-n_{\boldsymbol{q}\lambda}\,f_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}+q})\Big\}\times\delta(\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}-\varepsilon_{\boldsymbol{k}}-\hbar\omega_{\boldsymbol{q}\lambda})\quad .\end{aligned} \tag{5.312}$$

Phonon equilibrium can be achieved via a number of mechanisms we have not considered here, such as impurity or lattice defect scattering, anharmonic effects (*i.e.* phonon-phonon scattering), or grain boundary scattering. At low temperatures,

$$\frac{1}{\tau(\omega)} = \begin{cases} A\,\omega^2 & \text{impurity scattering} \\ B\,\omega^2\,T^3 & \text{anharmonic phonon scattering} \\ C/L & \text{boundary scattering } (L=\text{grain size})\end{cases} \tag{5.313}$$

where A, B, and C are constants.

Of course phonons and electrons scatter from each other – this is the process we are studying – and in principle we should write $f_{\boldsymbol{k}} = f^0_{\boldsymbol{k}} + \delta f_{\boldsymbol{k}}$ and $n_{\boldsymbol{q}\lambda} = n^0_{\boldsymbol{q}\lambda} + \delta n_{\boldsymbol{q}\lambda}$, and linearize the two Boltzmann equations for the electron and phonon distributions in order to study how each species comes to equilibrium. To compute the phonon lifetime due to electron-phonon scattering, we adopt the simplifying assumption that the electrons are in equilibrium at $T=0$

and linearize in $\delta n_{\boldsymbol{q}\lambda}$. This gives a phonon scattering rate of

$$\begin{aligned}\frac{1}{\tau_{\boldsymbol{q}\lambda}} &= \frac{4\pi}{\hbar}\,|g_{\boldsymbol{q}\lambda}|^2\cdot\frac{1}{V}\sum_{\boldsymbol{k}\in\hat\Omega}\big(f^0_{\boldsymbol{k}+\boldsymbol{q}}-f^0_{\boldsymbol{k}}\big)\,\delta(\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}-\varepsilon_{\boldsymbol{k}}-\hbar\omega_{\boldsymbol{q}\lambda}) \\ &= \frac{4\pi}{\hbar^2}\,|g_{\boldsymbol{q}\lambda}|^2\int\limits_{\hat\Omega}\frac{d^3k}{(2\pi)^3}\,\Big[\Theta\big(k_{\rm F}-|\boldsymbol{k}+\boldsymbol{q}|\big)-\Theta\big(k-k_{\rm F}\big)\Big]\delta\Big(\omega_{\boldsymbol{q}\lambda}-\frac{\hbar\boldsymbol{q}^2}{2m^*}-\frac{\hbar\boldsymbol{k}\cdot\boldsymbol{q}}{m^*}\Big) \\ &= \frac{4\pi}{\hbar^2}\,|g_{\boldsymbol{q}\lambda}|^2\,S\big(\boldsymbol{q},\omega_{\boldsymbol{q}\lambda}\big)\quad,\end{aligned}\tag{5.314}$$

where we assume a spherical Fermi surface and isotropic effective mass m^*. Here, $S(\boldsymbol{q},\omega)$ is the *dynamic structure factor* (dsf) of the filled Fermi sphere – we will compute this in detail in chapter three. For now, all we need to know is that

$$S(\boldsymbol{q},\omega)=g(\varepsilon_{\rm F})\,\frac{\pi\omega}{2v_{\rm F}q}\quad\text{for}\quad\omega<v_{\rm F}q\Big(1-\frac{q}{2k_{\rm F}}\Big)\quad.\tag{5.315}$$

We then obtain, for longitudinal acoustic phonons,

$$\frac{1}{\tau_{{\rm L},\boldsymbol{q}}}=2\pi^2\lambda_{\rm el-ph}\,\frac{c_{\rm L}^2}{v_{\rm F}}\,q\quad,\tag{5.316}$$

where $c_{\rm L}$ is the acoustic phonon velocity. Thus, $\tau_{\rm L}^{-1}(\omega)=2\pi^2\lambda_{\rm el-ph}\,(c_{\rm L}/v_{\rm F})\,\omega$.

To compute the electron lifetime due to electron-phonon scattering, we first make the simplifying assumption that the phonons are in equilibrium, *i.e.* $n_{\boldsymbol{q}\lambda}=n^0_{\boldsymbol{q}\lambda}$. We then write $f_{\boldsymbol{k}}=f^0_{\boldsymbol{k}}+\delta f_{\boldsymbol{k}}$ and linearize $\mathcal{I}_{\boldsymbol{k}}[f]$, to obtain

$$\begin{aligned}\mathcal{L}\,\delta f=\frac{2\pi}{\hbar V}\sum_{\boldsymbol{q}\lambda}\big|g_{\boldsymbol{q}\lambda}\big|^2\Big\{&\Big[(1-f^0_{\boldsymbol{k}}+n^0_{\boldsymbol{q}\lambda})\delta f_{\boldsymbol{k}+\boldsymbol{q}}-(f^0_{\boldsymbol{k}+\boldsymbol{q}}+n^0_{\boldsymbol{q}\lambda})\delta f_{\boldsymbol{k}}\Big]\,\delta(\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}-\varepsilon_{\boldsymbol{k}}-\hbar\omega_{\boldsymbol{q}\lambda}) \\ &-\Big[(1-f^0_{\boldsymbol{k}+\boldsymbol{q}}+n^0_{-\boldsymbol{q}\lambda})\delta f_{\boldsymbol{k}}-(f^0_{\boldsymbol{k}}+n^0_{-\boldsymbol{q}\lambda})\delta f_{\boldsymbol{k}+\boldsymbol{q}}\Big]\,\delta(\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}-\varepsilon_{\boldsymbol{k}}+\hbar\omega_{-\boldsymbol{q}\lambda})\Big\}\quad.\end{aligned}\tag{5.317}$$

This integral operator must be inverted in order to solve for $\delta f_{\boldsymbol{k}}$ in

$$\mathcal{L}\,\delta f=e\,\boldsymbol{v}\cdot\boldsymbol{\mathcal{E}}\left(-\frac{\partial f^0}{\partial\varepsilon}\right)\quad.\tag{5.318}$$

Unfortunately, the inversion is analytically intractable – there is no simple solution of the form $\delta f_{\boldsymbol{k}}=e\tau_{\boldsymbol{k}}\boldsymbol{v}_{\boldsymbol{k}}\cdot\boldsymbol{\mathcal{E}}\,(\partial f^0/\partial\varepsilon)$ as there was in the case of isotropic impurity scattering. However, we can still identify the coefficient of $-\delta f_{\boldsymbol{k}}$ in $\mathcal{L}\,\delta f$ as the scattering rate $\tau_{\boldsymbol{k}}^{-1}$. As before, $\tau_{\boldsymbol{k}}$ in fact is a function of the energy $\varepsilon(\boldsymbol{k})$:

$$\begin{aligned}\frac{1}{\tau(\varepsilon)}=\frac{1}{4\pi^2\hbar^2}\int d\varepsilon'\!\int dS_{\varepsilon'}\,\frac{|g_{\boldsymbol{k}'-\boldsymbol{k}}|^2}{|\boldsymbol{v}_{\boldsymbol{k}'}|}\Big\{&\big[f^0(\varepsilon')+n^0_{\boldsymbol{k}'-\boldsymbol{k}}\big]\;\delta(\varepsilon'-\varepsilon-\hbar\omega_{\boldsymbol{k}'-\boldsymbol{k}}) \\ &+\big[1+f^0(\varepsilon')+n^0_{\boldsymbol{k}-\boldsymbol{k}'}\big]\;\delta(\varepsilon'-\varepsilon+\hbar\omega_{\boldsymbol{k}-\boldsymbol{k}'})\Big\}\end{aligned}\tag{5.319}$$

In an isotropic system, $\tau(\varepsilon(\boldsymbol{k}))$ is independent of $\hat{\boldsymbol{k}}$. This means we can take $\boldsymbol{k} = \sqrt{2m^*\varepsilon/\hbar^2}\,\hat{\boldsymbol{z}}$ in performing the above integral.

It is convenient to define the dimensionless function

$$\alpha^2F(\omega) \equiv \frac{1}{8\pi^3\hbar^2}\int dS_{\varepsilon'}\,\frac{|g_{\boldsymbol{k}'-\boldsymbol{k}}|^2}{|\boldsymbol{v}_{\boldsymbol{k}'}|}\,\delta(\omega-\omega_{\boldsymbol{k}'-\boldsymbol{k}}) \quad . \tag{5.320}$$

For parabolic bands, one obtains

$$\begin{aligned}\alpha^2F(\omega) &= \frac{1}{8\pi^3\hbar^2}\frac{\lambda_{\rm el-ph}\,\hbar\omega}{m^*k_{\rm F}/\pi^2\hbar^2}\,\frac{m^*}{\hbar k_{\rm F}}\,k_{\rm F}^2\int d\hat{\boldsymbol{k}}'\,\delta\big(\omega - c_{\rm L}k_{\rm F}|\hat{\boldsymbol{k}}'-\hat{\boldsymbol{z}}|\big) \\ &= \lambda_{\rm el-ph}\left(\frac{\hbar\omega}{k_{\rm B}\Theta_{\rm s}}\right)^2\Theta(2k_{\rm B}\Theta_{\rm s}-\hbar\omega) \quad .\end{aligned} \tag{5.321}$$

The scattering rate is given in terms of $\alpha^2F(\omega)$ as

$$\frac{1}{\tau(\varepsilon)} = 2\pi\int_0^\infty d\omega\,\alpha^2F(\omega)\left\{f^0(\varepsilon+\hbar\omega) - f^0(\varepsilon-\hbar\omega) + 2n^0(\omega) + 1\right\} \quad . \tag{5.322}$$

At $T=0$ we have $f^0(\varepsilon) = \Theta(\varepsilon_{\rm F}-\varepsilon)$ and $n^0(\omega)=0$, whence

$$\begin{aligned}\frac{1}{\tau(\varepsilon)} &= 2\pi\int_0^\infty d\omega\,\alpha^2F(\omega)\,\left\{\Theta(\varepsilon_{\rm F}-\varepsilon-\hbar\omega) - \Theta(\varepsilon_{\rm F}-\varepsilon+\hbar\omega) + 1\right\} \\ &= \begin{cases} \frac{\lambda_{\rm el-ph}}{12}\,\frac{2\pi}{\hbar}\cdot\frac{|\varepsilon-\varepsilon_{\rm F}|^3}{(k_{\rm B}\Theta_{\rm s})^2} & \text{if } |\varepsilon-\varepsilon_{\rm F}| < 2k_{\rm B}\Theta_{\rm s} \\ \frac{2\lambda_{\rm el-ph}}{3}\,\frac{2\pi}{\hbar}\cdot(k_{\rm B}\Theta_{\rm s}) & \text{it } |\varepsilon-\varepsilon_{\rm F}| > 2k_{\rm B}\Theta_{\rm s} \quad .\end{cases}\end{aligned} \tag{5.323}$$

Note that $\tau(\varepsilon_{\rm F}) = \infty$, unlike the case of impurity scattering. This is because at $T=0$ there are no phonons! For $T\neq 0$, the divergence is cut off, and one obtains

$$\frac{1}{\tau(\mu)} = \frac{2\pi\lambda_{\rm el-ph}}{\hbar}\,\frac{k_{\rm B}T^3}{\Theta_{\rm s}^2}\,G\!\left(\frac{2\Theta_{\rm s}}{T}\right) \tag{5.324}$$

with

$$G(y) = \int_0^y dx\,\frac{x^2}{2\sinh x} = \begin{cases} \frac{7}{4}\,\zeta(3) & \text{if } y=\infty \\ \frac{1}{4}\,y & \text{if } y\ll 1 \quad .\end{cases} \tag{5.325}$$

Thus,

$$\frac{1}{\tau(\mu)} = \begin{cases} \frac{7\pi\zeta(3)}{2\hbar}\,\frac{k_{\rm B}T^3}{\Theta_{\rm s}^2}\,\lambda_{\rm el-ph} & \text{if } T\ll\Theta_{\rm s} \\ \frac{2\pi}{\hbar}\,k_{\rm B}T\,\lambda_{\rm el-ph} & \text{if } T\gg\Theta_{\rm s} \quad .\end{cases} \tag{5.326}$$

This calculation predicts that $\tau \propto T^{-3}$ at low temperatures. This is correct if τ is the *thermal lifetime*. However, a more sophisticated calculation shows that the *transport lifetime* behaves as $\tau_{\rm tr} \propto T^{-5}$ at low T. The origin of the discrepancy is our neglect of the $(1-\cos\vartheta)$ factor present in the average of the momentum relaxation time. At low T, there is only small angle scattering from the phonons, and $\langle\vartheta^2\rangle \propto \langle \boldsymbol{q}^2/k_{\rm F}^2\rangle \propto T^2$. The Wiedemann-Franz law, $\tau_\sigma = \tau_\kappa$, is valid for $k_{\rm B}T \gtrsim \hbar c_{\rm L} k_{\rm F}$, as well as at low T in isotropic systems, where impurity scattering is the dominant mechanism. It fails at intermediate temperatures.

Chapter 6

Semiconductors and Insulators

6.1 Introduction

The Bloch energy band structure of noninteracting electrons in a periodic potential leads us to a broad classification of crystalline solids: (i) *metals*, in which the density of states $g(\varepsilon_{\text{F}})$ at the Fermi level is nonzero, and (ii) *insulators*, where $g(\varepsilon_{\text{F}}) = 0$. In an insulator, each Bloch band is either completely filled, corresponding to two electrons per unit cell, or completely empty. Thus, band insulators necessarily have an even number of electrons per unit cell.

6.1.1 Band gaps and transport

In the presence of a uniform electric field $\boldsymbol{E}$ at $T = 0$, a metal respond by generating a current density $\boldsymbol{j} = \sigma \boldsymbol{E}$, where σ is the conductivity matrix. In isotropic systems[1], $\sigma = ne^2\tau/m^*$, as we have seen. In an insulator, there is an energy gap, which is typically taken to mean $\sigma(T = 0, E) = 0$. In fact, this is not quite right, because there can be quantum tunneling between valence and conduction bands at finite E, a process known as *Zener tunneling*. For a direct gap between isotropic valence and conduction bands, the tunneling rate $T = 0$, *i.e.* the number density of electrons tunneling from valence to conduction band per unit time, is found to be[2]

$$\gamma(T = 0, E) = \frac{e^2 E^2 m_{\text{r}}^{1/2}}{18\pi\hbar^2 \Delta^{1/2}} \exp\left(- \frac{\pi m_{\text{r}}^{1/2} \Delta^{3/2}}{2\hbar e E} \right) \tag{6.1}$$

where $m_r = m_{\text{c}}^* m_{\text{v}}^* / (m_{\text{c}}^* + m_{\text{v}}^*)$ is the reduced mass of the valence holes and conduction electrons. Note that the current response is highly nonlinear, and furthermore that the exponential factor

[1] With regard to tensors of rank two like $\sigma_{\alpha\beta}$, cubic symmetry is sufficient in order that the conductivity tensor be a multiple of the unit matrix.

[2] See E. O. Kane, *J. Phys. Chem. Solids* **12**, 181 (1959).

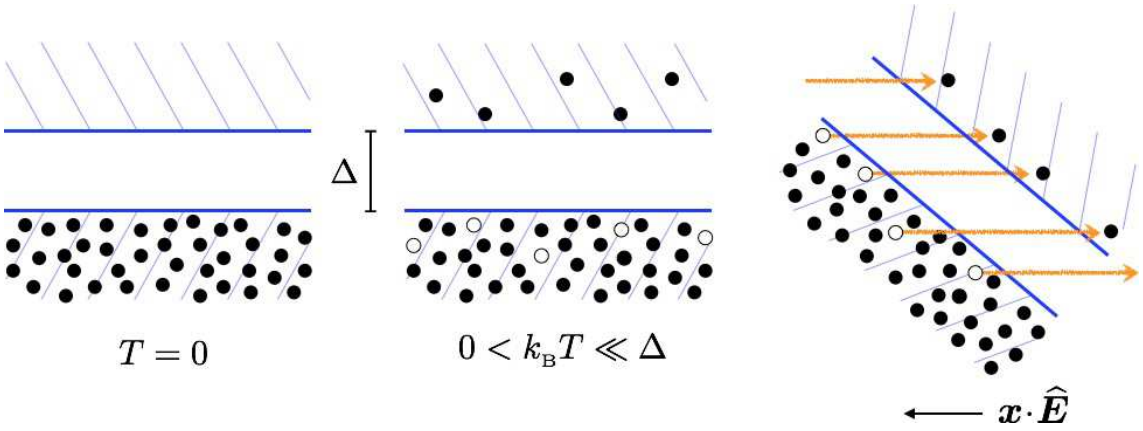


Figure 6.1: Schematics of electron occupation in semiconductor valence and conduction bands. Left: $T = 0$. Right: $0 < k_{\rm B}T \ll \Delta$. Right: Exaggerated schematic of Zener tunneling ($T = 0$).

$\exp(-E_0/E)$, with $E_0 \equiv \pi m_{\rm r}^{1/2}\Delta^{3/2}/2e\hbar$, overwhelms the E^2 prefactor in the $E \to 0$ limit. Here,

$$E_0 = 5.7 \times 10^7 \, \frac{\rm V}{\rm cm} \times \big(\Delta[{\rm eV}]\big)^{3/2} \sqrt{\frac{m_{\rm r}}{m_{\rm e}}} \quad . \tag{6.2}$$

The current density is then $j = e\gamma d$, where d is the thickness over which the field extends. At the level of *linear response*, though, the $T = 0$ conductivity of all insulators is zero.

At finite temperature, due to thermal fluctuations there is a finite electron density $n_{\rm c}$ in the conduction band, and a finite hole density $p_{\rm v}$ in the valence band, with, as we shall see, $n_{\rm c}(T) = p_{\rm v}(T) \propto \exp(-\Delta/2k_{\rm B}T)$. The conductivity is

$$\sigma(T) = \frac{n_{\rm c}(T)\, e^2\tau_{\rm c}}{m_{\rm c}^*} + \frac{p_{\rm v}(T)\, e^2\tau_{\rm v}}{m_{\rm v}^*} \propto e^{-\Delta/2k_{\rm B}T} \quad . \tag{6.3}$$

At $T = 300\,{\rm K}$, we have $k_{\rm B}T = 0.0258\,{\rm eV}$, so for an insulator like carbon diamond, for which $\Delta_{\rm Si} = 5.47\,{\rm eV}$ (indirect gap), $\Delta_{\rm Si}/2k_{\rm B}T = 106$, and the room temperature conductivity is essentially zero. Germanium, however, has a gap of $\Delta_{\rm Ge} = 0.66\,{\rm eV}$ (also indirect), hence $\Delta_{\rm Ge}/2k_{\rm B}T = 12.8$, and the Boltzmann weight is not nearly as small. You should know that the energy gap varies with temperature, mostly because anharmonic lattice vibrations cause the lattice to expand and the atomic positions to fluctuate. Typically one has $\Delta(T) \simeq \Delta(0) - ak_{\rm B}T$ with $a \approx 5$. Band gaps are also pressure-dependent, with $\Delta(p) \simeq \Delta(0) + bp$, with $b \approx 7 \times 10^{-9}\,{\rm eV/cm^2\,g}$. While there is no sharp distinction between semiconductors and insulators, at room temperature one typically classifies solids according to their conductivity, *viz.*

$$\begin{aligned} {\rm metals}: &\ \ \rho(300\,{\rm K}) \lesssim 10^{-6}\,\Omega\cdot{\rm cm} \\ {\rm semiconductors}: &\ \ \rho(300\,{\rm K}) \in \big[10^{-3}\,\Omega\cdot{\rm cm}\,,\ 10^{9}\,\Omega\cdot{\rm cm}\big] \\ {\rm insulators}: &\ \ \rho(300\,{\rm K}) \gtrsim 10^{12}\,\Omega\cdot{\rm cm} \quad . \end{aligned}$$

group	formula	Δ (eV)	gap	m_c^*/m_e	m_v^*/m_e	ϵ	lattice const. (Å)	type
IV	C	5.47	I	0.2	0.25	5.7	3.567	D
IV	Si	1.12	I	1.64(l)/0.082(t)	0.1,(l)/0.49(t)	11.9	5.431	D
IV	Ge	0.66	I	0.9,(l)/0.19(t)	0.04(l)/0.28(t)	16.0	5.646	D
IV−IV	SiC	3.00	I	0.60	1.00	9.66	$3.09(a)/15.1(c)$	W
III−V	AlAs	2.36	I	0.11	0.22	10.1	5.661	Z
III−V	AlP	2.42	I	0.212	0.145	9.8	5.464	Z
III−V	AlSb	1.58	I	0.12	0.98	14.4	6.136	Z
III−V	GaAs	1.42	D	0.063	0.076(lh)/0.5(hh)	12.9	5.653	Z
III−V	GaN	3.44	D	0.27	0.8	10.4	$3.19(a)/10.4(c)$	W
III−V	GaP	2.26	I	0.82	0.60	11.1	5.451	Z
III−V	GaSb	0.72	D	0.042	0.40	15.7	6.096	Z
III−V	InAs	0.36	D	0.023	0.40	15.1	6.058	Z
III−V	InP	1.35	D	0.077	0.64	12.6	5.869	Z
III−V	InSb	0.17	D	0.0145	0.40	16.8	6.479	Z
II−VI	CdS	2.5	D	0.14	0.51	5.4	5.825	Z
II−VI	CdS	2.49	D	0.20	0.7	9.1	$4.14(a)/7.71(c)$	W
II−VI	CdSe	1.70	D	0.13	0.45	10.0	6.050	Z
II−VI	ZnS	3.66	D	0.39	0.23	8.4	5.410	Z
II−VI	ZnS	3.78	D	0.287	0.49	9.6	$3.82(a)/6.26(c)$	W
IV−VI	PbS	041	I	0.25	0.25	17.0	5.936	R
IV−VI	PbTe	0.31	I	0.17	0.20	30.0	6.462	R

Table 6.1: Common semiconductors and their properties at $T = 3000\,\mathrm{K}$. Gap types: D (direct) and I (indirect). Structure: D (Diamond), W (Wurtzite), Z (Zincblende), R (Rocksalt). Hole masses: hh (heavy hole), lh (light hole). Source: S. M. Sze, *Physics of Semiconductors*.

The resistivity of metals and semiconductors depends on the scattering mechanisms which are responsible for momentum relaxation among the charge carriers.

Most semiconductors are covalently bonded crystals coming from column IV of the periodic table (*e.g.*, elemental semiconductors Si, Ge, and grey Sn), or compounds such as III-V materials (GaAs, GaP, InS, InP, GaSb, AlSb, *etc.*) and II-VI materials (PbS, PbSe, SnTe, *etc.*).

6.1.2 Hall effect

High field Hall effect measurements, which give $\sigma_{xy} = (p_v - n_c)\, ec/B$, may be used to obtain an independent measurement of the carrier concentration without requiring knowledge of the

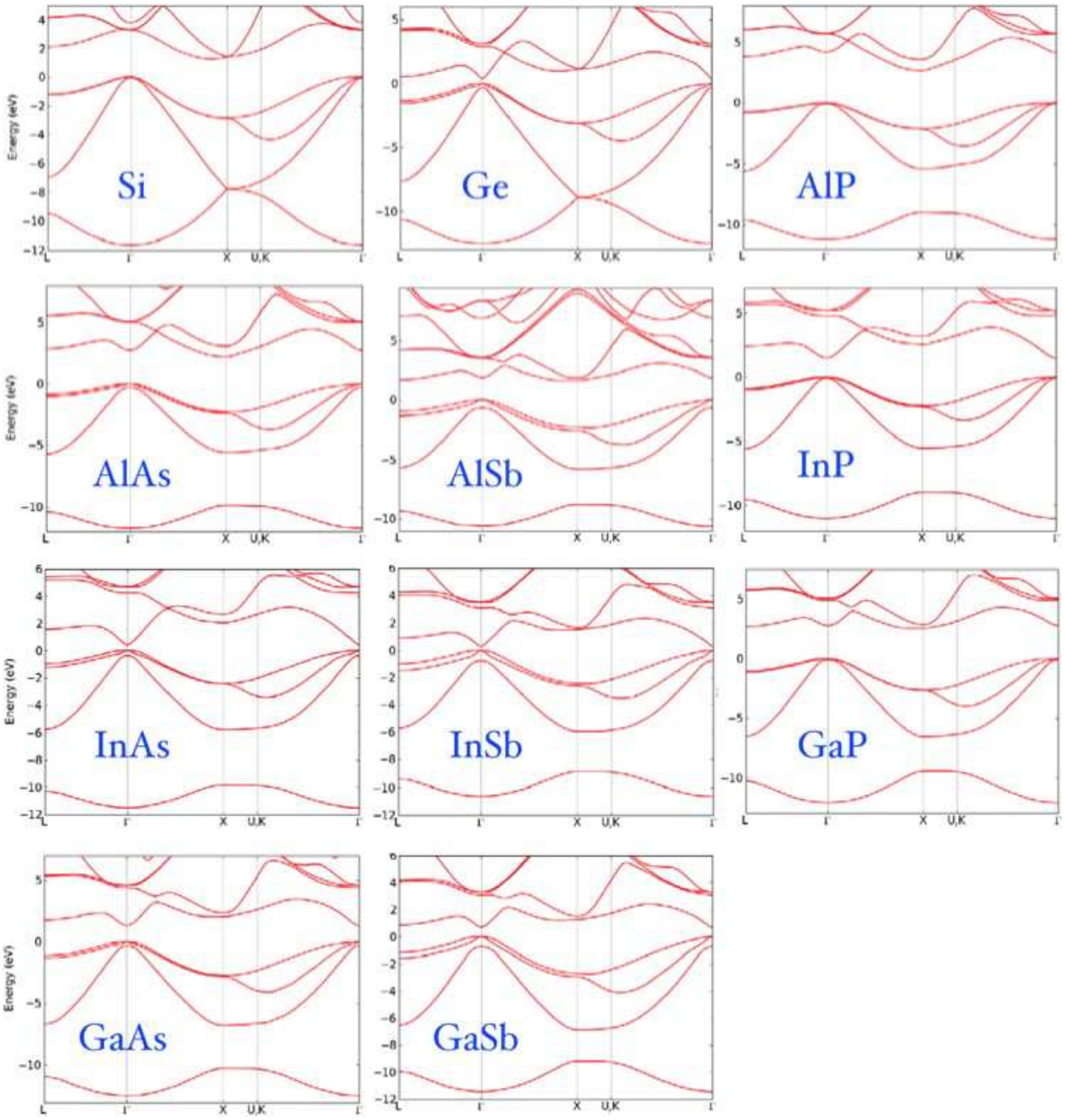
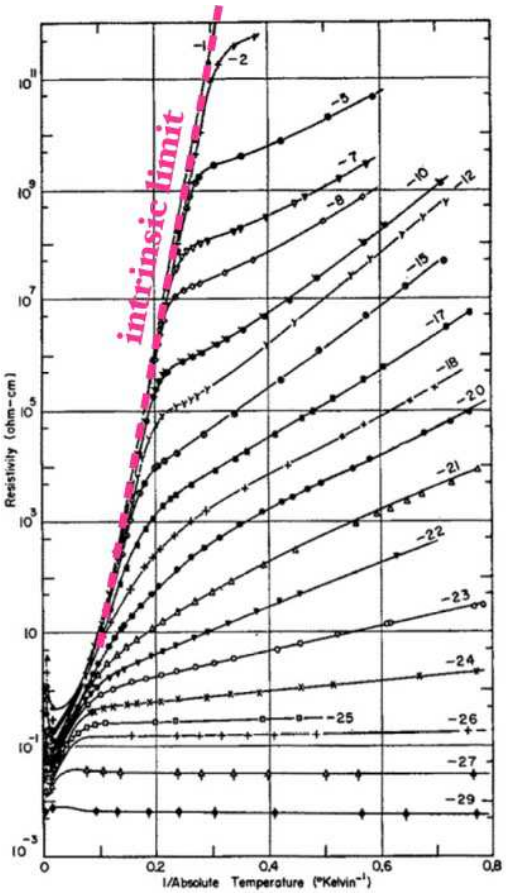


Figure 6.2: Pseudopotential calculation of band structures of diamond and zincblende semiconductors, with spin-orbit effects included. From B. D. Malone and M. L. Cohen, *J. Phys. Condens. Matter* **25**, 105503 (2013).

scattering times $\tau_{\rm v}$ and $\tau_{\rm c}$, which appear in the diagonal conductivity σ_{xx}. Of course, for a pure (*i.e. intrinsic*) semiconductor, $n_{\rm c} = p_{\rm v}$, but in the *extrinsic* case, impurities (*i.e. dopants*) lead to the condition $n_{\rm c} \neq p_{\rm v}$, as we shall see. Such independent measurements of carrier



Sample code number	N_D (cm^{-3})
−1	$5{\cdot}3\times10^{14}$
−2	$9{\cdot}3\times10^{14}$
−5	$1{\cdot}6\times10^{15}$
−7	$2{\cdot}3\times10^{15}$
−8	$3{\cdot}0\times10^{15}$
−10	$5{\cdot}2\times10^{15}$
−12	$8{\cdot}5\times10^{15}$
−15	$1{\cdot}3\times10^{16}$
−17	$2{\cdot}4\times10^{16}$
−18	$3{\cdot}5\times10^{16}$
−20	$4{\cdot}5\times10^{16}$
−21	$5{\cdot}5\times10^{16}$
−22	$6{\cdot}4\times10^{16}$
−23	$7{\cdot}4\times10^{16}$
−24	$8{\cdot}4\times10^{16}$
−25	$1{\cdot}2\times10^{17}$
−26	$1{\cdot}3\times10^{17}$
−27	$2{\cdot}7\times10^{17}$
−29	$9{\cdot}5\times10^{17}$

Figure 6.3: Resistivity of antimony-doped germanium as a function of $1/T$ for varying impurity concentrations. Table at right correlates donor density N_D with sample code number. The magenta "intrinsic limit" line corresponds to the limit $N_D \to 0$. Data from H. J. Fritzsche, *J. Phys. Chem. Solids* **6**, 69 (1958).

concentration confirm that the rapid changes in $\sigma_{xx}(T)$ are predominantly due to variations in carrier concentration.

6.1.3 Optical absorption

The energy gap Δ in a semiconductor may be measured by the temperature dependence of $\ln \sigma_{xx}(T)$, but more directly by optical absorption. When a photon of energy $\hbar\omega > \Delta$ is absorbed by a semiconductor, it creates an conduction electron - valence hole pair, as depicted in Fig. 6.4. At the simplest level of description, the absorption edge coincides with the band gap. At a greater level of refinement, the Coulomb interaction between conduction electron and valence hole must be accounted for, and results in structure to the absorption curve below the

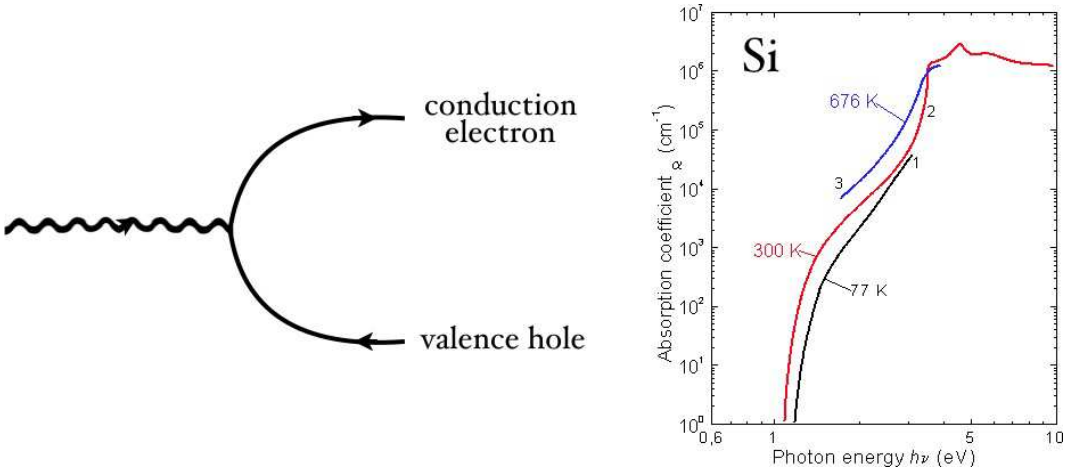


Figure 6.4: Left: A photon with $\hbar\omega > \Delta$ creating an electron-hole pair. Right: Optical conductivity of Si ($\Delta(0\,\mathrm{K}) = 1.17\,\mathrm{eV}$).

Δ threshold. Since $\hbar c = 1973\,\mathrm{eV}\cdot\mathrm{\AA}$, the wavelength of light at the band gap energy is

$$\lambda = \frac{2\pi\hbar c}{\Delta} = \frac{12400\,\mathrm{\AA}}{\Delta[\mathrm{eV}]} \quad . \tag{6.4}$$

Since both energy and momentum must be conserved[3], we have separately $\boldsymbol{k}_\gamma = \boldsymbol{k}_\mathrm{e} - \boldsymbol{k}_\mathrm{h}$ as well as $\hbar c k_\gamma = \Delta$. Thus, $|\boldsymbol{k}_\mathrm{e} - \boldsymbol{k}_\mathrm{h}| = \Delta/\hbar c = \Delta[\mathrm{eV}]\,/\,1973\,\mathrm{eV}\cdot\mathrm{\AA}$. Typically Δ is on the order of eV, hence the difference in electron and hole wavevectors is on the order of a milli-Ångstrom, which is insignificant on the scale of the Brillouin zone. Thus, *optical transitions are vertical*, meaning they involve no change in wavevector for the electron as it is excited from valence to conduction band. The reason is that the speed of light is very big.

Under a constant flux of light, the carrier density $n_\mathrm{c} = p_\mathrm{v} \equiv n$ obeys

$$\frac{dn}{dt} = \alpha - \beta n^2 \quad . \tag{6.5}$$

The first term accounts for the creation of $\mathrm{e}-\mathrm{h}$ pairs due to photoexcitation. The second term accounts for the *recombination* of photoexcited $\mathrm{e}-\mathrm{h}$ pairs. In equilibrium, $n = (\alpha/\beta)^{1/2}$.

6.1.4 Direct *versus* indirect gaps

Fig. 6.5 shows the cases of *direct* and *indirect* gap semiconductors. In a direct gap material, the conduction band minimum and the valence band maximum occur at the same point (or points)

[3]In fact, only *crystal momentum* must be conserved, meaning the wavevector $\boldsymbol{k}$ is conserved modulo any reciprocal lattice vector.

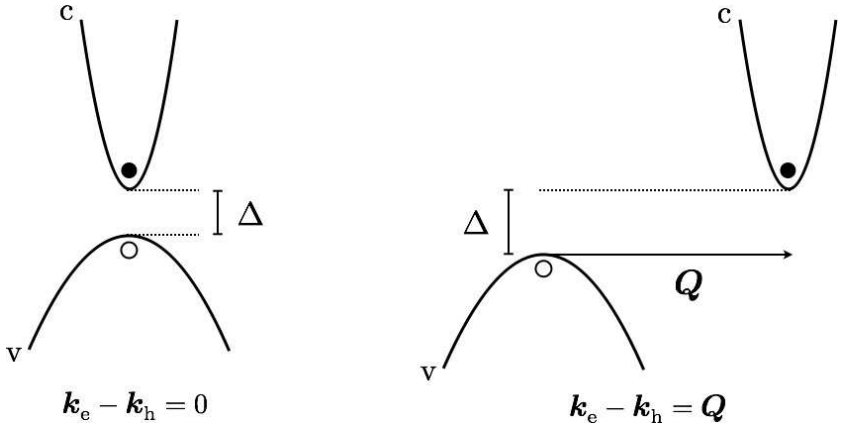


Figure 6.5: Left: A direct gap semiconductor with a low-energy particle-hole pair. The net crystal momentum is $\boldsymbol{K} = \boldsymbol{k}_{\rm e} - \boldsymbol{k}_{\rm h} = 0$. Right: An indirect gap semiconductor with a low-energy particle-hole pair. The net crystal momentum is $\boldsymbol{K} = \boldsymbol{k}_{\rm e} - \boldsymbol{k}_{\rm h} = \boldsymbol{Q}$, where $\boldsymbol{Q}$ is the wavevector difference between valence band maximum and conduction band minimum.

in the Brillouin zone. In an indirect gap material, there is a difference $\boldsymbol{Q} = \boldsymbol{k}_{\rm c}^{\rm min} - \boldsymbol{k}_{\rm v}^{\rm max}$ (modulo $\boldsymbol{G}$). Examples of direct gap materials include $\alpha-$Sn, Se, Te, GaAs, and ZnS. Examples of indirect gap materials include Si, Ge, AlSb, GaP, and PbTe. In an indirect gap material, something other then the photon must supply the missing momentum $\hbar\boldsymbol{Q}$ in order for the material to absorb light at the band gap, and that something is usually a phonon (*i.e.* a lattice vibration). Since phonon frequencies are on the order of meV (Debye temperatures hundreds of Kelvins, $k_{\rm B} = 86.2\,\mu{\rm eV/K}$), the additional phonon energy is small compared with Δ, except perhaps in narrow gap materials. Since the number of phonons vanishes as T^3 at low temperatures, the optical absorption at $\hbar\omega = \Delta$ will be temperature dependent.

6.1.5 Mobility

Mobility μ is defined by the relation of the diffusional drift velocity to the applied electric field strength:

$$\boldsymbol{v}_{\rm drift} = \mp\mu\boldsymbol{E} \quad , \tag{6.6}$$

with the upper sign holding for electrons and the lower sign for holes. The current density is $\boldsymbol{j} = nq\boldsymbol{v}_{\rm drift} = n|q|\mu\boldsymbol{E}$, where q is the charge. Hence we have $\sigma = n|q|\mu$. When both signs of carriers are present,

$$\sigma = n_{\rm c}\, e\, \mu_{\rm c} + p_{\rm v}\, e\, \mu_{\rm v} \quad . \tag{6.7}$$

6.1.6 Effective mass

Generally speaking, the dispersion in the vicinity of a local extremum in the conduction (maximum) or valence (minimum) band dispersions obeys

$$\begin{aligned} E_{\rm c}(\boldsymbol{k}) &= E_{\rm c}^0 + \tfrac{1}{2}\hbar^2\,(m_{\rm c}^*)^{-1}_{\alpha\beta}\,(k^\alpha - Q_{\rm c}^\alpha)(k^\beta - Q_{\rm c}^\beta) + \ldots \\ E_{\rm v}(\boldsymbol{k}) &= E_{\rm v}^0 - \tfrac{1}{2}\hbar^2\,(m_{\rm v}^*)^{-1}_{\alpha\beta}\,(k^\alpha - Q_{\rm v}^\alpha)(k^\beta - Q_{\rm v}^\beta) + \ldots \quad , \end{aligned} \tag{6.8}$$

where $\boldsymbol{Q}_{\rm c,v}$ is the wavevector of the extremum, and $m^*_{\rm c,v}$ is the effective mass tensor. The band gap is given by $\Delta = E_{\rm c}^0 - E_{\rm v}^0$.

Since the effective mass tensors are each symmetric, they may be diagonalized along their principal axes, in which case we may write

$$E(\boldsymbol{k}) = E^0 \pm \left(\frac{\hbar^2(\Delta k_1)^2}{2m_1^*} + \frac{\hbar^2(\Delta k_2)^2}{2m_2^*} + \frac{\hbar^2(\Delta k_3)^2}{2m_3^*}\right) + \mathcal{O}\big((\Delta k)^4\big) \quad , \tag{6.9}$$

where $\Delta\boldsymbol{k} = \boldsymbol{k} - \boldsymbol{Q}$. *Nota bene* : the tensors $m_{\rm c}^*$ and $m_{\rm v}^*$ may not commute, in which case their principal axes do not coincide, and they may not both be rendered diagonal in the same basis. Furthermore, since the extrema of $E_{\rm c,v}(\boldsymbol{k})$ may not lie at the Γ point, the level sets of $E_{\rm c,v}(\boldsymbol{k})$ may not be connected, *i.e.* they may consist of several disjoint components. Indeed this is the case in silicon, where the six equivalent conduction band minima lie along the ΓX directions ($\langle 100\rangle$). In germanium, the conduction band minima occur at the fourfold degenerate L point. Oftentimes, as in the cases of Si, CdTe, InSb, and several other materials, there are more than one electron or hole bands in play.

6.2 Number of Carriers in Thermal Equilibrium

We define

$$\begin{aligned} n_{\rm c}(T,\mu) &= \text{number density of electrons in the conduction band} \\ p_{\rm v}(T,\mu) &= \text{number density of holes in the valence band} \quad . \end{aligned}$$

Quantum thermodynamics in the grand canonical ensemble then says

$$n_{\rm c}(T,\mu) = \int\limits_{E_{\rm c}^0}^{\infty} d\varepsilon\; g_{\rm c}(\varepsilon)\, f(\varepsilon-\mu) \tag{6.10}$$

and

$$p_{\rm v}(T,\mu) \int\limits_{-\infty}^{E_{\rm v}^0} d\varepsilon\; g_{\rm v}(\varepsilon)\,\Big\{1 - f(\varepsilon-\mu)\Big\} = \int\limits_{-\infty}^{E_{\rm v}^0} d\varepsilon\; g_{\rm v}(\varepsilon)\, f(\mu-\varepsilon) \quad , \tag{6.11}$$

where $f(u) = \left[\exp(u/k_{\rm B}T)+1\right]^{-1}$ is the Fermi distribution. For quadratic extrema, the conduction and valence band densities of states behave as $g_{\rm c}(\varepsilon) \propto (\varepsilon - E^0_{\rm c})^{1/2}$ and $g_{\rm v}(\varepsilon) \propto (E^0_{\rm v} - \varepsilon)^{1/2}$. We can define the function $\bar{f}(u) \equiv f(-u)$ to be the Fermi distribution function for holes.

The dependence of $n_{\rm c}(T,\mu)$ and $p_{\rm v}(T\,\mu)$ on the chemical potential in Eqns. 6.10 and 6.11 is complicated. In the Maxwell-Boltzmann limit, however, where $|E^0_{\rm c,v} - \mu| \gg k_{\rm B}T$, the Fermi functions become $f(\varepsilon-\mu) \simeq e^{-(\varepsilon-\mu)/k_{\rm B}T}$ for $\varepsilon > E^0_{\rm c}$ and $f(\mu-\varepsilon) \simeq e^{-(\mu-\varepsilon)/k_{\rm B}T}$ for $\varepsilon < E^0_{\rm v}$. Thus, we may write

$$\begin{aligned} n_{\rm c}(T,\mu) &= N_{\rm c}(T)\, e^{-(E^0_{\rm c}-\mu)/k_{\rm B}T} \\ p_{\rm v}(T,\mu) &= P_{\rm v}(T)\, e^{-(\mu-E^0_{\rm v})/k_{\rm B}T} \quad , \end{aligned} \tag{6.12}$$

where

$$\begin{aligned} N_{\rm c}(T) &= \int\limits_{E^0_{\rm c}}^{\infty} d\varepsilon\; g_{\rm c}(\varepsilon)\, e^{-(\varepsilon-E^0_{\rm c})/k_{\rm B}T} \\ P_{\rm v}(T) &= \int\limits_{-\infty}^{E^0_{\rm v}} d\varepsilon\; g_{\rm v}(\varepsilon)\, e^{-(E^0_{\rm v}-\varepsilon)/k_{\rm B}T} \quad . \end{aligned} \tag{6.13}$$

Now for ellipsoidal bands as in Eqn. 6.9, the density of states, including spin degeneracy, is given by

$$g(\varepsilon) = \frac{\sqrt{2}}{\pi^2\hbar^3}\,\sqrt{m_1^*\,m_2^*\,m_3^*}\;\varepsilon^{1/2}\,\Theta(\varepsilon) \quad , \tag{6.14}$$

in which case

$$N_{\rm c}(T) = 2\,\lambda_{T,{\rm c}}^{-3} \qquad , \qquad P_{\rm v}(T) = 2\,\lambda_{T,{\rm v}}^{-3} \quad , \tag{6.15}$$

where the thermal wavelengths are given by $\lambda_{T,{\rm c/v}} = \left(2\pi\hbar^2/m^*_{\rm c/v}k_{\rm B}T\right)^{1/2}$, with the DOS mass $m^*_{\rm c/v}$ given by $m^*_{\rm c/v} = \left(m^*_{1,{\rm c/v}}\,m^*_{2,{\rm c/v}}\,m^*_{3,{\rm c/v}}\right)^{1/3}$. It is convenient to express the quantities $N_{\rm c}(T)$ and $P_{\rm v}(T)$ as

$$\begin{aligned} N_{\rm c}(T) &= 2.51\times10^{19}\,{\rm cm}^{-3}\left(\frac{m^*_{\rm c}}{m_{\rm e}}\right)^{3/2}\left(\frac{T}{300\,{\rm K}}\right)^{3/2} \\ P_{\rm v}(T) &= 2.51\times10^{19}\,{\rm cm}^{-3}\left(\frac{m^*_{\rm v}}{m_{\rm e}}\right)^{3/2}\left(\frac{T}{300\,{\rm K}}\right)^{3/2} \quad , \end{aligned} \tag{6.16}$$

which, along with Eqn. 6.12, tells us that $10^{19}\,{\rm cm}^{-3}$ is an approximate upper limit to the carrier concentration in a nondegenerate semiconductor[4].

[4] "Nondegenerate" means $|E^0_{\rm c,v} - \mu| \gg k_{\rm B}T$.

6.2.1 Intrinsic semiconductors

How do we find the chemical potential $\mu(T)$? In an *intrinsic* semiconductor, the number of conduction electrons must be equal to the number of valence holes, hence

$$n_{\rm c}(T,\mu) = p_{\rm v}(T,\mu) \quad , \tag{6.17}$$

which is one equation in the two unknowns (T,μ). The solution set is thus the desired function $\mu(T)$. The above equation is difficult to solve owing to the complicated dependence of the integrals in Eqns. 6.10 and 6.11 on μ, but if we are in the Maxwell-Boltzmann limit, a solution is readily available. Writing

$$n_{\rm c}(T,\mu) = N_{\rm c}(T)\,e^{-(E^0_{\rm c}-\mu)/k_{\rm B}T} = P_{\rm v}(T)\,e^{-(\mu-E^0_{\rm c})/k_{\rm B}T} = p_{\rm v}(T,\mu) \quad , \tag{6.18}$$

we have

$$e^{(2\mu-E^0_{\rm c}-E^0_{\rm v})/k_{\rm B}T} = \frac{P_{\rm v}(T)}{N_{\rm c}(T)} = \left(\frac{m^*_{\rm v}}{m^*_{\rm c}}\right)^{3/2} \quad , \tag{6.19}$$

and we find

$$\mu(T) = \tfrac{1}{2}\big(E^0_{\rm c}+E^0_{\rm v}\big) + \tfrac{3}{4}\,k_{\rm B}T\,\ln\left(\frac{m^*_{\rm v}}{m^*_{\rm c}}\right) \quad . \tag{6.20}$$

As $T\to 0$, the chemical potential tends to the average $\mu(0) = \frac{1}{2}\big(E^0_{\rm c}+E^0_{\rm v}\big)$. For finite temperature T, $\mu(T)$ increases with temperature if $m^*_{\rm v} > m^*_{\rm c}$ and decreases with T if $m^*_{\rm v} < m^*_{\rm c}$. Since the ratio $m^*_{\rm v}/m^*_{\rm c}$ is of order unity, we have $|\mu - E^0_{\rm c,v}| = \frac{1}{2}\,\Delta + \mathcal{O}(1)\cdot k_{\rm B}T$, and provided $k_{\rm B}T \ll \Delta$, the degeneracy condition applies. In most semiconductors, $\Delta \gg k_{\rm B}T_{\rm room}$, so intrinsic semiconductors are almost always degenerate at and below room temperature.

6.2.2 Extrinsic semiconductors

In an extrinsic semiconductor, *dopant* ions with energy levels just above the valence band (acceptors) or just below the conduction band (donors) contribute valence holes or conduction electrons, respectively, and there is an imbalance $n_{\rm c} - p_{\rm v} = \Delta n$. When acceptors are present, the chemical potential lies close to the valence band maximum, and the system is said to be p-type. The charge carriers are then valence holes. When donors are present, the chemical potential lies close to the conduction band minimum, and the system is said to be n-type. The charge carriers are then conduction electrons. The densities $n_{\rm c}$ and $p_{\rm v}$ denote only conduction electrons and valence holes, and do not include contributions from impurity states. Regardless of the shift in μ due to extrinsic effects, in the Maxwell-Boltzmann limit the product $n_{\rm c}(T,\mu)\,p_{\rm v}(T,\mu)$ is independent of μ, and given by

$$n_{\rm c}(T,\mu)\,p_{\rm v}(T,\mu) = N_{\rm c}(T)\,P_{\rm v}(T)\,e^{-(E^0_{\rm c}-E^0_{\rm v})/k_{\rm B}T} \equiv n_{\rm i}^2(T) \quad , \tag{6.21}$$

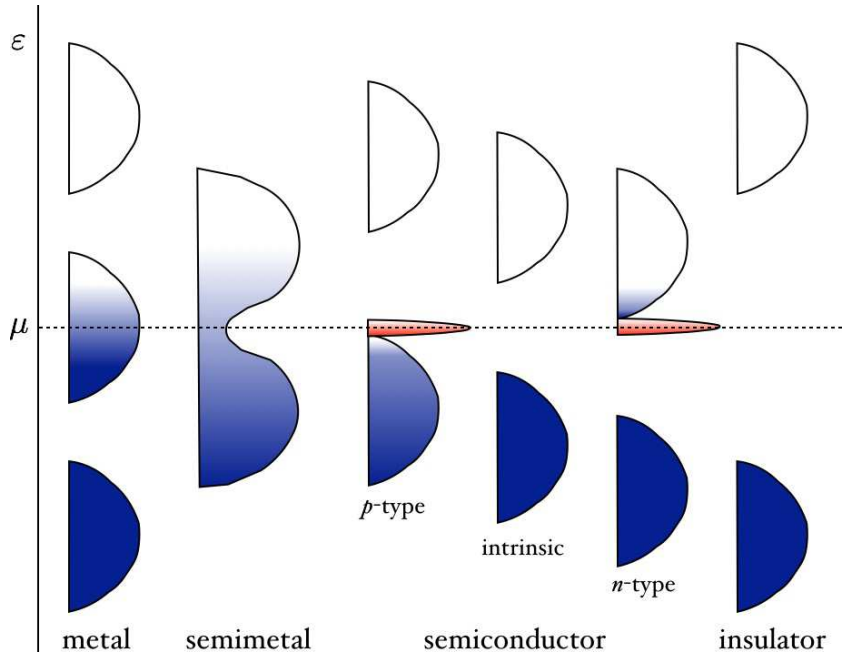


Figure 6.6: Bands and their fillings in metals, semiconductors, and insulators. Narrow red-shaded regions denote impurity bands of acceptors (p-type) and donors (n-type).

where

$$\begin{aligned} n_{\rm i}(T) &\equiv 2\,\bar{\lambda}_T^{-3/2}\,e^{-\Delta/2k_{\rm B}T} \\ &= 2.5\times10^{19}\,{\rm cm}^{-3}\left(\frac{\sqrt{m_{\rm v}^*\,m_{\rm c}^*}}{m_{\rm e}}\right)^{3/2}\left(\frac{T}{300\,{\rm K}}\right)^{3/2}e^{-\Delta/2k_{\rm B}T} \quad , \end{aligned} \tag{6.22}$$

with $\bar{\lambda}_T \equiv \sqrt{\lambda_{T,{\rm c}}\,\lambda_{T,{\rm v}}}$.

In the extrinsic case, then, $n_{\rm c} - p_{\rm v} = \Delta n$ and $n_{\rm c}\,p_{\rm v} = n_{\rm i}^2$ are two equations in two unknowns, with the solution

$$\begin{aligned} n_{\rm c} &= \sqrt{n_{\rm i}^2 + \tfrac{1}{4}(\Delta n)^2} + \tfrac{1}{2}\,\Delta n \\ p_{\rm v} &= \sqrt{n_{\rm i}^2 + \tfrac{1}{4}(\Delta n)^2} - \tfrac{1}{2}\,\Delta n \quad . \end{aligned} \tag{6.23}$$

If we furthermore define the quantity $\mu_{\rm i}$ such that

$$n_{\rm c}(T,\mu) \equiv n_{\rm i}(T)\,e^{(\mu-\mu_{\rm i})/k_{\rm B}T} \qquad , \qquad p_{\rm v}(T,\mu) \equiv n_{\rm i}(T)\,e^{(\mu_{\rm i}-\mu)/k_{\rm B}T} \quad , \tag{6.24}$$

then the quantities $(\Delta n\,,\,\mu_{\rm i})$ are related by

$$\Delta n = 2\,n_{\rm i}(T)\,\sinh\left(\frac{\mu-\mu_{\rm i}}{k_{\rm B}T}\right) \quad . \tag{6.25}$$

II	III	IV	V	VI
	B [He] $2s^2\,2p^1$	C [He] $2s^2\,2p^2$	N [He] $2s^2\,2p^3$	O [He] $2s^2\,2p^4$
	Al [Ne] $3s^2\,3p^1$	Si [Ne] $3s^2\,3p^2$	P [Ne] $3s^2\,3p^3$	S [Ne] $3s^2\,3p^4$
Zn [Ar] $4s^2\,3d^{10}$	Ga [Ar] $4s^2\,3d^{10}\,4p^1$	Ge [Ar] $4s^2\,3d^{10}\,4p^2$	As [Ar] $4s^2\,3d^{10}\,4p^3$	Se [Ar] $4s^2\,3d^{10}\,4p^4$
Cd [Kr] $5s^2\,4d^{10}$	In [Kr] $5s^2\,4d^{10}\,5p^1$	Sn [Kr] $5s^2\,4d^{10}\,5p^2$	Sb [Kr] $5s^2\,4d^{10}\,5p^3$	Te [Kr] $5s^2\,4d^{10}\,5p^4$
Hg [Xe] $6s^2\,5d^{10}$	Tl [Xe] $6s^2\,5d^{10}\,6p^1$	Pb [Xe] $6s^2\,5d^{10}\,6p^2$	Bi [Xe] $6s^2\,5d^{10}\,6p^3$	Po [Xe] $6s^2\,5d^{10}\,6p^4$

Figure 6.7: Relevant group II through group VI elements and their electronic configurations.

Now if $\Delta n/n_{\rm i}$ is small, then $|\mu - \mu_{\rm i}| \ll k_{\rm B}T$, and if the degeneracy condition applies, this means that μ is far from $E^0_{\rm c,v}$, and both $n_{\rm c} \approx p_{\rm v} \approx n_{\rm i}$. This remains the case so long as $|\Delta n| \ll n_{\rm i}$.

In the opposite limit, when $|\Delta n| \gg n_{\rm i}$, we have

$$\sqrt{n_{\rm i}^2 + \tfrac{1}{4}(\Delta n)^2} = \tfrac{1}{2}|\Delta n| + \frac{n_{\rm i}^2}{|\Delta n|} + \ldots \quad , \tag{6.26}$$

and therefore

$$\begin{aligned} n\text{-type} \quad &: \quad \frac{\Delta n}{n_{\rm i}} \gg +1 \quad \Rightarrow \quad n_{\rm c} = \Delta n \quad , \quad p_{\rm v} = \frac{n_{\rm i}^2}{\Delta n} \\ p\text{-type} \quad &: \quad \frac{\Delta n}{n_{\rm i}} \ll -1 \quad \Rightarrow \quad n_{\rm c} = \frac{n_{\rm i}^2}{|\Delta n|} \quad , \quad p_{\rm v} = |\Delta n| \quad . \end{aligned} \tag{6.27}$$

6.3 Donors and Acceptors

6.3.1 Impurity charges in a semiconductor

Silicon is a group IV element in the periodic table. To its left sits aluminum and to its right sits phosphorus. Consider a Si crystal in which one of the Si atoms has been replaced by a P atom. Compared to silicon, phosphorus has one extra nuclear charge and one additional electron. In free space, this last P electron has a binding energy of $10.5\,\text{eV}$, the first ionization potential of P. In a crystal, this binding energy is drastically reduced, due to two effects:

- The static dielectric constant of the semiconductor crystal is typically large ($\epsilon_{\rm Si} = 11.9$, $\epsilon_{\rm InSb} = 16.8$, $\epsilon_{\rm PbTe} = 30$). Small gaps lead to large dielectric constants[5]. Later on we shall derive the expression
$$\epsilon \lesssim 1 + \left(\frac{\hbar\omega_{\rm pv}}{\Delta}\right)^2 \quad , \tag{6.28}$$
where $\omega_{\rm pv} = (4\pi n_{\rm v} e^2/m_{\rm e})^{1/2}$, with $n_{\rm v}$ the number density of valence electrons. So the attraction between the phosphorus core and the added electron is reduced by a factor of ϵ, provided the radius of the electronic orbit is on the order of several lattice spacings.
- The effective mass of the electrons in the conduction band is $m_{\rm c}^*$, which is generally about a tenth of the electron mass $m_{\rm e}$.

The radius of the orbit is thus not $a_{\rm B} = \hbar^2/me^2$, but rather
$$r_0 = \frac{\epsilon\,\hbar^2}{m_{\rm c}^* e^2} = \frac{m_{\rm e}}{m_{\rm c}^*}\,\epsilon\,a_{\rm B} \quad . \tag{6.29}$$
If $\epsilon \approx 10$ and $m_{\rm c}^* \approx 0.1\,m_{\rm e}$, we have $r_0 \approx 100\,a_{\rm B}$. The binding energy W of the lowest hydrogenic state is then given by
$$W_{\rm d} = \frac{e^2}{2\epsilon\,r_0} = \overbrace{\frac{e^2}{2a_{\rm B}}}^{13.6\,\text{eV}} \cdot \overbrace{\frac{m_{\rm c}^*}{m_{\rm e}}\,\frac{1}{\epsilon^2}}^{\approx 10^{-3}} \tag{6.30}$$
Thus, the donor binding energy is $W_{\rm d} \approx 10^{-3}\,\text{Ry}$, which is usually very small in comparison with the band gap Δ. This means that the donor levels lie just below the conduction band minimum $E_{\rm c}^0$, as depicted in Fig. 6.8. The calculation for hole binding by acceptors is identical, aside from the replacement of $m_{\rm c}^*$ by $m_{\rm v}^*$. Thus,
$$\begin{aligned} E_{\rm d} &= E_{\rm c}^0 - \frac{e^2}{2a_{\rm B}} \cdot \frac{m_{\rm c}^*}{m_{\rm e}}\,\frac{1}{\epsilon^2} \\ E_{\rm a} &= E_{\rm v}^0 + \frac{e^2}{2a_{\rm B}} \cdot \frac{m_{\rm v}^*}{m_{\rm e}}\,\frac{1}{\epsilon^2} \quad . \end{aligned} \tag{6.31}$$

[5]In a metal, where ther is no gap, $\epsilon = \infty$.

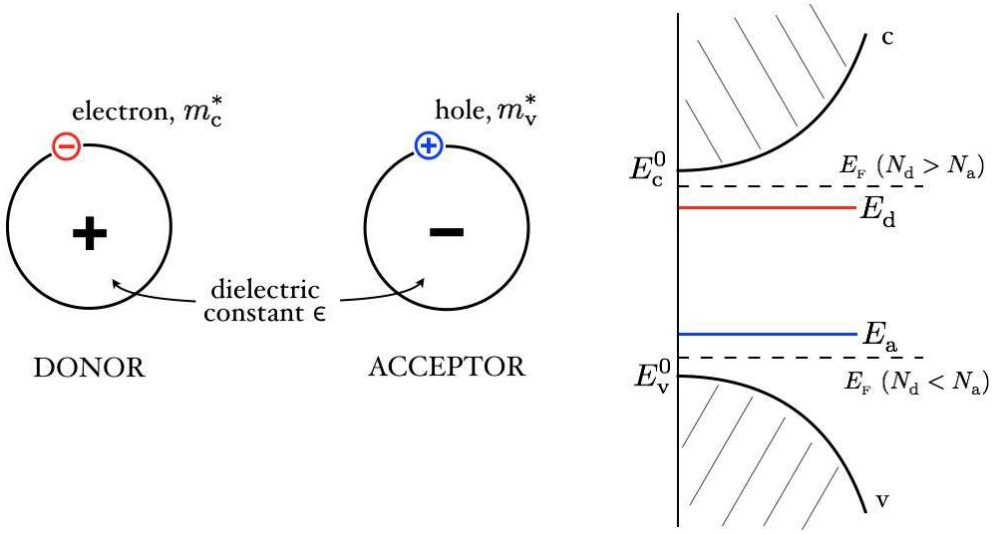


Figure 6.8: Donors and acceptors.

		group III : acceptors						group V : donors					
	$\Delta(300\,\mathrm{K})$	W_a	B	Al	Ga	In	Tl	W_d	N	P	As	Sb	Bi
Si	1.12	48	45.0	68.5	71	155	245	113	140	45.3	53.7	42.7	70.6
Ge	0.67	15	10.8	11.1	11.3	12.0	13.5	28	–	12.9	14.2	10.3	12.8

Table 6.2: Donor and acceptor binding energies in Si and Ge (in meV).

6.3.2 Donor and acceptor quantum statistics

In the presence of donor and acceptor ions, the net change in the background ionic charge density is given by $\Delta\rho_{\rm ion} = e\,(N_{\rm d} - N_{\rm a})$. Since the net system is charge neutral, this must be balanced by the net electronic charge density, $\Delta\rho_{\rm elec} = -e\,(n_{\rm c} + n_{\rm d} - p_{\rm v} - p_{\rm a})$ where $n_{\rm d}$ and $p_{\rm a}$ are the number densities of donor electrons and acceptor holes in equilibrium, respectively. The charge neutrality condition is then $\Delta\rho_{\rm ion} + \Delta\rho_{\rm elec} = 0$, which requires

$$n_{\rm c} - p_{\rm v} + n_{\rm d} - p_{\rm a} = N_{\rm d} - N_{\rm a} \quad . \tag{6.32}$$

The question now arises of how to compute $n_{\rm d}$ and $p_{\rm a}$. The simplest assumption is to assume the donor and acceptor levels of each spin are independently occupied according to a Fermi distribution, just like the conduction and valence band levels. Under this assumption,

$$n_{\rm d} = 2N_{\rm d}\,f(E_{\rm d} - \mu) \qquad , \qquad p_{\rm a} = 2N_{\rm a}\,f(\mu - E_{\rm a}) \quad , \tag{6.33}$$

where the factor of 2 comes from spin degeneracy. However, donor and acceptor state wavefunctions are *localized* in space, and so donor states with two electrons and acceptor states with

$\vert\,\Psi\,\rangle$	E	$\hat{n}$
$\vert\,0\,\rangle$	0	0
$\vert\uparrow\rangle\,,\vert\downarrow\rangle$	$E_{\rm d}$	1
$\vert\uparrow\downarrow\rangle$	$2E_{\rm d}+U$	2

Table 6.3: Donor states and their energies.

two holes are energetically disfavored. Consider the case of a donor level. When one electron of either spin polarization is present $\big(\vert\uparrow\rangle$ or $\vert\downarrow\rangle\big)$, the energy is $E = E_{\rm d}$. When two electrons are present in the state $\vert\uparrow\downarrow\rangle$, the energy is $E = 2E_{\rm d}+U$, where U is an extra Coulomb repulsion energy between the two electrons. Thus, in thermal equilibrium at temperature T, the average donor occupancy is

$$\langle\hat{n}\rangle = \frac{2\,e^{-\beta(E_{\rm d}-\mu)} + 2\,e^{-\beta(2E_{\rm d}-2\mu+U)}}{1+2\,e^{-\beta(E_{\rm d}-\mu)}+e^{-\beta(2E_{\rm d}-2\mu+U)}} \quad . \tag{6.34}$$

When $U=0$, we obtain $\langle n\rangle = 2f(E_{\rm d}-\mu)$. When $e^{-\beta U}\ll 1$, we have

$$\langle\hat{n}\rangle = \frac{2}{e^{\beta(E_{\rm d}-\mu)}+2} \quad . \tag{6.35}$$

More generally, for a donor with a $g_{\rm d}$-fold degeneracy of the $\hat{n}=1$, and for acceptor states with a $g_{\rm a}$-fold degeneracy of the $\hat{p}=1$ level (*i.e.* one hole), we have that the average occupancy is

$$\langle\hat{n}\rangle = \frac{g_{\rm d}}{e^{\beta(E_{\rm d}-\mu)}+g_{\rm d}} \quad , \quad \langle\hat{p}\rangle = \frac{g_{\rm a}}{e^{\beta(\mu-E_{\rm a})}+g_{\rm a}} \quad . \tag{6.36}$$

This means that the donor electron density and acceptor hole density are given by

$$n_{\rm d}(T,\mu) = \frac{g_{\rm d}N_{\rm d}}{e^{\beta(E_{\rm d}-\mu)}+g_{\rm d}} \quad , \quad p_{\rm a}(T,\mu) = \frac{g_{\rm a}N_{\rm a}}{e^{\beta(\mu-E_{\rm a})}+g_{\rm a}} \quad . \tag{6.37}$$

Typically $g_{\rm d}=2$, but in many cases there is an extra degeneracy of the acceptor states. We are left with the following equation to be solved for $\mu(T)$:

$$N_{\rm c}(T)\,e^{-(E^0_{\rm c}-\mu)/k_{\rm B}T} - P_{\rm v}(T)\,e^{-(\mu-E^0_{\rm v})/k_{\rm B}T} = \frac{N_{\rm d}}{g_{\rm d}\,e^{\beta(\mu-E_{\rm d})}+1} - \frac{N_{\rm a}}{g_{\rm a}\,e^{\beta(E_{\rm a}-\mu)}+1} \quad . \tag{6.38}$$

Consider now the case where $E_{\rm d}-\mu \gg k_{\rm B}T$ and $\mu - E_{\rm a} \gg k_{\rm B}T$. In this case, Eqn. 6.38 becomes

$$N_{\rm c}(T)\,e^{-(E^0_{\rm c}-\mu)/k_{\rm B}T} - P_{\rm v}(T)\,e^{-(\mu-E^0_{\rm v})/k_{\rm B}T} = N_{\rm d}-N_{\rm a} \quad . \tag{6.39}$$

From Eqn. 6.23, we have the solution

$$\left\{\begin{matrix} n_{\rm c} \\ p_{\rm v}\end{matrix}\right\} = \tfrac{1}{2}\sqrt{(N_{\rm d}-N_{\rm a})^2+4n_{\rm i}^2} \;\pm\; \tfrac{1}{2}(N_{\rm d}-N_{\rm a}) \quad . \tag{6.40}$$

with $n_{\rm i}(T) = 2\,\bar\lambda_T^{-3/2}\,e^{-\Delta/2k_{\rm B}T}$. For light doping, where $|N_{\rm d} - N_{\rm a}| \ll n_{\rm i}$, we have

$$\left\{\begin{matrix} n_{\rm c} \\ p_{\rm v} \end{matrix}\right\} \approx n_{\rm i} \pm \tfrac{1}{2}(N_{\rm d} - N_{\rm a}) \quad . \tag{6.41}$$

In the opposite limit, where $|N_{\rm d} - N_{\rm a}| \gg n_{\rm i}$, we have

$$N_{\rm d} > N_{\rm a} \quad : \quad n_{\rm c} \approx N_{\rm d} - N_{\rm a} \qquad , \qquad p_{\rm v} \approx \frac{n_{\rm i}^2}{N_{\rm d} - N_{\rm a}} \tag{6.42}$$

and

$$N_{\rm d} < N_{\rm a} \quad : \quad n_{\rm c} \approx \frac{n_{\rm i}^2}{N_{\rm a} - N_{\rm d}} \qquad , \qquad p_{\rm v} \approx N_{\rm a} - N_{\rm d} \quad . \tag{6.43}$$

6.3.3 Chemical potential *versus* temperature in doped semiconductors

How does the chemical potential $\mu(T)$ behave as a function of temperature in a doped semiconductor? In n-doped materials, charge neutrality requires that all the acceptor levels are singly occupied at $T = 0$, to compensate for the extra background charge. Thus, $\varepsilon_{\rm F} = \mu(T = 0)$ must lie between the donor energy $E_{\rm d}$ and the conduction band minimum $E_{\rm c}^0$. For p-doped materials, $\varepsilon_{\rm F} = \mu(T = 0)$ lies between the acceptor energy $E_{\rm a}$ and the valence band maximum $E_{\rm v}^0$.

What happens for large T? The answer depends on what we mean by "large", *i.e.* large compared to what? Assuming $\Delta \gg |E_{\rm d,a} - \mu|$, we have, from Eqn. 6.20, $\mu(T) = \frac{1}{2}\big(E_{\rm c}^0 + E_{\rm v}^0\big) + \frac{3}{4}\,k_{\rm B}T\,\ln\big(m_{\rm v}^*/m_{\rm c}^*\big)$. At still higher temperatures, if we make the dubious assumption that there is a further separation of energy scales which allows us to consider only the valence and conduction bands, we may write

$$2 = \int\limits_{E_{\rm v}^-}^{E_{\rm c}^+} d\varepsilon\,\overline{g}(\varepsilon)\,f(\varepsilon-\mu) = \int\limits_{E_{\rm v}^-}^{E_{\rm c}^+} d\varepsilon\,\frac{\overline{g}(\varepsilon)}{2+\beta(\varepsilon-\mu)+\ldots} = 2 - \beta\mu + \tfrac{1}{4}\beta\int\limits_{E_{\rm v}^-}^{E_{\rm c}^+} d\varepsilon\,\overline{g}(\varepsilon)\,\varepsilon + \mathcal{O}(\beta^2) \quad , \tag{6.44}$$

where $\overline{g}(\varepsilon) = \overline{g}_{\rm v}(\varepsilon) + \overline{g}_{\rm c}(\varepsilon)$ is the total density of states per unit cell for both bands, and where $E_{\rm v}^-$ and $E_{\rm c}^+$ denote the *lowest* energy of the valence band and *highest* energy of the conduction band, respectively. Recall that $\int\limits_{-\infty}^{\infty} d\varepsilon\,\overline{g}_{\rm v,c}(\varepsilon) = 2$, *i.e.* each band accommodates a maximum of two electrons per cell. Thus, we conclude

$$\mu(T \gg \Delta) = \tfrac{1}{2}\langle\varepsilon\rangle_{\rm v} + \tfrac{1}{2}\langle\varepsilon\rangle_{\rm c} \equiv \bar E_{\rm cv} \quad , \tag{6.45}$$

where

$$\langle\varepsilon\rangle_{\rm v} = \tfrac{1}{2}\int\limits_{E_{\rm v}^-}^{E_{\rm v}^0} d\varepsilon\,\overline{g}_{\rm v}(\varepsilon)\,\varepsilon \qquad , \qquad \langle\varepsilon\rangle_{\rm c} = \tfrac{1}{2}\int\limits_{E_{\rm c}^0}^{E_{\rm c}^+} d\varepsilon\,\overline{g}_{\rm c}(\varepsilon)\,\varepsilon \tag{6.46}$$

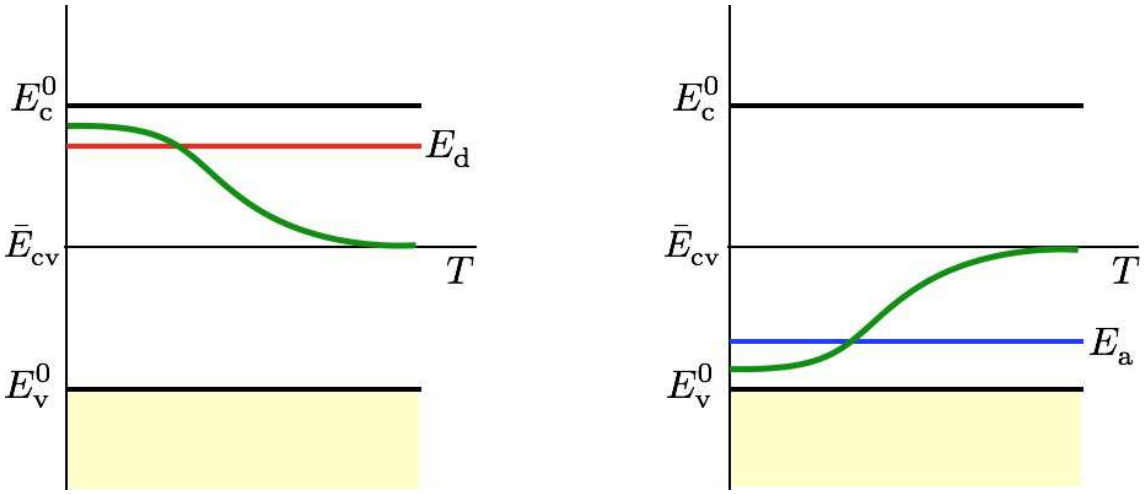


Figure 6.9: Evolution of the chemical potential with temperature in extrinsic semiconductors. The high temperature limit still assumes that only the valence and conduction bands need be considered.

are the *average* band energies. The situation is depicted in Fig. 6.9.

Of course, eventually other bands will enter the picture. In the infinite temperature limit, even the crystalline potential is irrelevant, and we can appeal to the classical result $n_\sigma = \lambda_T^{-3} \exp(\mu/k_{\rm B}T)$, for each spin polarization σ, where $\lambda_T = (2\pi\hbar^2/m_{\rm e}k_{\rm B}T)^{1/2}$. One then has $\mu \sim -\frac{3}{2}T\ln T$ as $T \to \infty$.

6.4 Inhomogeneous Semiconductors

Most of the technological uses of semiconductors are associated with materials which have inhomogeneous doping profiles: p-n junctions, MOSFETs, heterojunctions, *etc.* The parade example is the p-n junction, depicted in Fig. 6.10. We imagine a doping gradient such that the

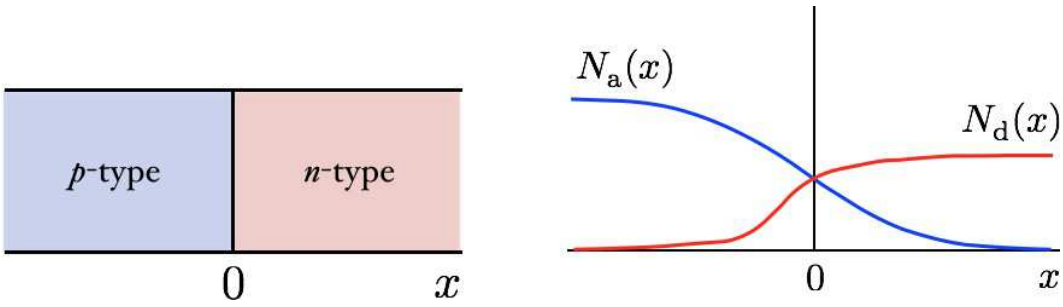


Figure 6.10: The p-n junction. Left: idealized case. Right: typical doping profile.

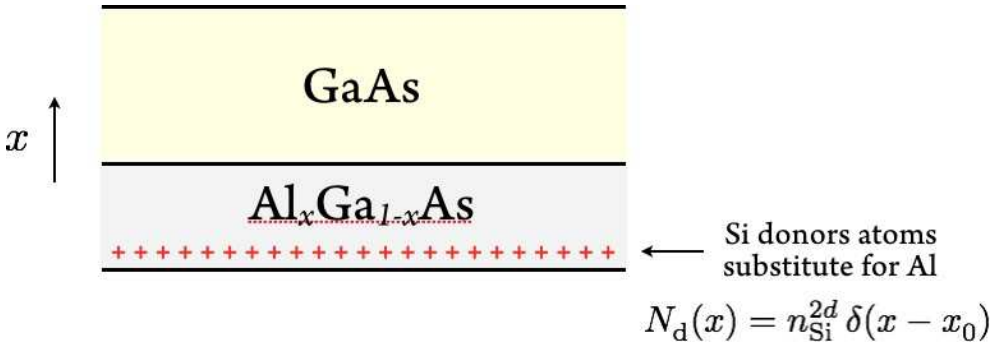


Figure 6.11: A GaAs - Al_xGa_xAs heterostructure formed by δ-doping.

system is p-doped for $x < 0$ and n-doped for $x > 0$. Typically this is accomplished by varying the impurity concentration in a melt from which the solid is formed. Advances in growth techniques such as MBE (molecular beam epitaxy) now allow for layer-by-layer growth and doping profiles with nearly atomic precision. An important example of this is the δ-doped GaAs - Al_xGa_xAs heterostructure sketched in Fig. 6.11. As we shall discuss further below, Al_xGa_x has a larger band gap than GaAs. Substituting Si for an Al atom results in a donor, but, as we shall see, the valence electrons in the Al_xGa_x migrate over to the GaAs side of the heterostructure. By placing the Si dopants far from the interface, one thereby creates a two-dimensional electron gas (2DEG) with extremely high mobility. This feature was crucial to the 1982 discovery of the fractional quantum Hall effect in by Tsui, Störmer, and Gossard.

6.4.1 Modeling the p - n junction

Here we follow the rather clear discussion in chapter 29 of Ashcroft and Mermin, *Solid State Physics*.

We assume that the acceptor and donor densities are spatially varying according to

$$N_{\rm d}(x) = N_{\rm d}\,\Theta(x) \qquad , \qquad N_{\rm a}(x) = N_{\rm a}\,\Theta(-x) \quad . \tag{6.47}$$

In general, inhomogeneous doping along the x-direction in space will lead to a spatially varying electrostatic potential $\phi(x)$. Semiclassically, Bloch electrons in such a spatially varying potential are described by the Hamiltonian

$$H_n = E_n\left(\frac{\boldsymbol{p}}{\hbar} + \frac{e}{\hbar c}\,\boldsymbol{A}(\boldsymbol{r})\right) - e\,\phi(x) \quad . \tag{6.48}$$

We will consider the case with $\boldsymbol{B} = 0$, in which case we may choose a gauge in which $\boldsymbol{A} = 0$. Notice that the crystalline potential is not present explicitly, but rather is embodied in the Bloch dispersion $E_n(\boldsymbol{k})$. Such a description is valid provided the potential $\phi(x)$ varies slowly

on atomic scales., *i.e.* $|\boldsymbol{\nabla}\phi| \ll \Delta/ae$. If we further assume that the nondegeneracy condition $|E^0_{\rm c,v} - \mu| \gg k_{\rm B}T$ holds, then we have

$$\begin{aligned} n_{\rm c}(x) &= 2\,\lambda_{T,{\rm c}}^{-3}\,\exp\left(-\frac{E^0_{\rm c} - e\,\phi(x) - \mu}{k_{\rm B}T}\right) \\ p_{\rm v}(x) &= 2\,\lambda_{T,{\rm v}}^{-3}\,\exp\left(-\frac{\mu + e\,\phi(x) - E^0_{\rm v}}{k_{\rm B}T}\right) \quad . \end{aligned} \tag{6.49}$$

Thus, at the ends of the junction, we have

$$\begin{aligned} n_{\rm c}(x=+\infty) &= 2\,\lambda_{T,{\rm c}}^{-3}\,\exp\left(-\frac{E^0_{\rm c} - e\,\phi(+\infty) - \mu}{k_{\rm B}T}\right) = N_{\rm d} \\ p_{\rm v}(x=-\infty) &= 2\,\lambda_{T,{\rm v}}^{-3}\,\exp\left(-\frac{\mu + e\,\phi(-\infty) - E^0_{\rm v}}{k_{\rm B}T}\right) = N_{\rm a} \quad , \end{aligned} \tag{6.50}$$

where the second equality in each line follows from the analysis of §6.3.2, assuming that the condition $|N_{\rm d} - N_{\rm a}| \gg 2\,(\lambda_{T,{\rm v}}\,\lambda_{T,{\rm c}})^{-3/4}\,e^{-\Delta/2k_{\rm B}T} \equiv n_{\rm i}(T)$ holds. Multiplying these two equations yields the result

$$e\,\Delta\phi = \Delta + k_{\rm B}T\ln\left(\tfrac{1}{2}N_{\rm a}\lambda_{T,{\rm v}}^3\right) + k_{\rm B}T\ln\left(\tfrac{1}{2}N_{\rm d}\lambda_{T,{\rm c}}^3\right) \quad , \tag{6.51}$$

where $\Delta\phi \equiv \phi(+\infty) - \phi(-\infty)$ is the potential drop across the sample. We may now write

$$\begin{aligned} n_{\rm c}(x) &= N_{\rm d}\,e^{-e[\phi(+\infty)-\phi(x)]/k_{\rm B}T} \\ p_{\rm v}(x) &= N_{\rm a}\,e^{-e[\phi(x)-\phi(-\infty)]/k_{\rm B}T} \quad . \end{aligned} \tag{6.52}$$

Next we would like to determine the potential function $\phi(x)$ throughout the sample. To do this, we invoke Poisson's equation,

$$\frac{d^2\phi}{dx^2} = -\frac{4\pi\rho}{\epsilon} = \frac{4\pi e}{\epsilon}\Big\{N_{\rm a}(x) + n_{\rm c}(x) - N_{\rm d}(x) - p_{\rm v}(x)\Big\} \quad , \tag{6.53}$$

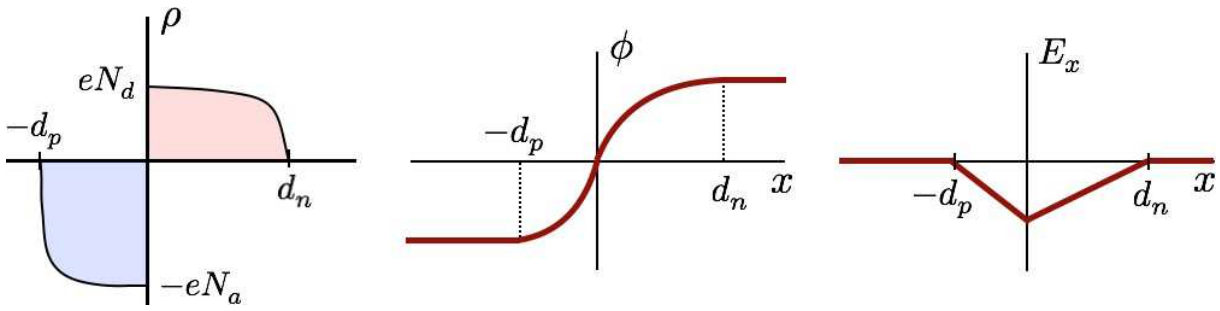


Figure 6.12: The p-n junction in equilibrium. Left: charge density $\rho(x)$. Middle: electrical potential $\phi(x)$. Right: electric field $E_x(x)$.

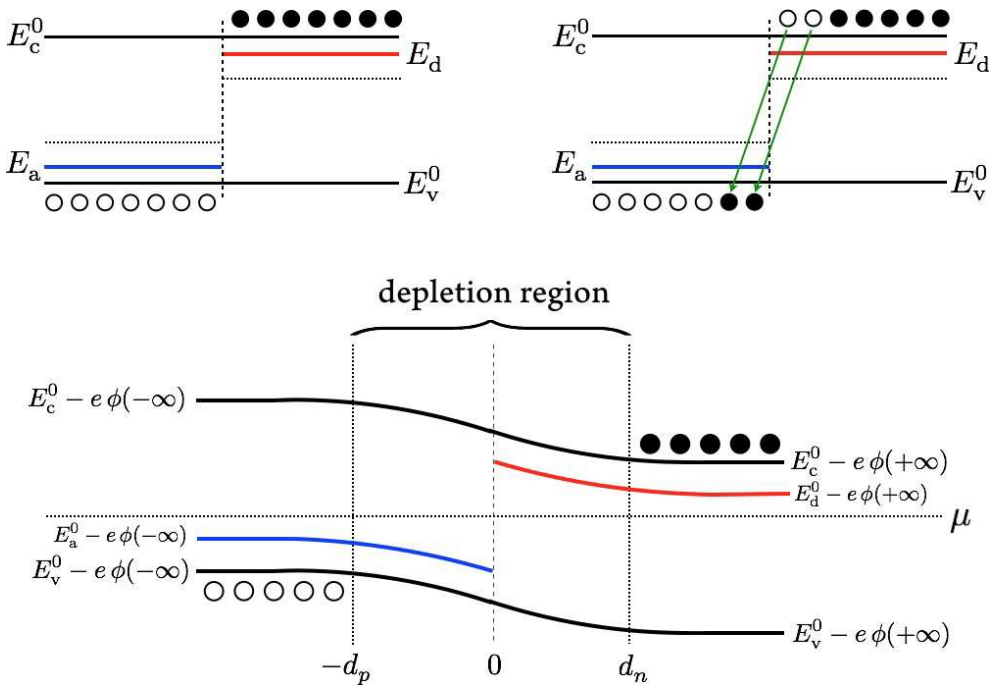


Figure 6.13: The p-n junction. Top left: p-type (left) and n-type (right) doped semiconductor at finite T with mismatched chemical potentials $\mu_{\rm n}(x)$ (dotted horizontal lines). Top right: Conduction band electrons fill valence band holes, thereby lowering the total energy and creating a depletion region devoid of mobile carriers, where the local imbalance of donor *versus* acceptor ions produces a local charge density $\rho(x)$ and electric potential $\phi(x)$. Bottom: The p-n junction in equilibrium. The electrochemical potential $\mu = \mu_{\rm n}(x) - e\phi(x)$ is constant throughout space. The energy bands bend with the local potential $\phi(x)$.

where we are further assuming $n_{\rm d}(x) \ll N_{\rm d}(x)$ and $p_{\rm a}(x) \ll N_{\rm a}(x)$, which are valid provided

$$\big|E_{\rm d,a} - \mu - e\,\phi(x)\big| \gg k_{\rm B}T \quad . \tag{6.54}$$

Since $n_{\rm c}(x)$ and $p_{\rm v}(x)$ depend on $\phi(x)$ through Eqn. 6.49, Eqn. 6.53 may be regarded as a nonlinear second order ODE for $\phi(x)$, rendered inhomogeneous through the appearance of the source terms $N_{\rm a}(x)$ and $N_{\rm d}(x)$. The self-consistent nature of Poisson's equation calls to mind the Debye-Hückel theory of screening in classical plasmas, except here we are not permitted to linearize in $\beta = (k_{\rm B}T)^{-1}$. To render our problem analytically tractable, we'll assume that $x > d_n$ that $n_{\rm c}(x) = n_{\rm c}(+\infty) = N_{\rm d}$ and thus $\rho(x > d_n) = 0$; similarly we assume that for $x < -d_p$ that $p_{\rm v}(x) = p_{\rm v}(-\infty) = N_{\rm a}$ and $\rho(x < -d_p) = 0$. In the course of our calculations, we shall determine the unknown distances $d_{n,p}$.

Outside of the region $x \in [-d_p\,,\,d_n]$, called the *depletion region* or the *space charge later*, the charge

density is zero. Within the space charge layer, we take

$$-\frac{\epsilon}{4\pi}\frac{d^2\phi}{dx^2} = \rho(x) \approx \begin{cases} 0 & \text{if} \quad x \leqslant -d_p \\ -eN_{\rm a} & \text{if} \quad -d_p < x \leqslant 0 \\ +eN_{\rm d} & \text{if} \quad 0 < x \leqslant d_n \\ 0 & \text{if} \quad x > d_n \quad . \end{cases} \tag{6.55}$$

Integrating, we have

$$\phi(x) = \begin{cases} \phi(-\infty) & \text{if} \quad x \leqslant -d_p \\ \phi(-\infty) + 2\pi\epsilon^{-1}eN_{\rm a}\,(x+d_p)^2 & \text{if} \quad -d_p < x \leqslant 0 \\ \phi(+\infty) - 2\pi\epsilon^{-1}eN_{\rm d}\,(x-d_n)^2 & \text{if} \quad 0 < x \leqslant d_n \\ \phi(+\infty) & \text{if} \quad x > d_n \quad . \end{cases} \tag{6.56}$$

We now match the potential $\phi(x)$ and its derivative $\phi'(x)$ at $x = 0$, obtaining

$$\phi(0^-) = \phi(0^+) \qquad \Rightarrow \qquad \Delta\phi = 2\pi\epsilon^{-1}e\big(N_{\rm d}d_n^2 + N_{\rm a}d_p^2\big) \tag{6.57}$$

and

$$\phi;(0^-) = \phi;(0^+) \qquad \Rightarrow \qquad d_nN_{\rm d} = d_pN_{\rm a} \quad . \tag{6.58}$$

Solving these two equations for the unknowns $d_{n,p}$, we have

$$d_p = \left[\frac{N_{\rm d}/N_{\rm a}}{N_{\rm d}+N_{\rm a}}\cdot\frac{\epsilon\,\Delta\phi}{2\pi e}\right]^{1/2} \quad , \qquad d_n = \left[\frac{N_{\rm a}/N_{\rm d}}{N_{\rm d}+N_{\rm a}}\cdot\frac{\epsilon\,\Delta\phi}{2\pi e}\right]^{1/2} \quad , \tag{6.59}$$

where $\Delta\phi$ is given in Eqn. 6.51. Typically $d_{n,p}$ are on the order of 100 Å to 1000 Å. The charge density ρ, electrical potential ϕ, and electric field $E_x = -d\phi/dx$ are all depicted in Fig. 6.12.

Let's reflect on the physics of why all this happens, following the sketches in Fig. 6.13. Conduction electrons on the n-type ($x > 0$) side of the junction can lower their energy by recombining with valence holes on the p-type side ($x < 0$). This leads to a departure from local charge neutrality, which thereby discourages further charge separation. Finally, a built-in potential $\phi(x)$ is established.

6.4.2 Rectification

Under equilibrium conditions, the electrochemical potential μ is constant across the junction. What happens if a bias voltage V is imposed? We'll define $V > 0$ (*forward bias*) as the condition where the external voltage source raises the electrical potential on the p side, and $V < 0$ (*reverse bias*) as the condition where the external voltage source raises the electrical potential on the n

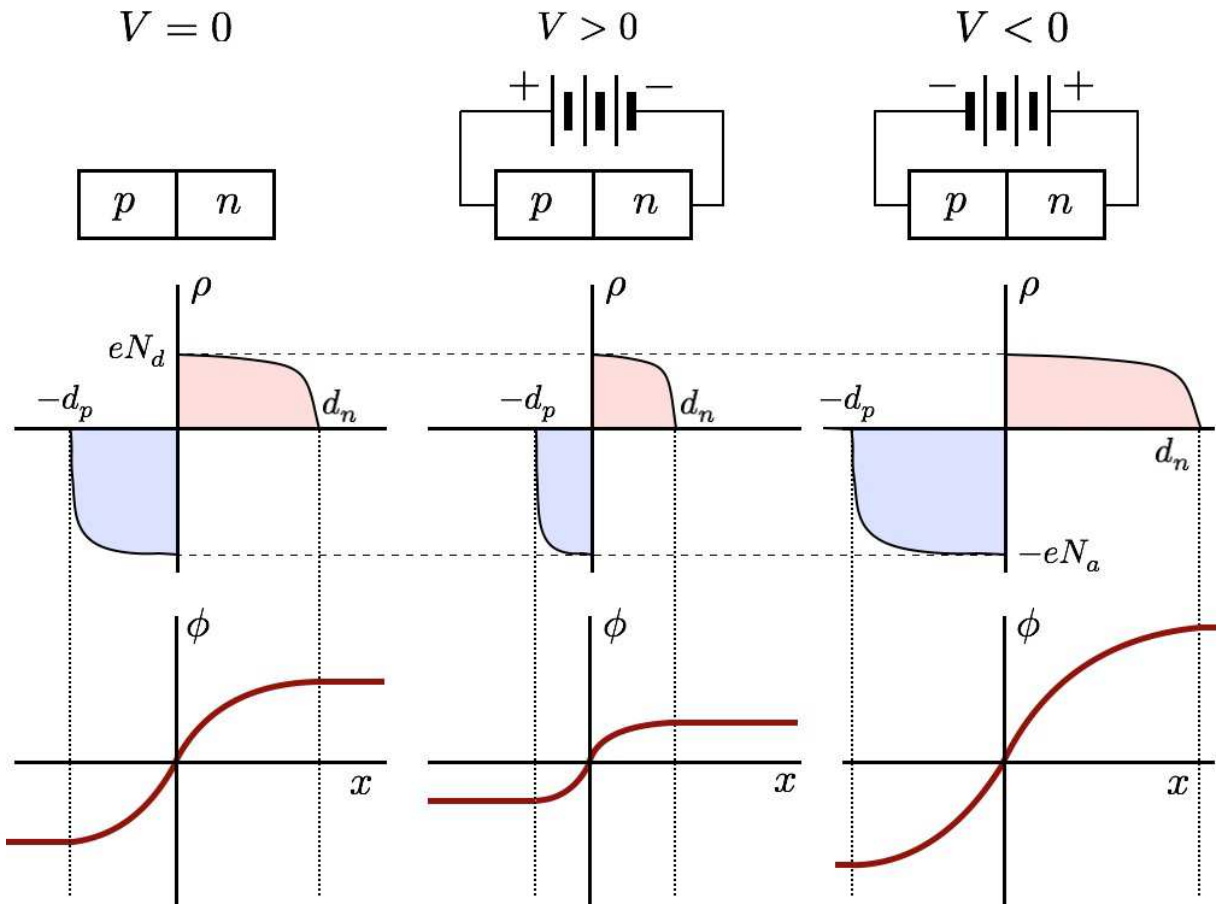


Figure 6.14: The biased p-n junction. Left: zero bias. Center: forward bias. Right: reverse bias.

side. Let $\phi_0(x)$ be the $V = 0$ solution we have just derived, and let $\phi(x)$ be the solution in the presence of the external voltage source. Then

$$\phi(+\infty) - \phi(-\infty) = \phi_0(+\infty) - \phi_0(-\infty) - V \quad . \tag{6.60}$$

Most of this potential drop still occurs in the depletion region. The reason is the depletion refion is depleted of charge carriers (duh!) and therefore has a higher electrical resistance than the bulk p-type and n-type regions. As you know, in a circuit comprised of several resistors in series, the greatest potential drop occurs across the largest resistor. Therefore we have from Eqn. 6.59,

$$d_{n,p}(V) = d_{n,p}(0) \cdot \left(1 - \frac{V}{\Delta\phi_0}\right)^{1/2} \quad . \tag{6.61}$$

As shown in the sketch in Fig. 6.13, there are no conduction electrons on the p-side of the junction, nor valence holes on the n-side. Strictly speaking, this is incorrect, since thermal fluctuations will produce particle-hole excitations across the gap. This is a small but crucial effect. In a homogeneous semiconductor, these excitations would simply recombine without

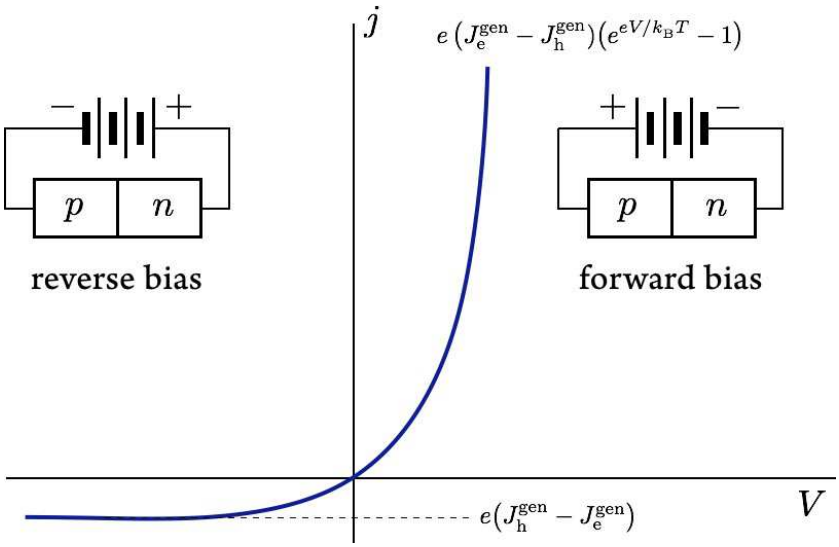


Figure 6.15: $j(V)$ for a biased p-n junction.

significant consequence, however in a p-n junction, there is an internal electric field pushing positive charges to the left and negative charges to the right. So valence holes move to the p side and conduction electrons to the n side, even at $V = 0$. This leads to a *generation current* $j_{\rm gen} = e\,(J_{\rm h} - J_{\rm e})$, where $J_{\rm h} > 0$ and $J_{\rm e} < 0$ are the number current densities of holes and electrons, respectively. Both these currents are proportional to $\exp(-\Delta/k_{\rm B}T)$ and are fairly insensitive to any bias V.

In equilibrium, there can be no net hole or electron current. Therefore there must be a counter-current of holes flowing from p to n, and of electrons flowing from n to p. These currents, which are akin to salmon swimming upstream, since they must flow in opposition to the electric field and overcome the built-in potential barrier $\Delta\phi = \Delta\phi_0 - V$, are called *recombination currents*. When $V = 0$, there is precise cancellation of the generation and recombination currents. Since $J^{\rm rec} \propto \exp\big[-e(\Delta\phi_0 - V)\big]$ and $J^{\rm rec}(V = 0) = -J^{\rm gen}$ (for each species), we conclude

$$J^{\rm rec}(V) = -J^{\rm gen}\, e^{eV/k_{\rm B}T} \quad . \tag{6.62}$$

The electrical current is then

$$\begin{aligned} j(V) &= e\,J_{\rm h}^{\rm rec} + e\,J_{\rm h}^{\rm gen} - e\,J_{\rm e}^{\rm rec} - e\,J_{\rm e}^{\rm gen} \\ &= e\,\big(J_{\rm h}^{\rm gen} - J_{\rm e}^{\rm gen}\big)\big(1 - e^{eV/k_{\rm B}T}\big) = e\,\big(|J_{\rm h}^{\rm gen}| + |J_{\rm e}^{\rm gen}|\big)\big(e^{eV/k_{\rm B}T} - 1\big) \quad . \end{aligned} \tag{6.63}$$

Note that $|J_{\rm h}^{\rm gen}|$ and $|J_{\rm e}^{\rm gen}|$ are each proportional to $\exp(-\Delta/k_{\rm B}T)$. The current-voltage relationship is sketched in Fig. 6.15. A p-n junction is thus a *current rectifier*. This is how a diode works: passing alternating current through such a junction yields a direct current.

6.4.3 MOSFETs and heterojunctions

In a metal, internal electric fields are efficiently screened and excess charge migrates rapidly to the surface, with charge density fluctuations attenuated exponentially as one enters the bulk. The Thomas-Fermi screening length, $\lambda_{\rm TF} = \big(4\pi e^2 g(\varepsilon_{\rm F})\big)^{-1/2}$, is short (a few Ångstroms) due to the large density of states at the Fermi level. In semiconductors, the Fermi level lies somewhere in the gap between valence and conduction bands, and the density of states at $\varepsilon_{\rm F}$ is quite low. Screening is due to thermally excited charge carriers, and since the carrier density is small in comparison to that in metals, the screening length is many lattice spacings.

Consider now a junction between a semiconductor and a metal, with an intervening insulating layer. This is called MIS, or metal-insulator-ssemiconductor, junction. If the metal is unbiased relative to the semiconductor, their chemical potentials will align. The situation for a p-type semiconductor - metal junction is depicted in the left panel of Fig. 6.16. Next consider the case in which the metal is biased negatively with respect to the semiconductor, *i.e.* the metal is placed at a negative voltage $-V$. There is then an electric field $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$ pointing *out* of the semiconductor. Electric fields point in the direction positive charges want to move, hence in this case valence holes are attracted to the interface, creating an *accumulation layer* of holes, as depicted in the right panel of Fig. 6.16. On the metallic side, electrons migrate to the interface for the same reason. *No charges move across the insulating barrier.* Thus, a dipole layer is created across the barrier, with the dipole moment pointing into the semiconductor. This creates an internal potential whose net difference $\phi_{\rm metal} - \phi_{\rm semiconductor}(-\infty) = V$ exactly cancels the applied bias. This condition in fact is what determines the width of the accumulation layer.

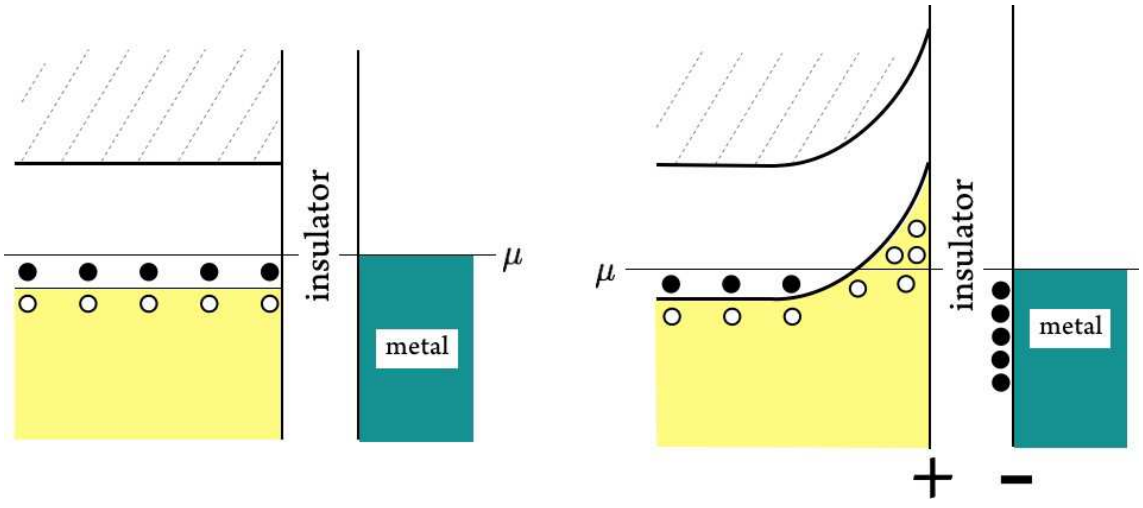


Figure 6.16: Junction between a p-type semiconductor and a metal. Left: Zero bias. Right: Metal biased negative with respect to semiconductor, creating an accumulation layer of holes and a net dipole moment at the interface.

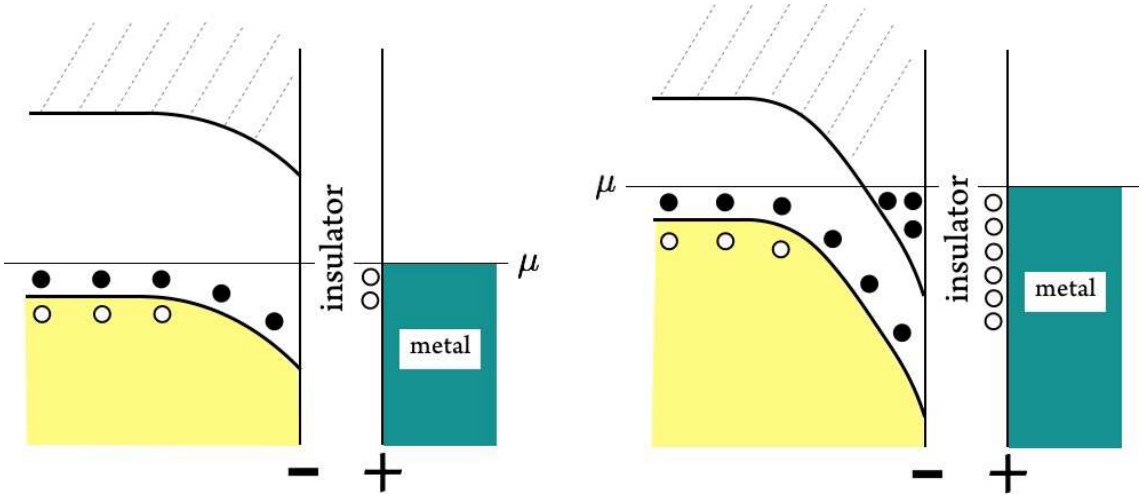


Figure 6.17: Junction between a p-type semiconductor and a metal. Left: Metal biased positive with respect to semiconductor, creating a space charge depletion layer. Right: Strong positive bias creates an inversion layer of n-type carriers in the p-type material.

What happens when the metal is biased positively? In this case, the electric field points into the semiconductor, and valence holes are repelled from the semiconductor surface, which is then negatively charged. This, in turn, repels electrons from the nearby metallic surface. The result is a space charge *depletion layer* in the semiconductor, which is devoid of charge carriers (*i.e.* valence holes). This situation is sketched in the left panel of Fig. 6.17.

Finally, what happens if the bias voltage on the metal exceeds the band gap? In this case, the field is so strong that not only are valence holes expelled from the surface, but conduction electrons are present, as shown in the right panel of Fig. 6.17. The presence of n-type carriers in a p-type semiconductor is known as *n-inversion*.

Remember this:

- *Accumulation* : presence of additional n-carriers in an n-type material, or additional p-carriers in a p-type material.
- *Depletion* : absence of n-carriers in an n-type material, or p-carriers in a p-type material.
- *Inversion* : presence of n-carriers in a p-type material, or p-carriers in an n-type material.

Inversion occurs when the presence of a depletion layer does not suffice to align the chemical potentials of the two sides of the junction.

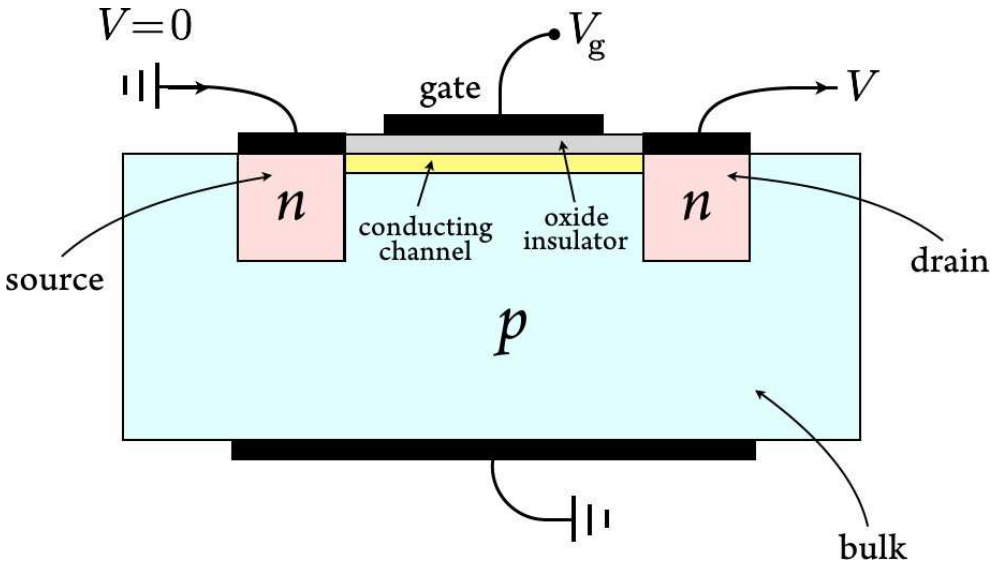


Figure 6.18: The MOSFET.

The MOSFET

A MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor) consists of two back-to-back p-n junctions, and, transverse to this, a gate-bulk-oxide capacitor. The situation is depicted in Fig. 6.18. If there is no gate voltage ($V_{\rm g} = 0$), then current will not flow at any bias voltage V because necessarily one of the p-n junction will be reverse-biased. The situation changes drastically if the gate is held at a high positive potential $V_{\rm g}$, for then an n-type accumulation layer forms at the bulk-gate interface, thereby connecting source and drain directly and resulting in a gate-controlled current flow. Although not shown in the figure, generally both source and drain are biased positively with respect to the bulk in order to avoid current leakage.

6.4.4 Heterojunctions

Potential uses of a junction formed from two distinct semiconductors were envisioned as early as 1951 by Shockley. Such devices, known as *heterojunctions*, have revolutionized the electronics industry and experimental solid state physics, the latter due to the advent of epitaxial technology which permits growth patterning to nearly atomic precision. Whereas the best inversion layer mobilities in Si MOSFETs are $\mu \approx 2 \times 10^4\,{\rm cm}^2/{\rm V}\cdot{\rm s}$, values as high as $10^7\,{\rm cm}^2/{\rm V}\cdot{\rm s}$ are possible in MBE-fabricated GaAs$-$Al$_x$Ga$_{1-x}$As heterostructures. There are three reasons for this:

(i) MBE (molecular beam epitaxy), as mentioned above, can produce layers which are smooth on an atomic scale. This permits exquisite control of layer thicknesses and doping profiles.

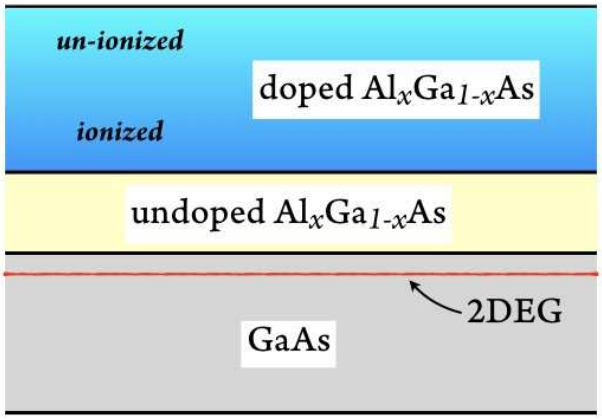


Figure 6.19: GaAs$-$Al$_x$Ga$_{1-x}$As heterojunction.

(ii) Use of ternary compounds such as Al$_x$Ga$_{1-x}$As makes for an excellent match in lattice constant across the heterojunction interface, *i.e.* on the order of or better than 1%. By contrast, the Si$-$SiO$_s$ interface is very poor, since SiO$_2$ is a glass.

(iii) By doping the Al$_x$Ga$_{1-x}$As layer far from the interface, Coulomb scattering between inversion layer electrons and dopant ions is suppressed.

Let's consider the chemical potential alignment problem in the case of an n-n heterojunction, sketched in Fig. 6.20. In the GaAs$-$Al$_x$Ga$_{1-x}$As heterojunction, GaAs has the smaller of the two band gaps. Initially there is a mismatch, as depicted in the left panel of the figure. By forming a depletion layer on the side with the larger band gap (Al$_x$Ga$_{1-x}$As), and an accumulation layer on the side with the smaller gap (GaAs), an internal potential $\phi(x)$ is established which aligns the chemical potentials.

Fig. 6.21 shows the phenomena of accumulation and inversion in different possible heterojunctions. There are four possibilities: (a) n-n, (b) p-p (c) n-p with the n-type material having the

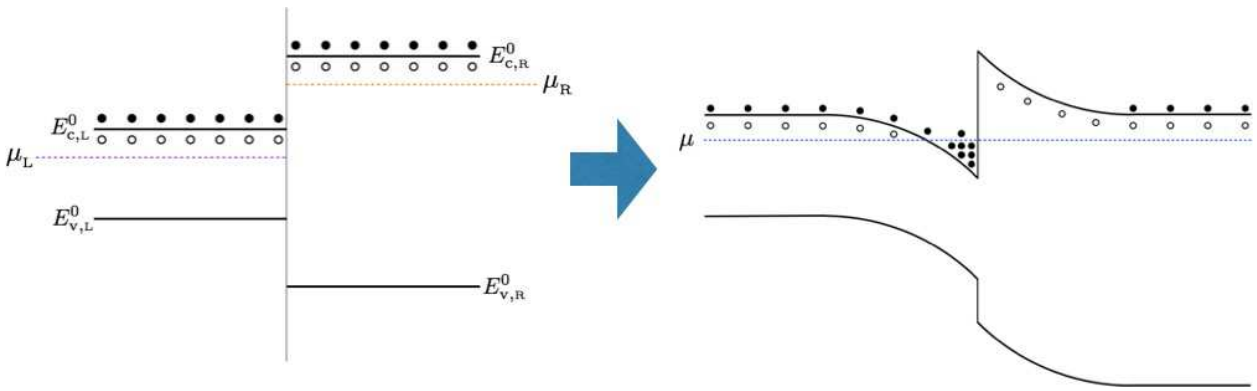


Figure 6.20: Accumulation layer formation in an n-n heterojunction.

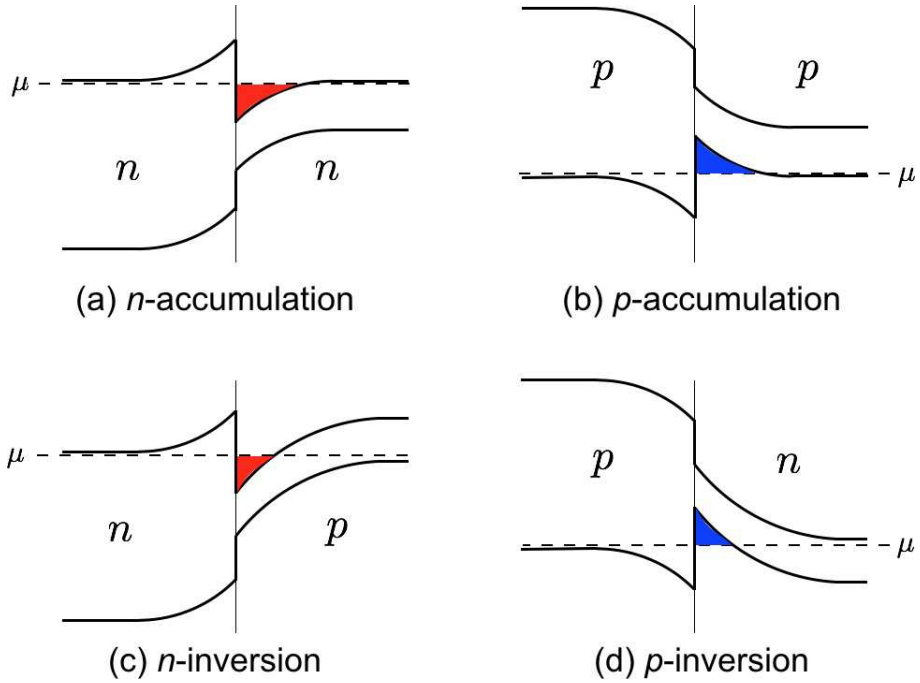


Figure 6.21: Accumulation and inversion in semiconductor heterojunctions. Red regions represent presence of conduction electrons. Blue regions represent presence of valence holes.

larger gap, and (d) n-p with the p-type material having the larger gap.

6.5 Insulators

An insulator is a system in which there is an energy gap for charged excitations at $T = 0$. Here we shall consider a subclass known as *band insulators*, *i.e.* materials which may be adequately idealized as noninteracting electrons in a crystalline potential with a Fermi level $\varepsilon_{\rm F}$ which lies in a band gap. Intrinsic semiconductors are then a subclass of band insulators. As we've seen above, extrinsic semiconductors can be modeled by including donor and acceptor levels to the intrinsic semiconductor band structure, and assuming that the dopant concentration is sufficiently low that the different donor/acceptor states associated with different ions may be considered noninteracting.

Insulators where the charge gap arises from strong electron-electron interactions are known as Mott-Hubbard insulators. Consider a tight-binding model with a single orbital per site, with the Hamiltonian

$$H = -\tfrac{1}{2}\sum_{i\neq j}\left(t_{\langle ij\rangle}\,c^\dagger_{i\sigma}\,c_{j\sigma} + t^*_{ij}\,c^\dagger_{j\sigma}\,c_{i\sigma}\right) + U\sum_i n_{i\uparrow}\,n_{i\downarrow} \quad , \tag{6.64}$$

The second term on the RHS imposes an energy cost U whenever there are two electrons on

the same site i. Consider now this model in which there is one electron per site. Assuming the crystal is a Bravais lattice, there is one band, and the Fermi level must cut through it in such a way that half the band is occupied and half is empty (since a filled band accommodates two electrons per site – one of each spin polarization). Thus, at $U = 0$ the system is a metal. But now consider the case where $U \gg W$, where W is the bandwidth at $U = 0$. Any state with one electron per site, with arbitrary spin polarization, *e.g.*, $|\,\psi\,\rangle = |\uparrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\cdots\rangle$ requires an energy $\Delta E \approx U - W$ to add an electron, since the added electron will necessarily result in double occupation of some site. This is called the *Mott-Hubbard gap*.

In this section, we will be concerned only with band insulators, and in particular with their dielectric properties. Mott-Hubbard systems are particularly interesting and important in condensed matter physics, but their analysis lies beyond the scope of this course.

6.5.1 Maxwell's equations in polarizable media

Maxwell's equations in the presence of sources are

$$\begin{aligned} \boldsymbol{\nabla}\cdot\mathbf{e} &= 4\pi\rho & \boldsymbol{\nabla}\times\mathbf{e} &= -\frac{1}{c}\,\frac{\partial\mathbf{b}}{\partial t} \\ \boldsymbol{\nabla}\cdot\mathbf{b} &= 0 & \boldsymbol{\nabla}\times\mathbf{b} &= \frac{4\pi}{c}\,\boldsymbol{j} + \frac{1}{c}\,\frac{\partial\mathbf{e}}{\partial t} \quad . \end{aligned} \tag{6.65}$$

Here, $\mathbf{e}(\boldsymbol{r})$, $\mathbf{b}(\boldsymbol{r})$, $\rho(\boldsymbol{r})$, and $\boldsymbol{j}(\boldsymbol{r})$ are the microscopic fields and sources. In a crystalline solid, the charge density $\rho(\boldsymbol{r})$ is a rapidly oscillating function of space, with positive delta-function like spikes at each ionic core and a negative contribution from the electron distribution which is localized for the core band states but more uniformly distributed between ions for the valence band states. The strongly fluctuating, microscopic description of electrodynamics at the atomic scale is not useful for our purposes[6].

Consider now a smoothing function $f(\boldsymbol{r})$ which satisfies $\int d^3r\, f(\boldsymbol{r}) = 1$, and define the smoothed fields

$$\boldsymbol{E}(\boldsymbol{r}) = \int d^3r'\; f(\boldsymbol{r}-\boldsymbol{r}')\,\mathbf{e}(\boldsymbol{r}') \qquad , \qquad \boldsymbol{B}(\boldsymbol{r}) = \int d^3r'\; f(\boldsymbol{r}-\boldsymbol{r}')\,\mathbf{b}(\boldsymbol{r}') \tag{6.66}$$

and the smoothed sources

$$\varrho(\boldsymbol{r}) = \int d^3r'\; f(\boldsymbol{r}-\boldsymbol{r}')\,\rho(\boldsymbol{r}') \qquad , \qquad \boldsymbol{J}(\boldsymbol{r}) = \int d^3r'\; f(\boldsymbol{r}-\boldsymbol{r}')\,\boldsymbol{j}(\boldsymbol{r}') \quad . \tag{6.67}$$

Now evaluate each of Eqns. 6.65 at $\boldsymbol{r}'$, multiply by $f(\boldsymbol{r}-\boldsymbol{r}')$, and integrate over $\boldsymbol{r}$. Integration

[6] In the following description of macroscopic electrodynamics, we follow the pellucid discussion in A. Garg, *Classical Electrodynamics in a Nutshell* (Princeton, 2012).

by parts then yields the *macroscopic Maxwell equations*,

$$\begin{aligned} \boldsymbol{\nabla}\cdot\boldsymbol{E} &= 4\pi\varrho & \boldsymbol{\nabla}\times\boldsymbol{E} &= -\frac{1}{c}\frac{\partial\boldsymbol{B}}{\partial t} \\ \boldsymbol{\nabla}\cdot\boldsymbol{B} &= 0 & \boldsymbol{\nabla}\times\boldsymbol{B} &= \frac{4\pi}{c}\boldsymbol{J} + \frac{1}{c}\frac{\partial\boldsymbol{E}}{\partial t} \quad . \end{aligned} \tag{6.68}$$

These take the same form as the microscopic Eqns. 6.65, however, the meanings of the fields and the sources are quite different. Unfortunately, without a theory for the sources ϱ and $\boldsymbol{J}$, the above version of the macroscopic Maxwell equations is rather useless.

Macroscopic charge density

In a solid, charges may be *free*, such as conduction electrons or valence holes, or *bound*, such as the core electrons. Free charges may execute macroscopic motion in response to external fields, while bound charges remain local. To grasp the significance of bound charges, consider a system of neutral atoms or molecules, each of which has a dipole moment $\boldsymbol{d}$. If their number density is n, then the dipole moment per unit volume is $\boldsymbol{P} = n\boldsymbol{d}$, which is called the *electrical polarization*.

The electrical potential $\phi(\boldsymbol{r})$ due to a dipole at the origin is given by $\phi(\boldsymbol{r}) = \boldsymbol{d}\cdot\boldsymbol{r}/r^3$. Now consider a region Ω containing dipolar matter is then

$$\begin{aligned} \phi(\boldsymbol{r}) &= \int\limits_\Omega d^3r'\,\frac{\boldsymbol{P}(\boldsymbol{r}')\cdot(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3} = \int\limits_\Omega d^3r'\,\boldsymbol{P}(\boldsymbol{r}')\cdot\boldsymbol{\nabla}'\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} \\ &= \int\limits_{\partial\Omega} d^2r'\,\frac{\boldsymbol{P}(\boldsymbol{r}')\cdot\hat{\boldsymbol{n}}'}{|\boldsymbol{r}-\boldsymbol{r}'|} - \int\limits_\Omega d^3r'\,\frac{\boldsymbol{\nabla}'\cdot\boldsymbol{P}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \\ &\equiv \int\limits_{\partial\Omega} d^2r'\,\frac{\sigma_{\rm pol}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} - \int\limits_\Omega d^3r'\,\frac{\varrho_{\rm pol}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \quad , \end{aligned} \tag{6.69}$$

where $\sigma_{\rm pol} = \boldsymbol{P}\cdot\hat{\boldsymbol{n}}$ is the surface charge density and $\varrho_{\rm pol} = -\boldsymbol{\nabla}\cdot\boldsymbol{P}$ is the polarization charge density. We may now write $\varrho = \varrho_{\rm free} + \varrho_{\rm pol}$, and defining the electrical displacement field $\boldsymbol{D} \equiv \boldsymbol{E} + 4\pi\boldsymbol{P}$, we obtain the relation

$$\boldsymbol{\nabla}\cdot\boldsymbol{D} = 4\pi\varrho_{\rm free} \quad . \tag{6.70}$$

The polarization charge density is the bound charge density. While a body may be electrically neutral overall, the local charge density may vary. One writes $\boldsymbol{D} = \epsilon\boldsymbol{E}$, where ϵ is the *electric permittivity* (dielectric constant). Note that

$$\boldsymbol{P} = \frac{\epsilon-1}{4\pi}\,\boldsymbol{E} \quad . \tag{6.71}$$

Macroscopic current density

The macroscopic current density may be written $\boldsymbol{j} = \boldsymbol{j}_{\rm free} + \boldsymbol{j}_{\rm pol} + \boldsymbol{j}_{\rm mag}$, *i.e.* as a sum of contributions from the motion of free charges, from polarization currents, and from Amperean (magnetization) currents[7]. From $\varrho_{\rm pol} = -\boldsymbol{\nabla}\cdot\boldsymbol{P}$, we have that $\boldsymbol{j}_{\rm pol} = \partial\boldsymbol{P}/\partial t$, in order that the continuity equation $\partial_t\varrho_{\rm pol} + \boldsymbol{\nabla}\cdot\boldsymbol{j}_{\rm pol} = 0$ be satisfied.

Let $\boldsymbol{M}(\boldsymbol{r})$ be the magnetic dipole moment density. The vector potential due to a magnetic dipole $\boldsymbol{m}$ located at the origin is $\boldsymbol{A}(\boldsymbol{r}) = \boldsymbol{m}\times\boldsymbol{r}/r^3$, hence

$$\begin{aligned}\boldsymbol{A}(\boldsymbol{r}) &= \int\limits_{\Omega} d^3r'\,\frac{\boldsymbol{M}(\boldsymbol{r}')\times(\boldsymbol{r}-\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|^3} = \int\limits_{\Omega} d^3r'\,\boldsymbol{M}(\boldsymbol{r}')\times\boldsymbol{\nabla}'\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} \\ &= \int\limits_{\partial\Omega} d^2r'\,\frac{\boldsymbol{M}(\boldsymbol{r}')\times\hat{\boldsymbol{n}}'}{|\boldsymbol{r}-\boldsymbol{r}'|} - \int\limits_{\Omega} d^3r'\,\frac{\boldsymbol{\nabla}'\times\boldsymbol{P}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \\ &\equiv \frac{1}{c}\int\limits_{\partial\Omega} d^2r'\,\frac{\boldsymbol{K}_{\rm mag}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} - \frac{1}{c}\int\limits_{\Omega} d^3r'\,\frac{\boldsymbol{j}_{\rm mag}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\quad ,\end{aligned} \tag{6.72}$$

where $\boldsymbol{K}_{\rm mag} \equiv c\,\boldsymbol{M}\times\hat{\boldsymbol{n}}$ is the surface current density and $\boldsymbol{j}_{\rm mag} = c\,\boldsymbol{\nabla}\times\boldsymbol{M}$ is the magnetization volume current. Defining the field $\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M}$, then we have

$$\boldsymbol{\nabla}\times\boldsymbol{H} = \frac{4\pi}{c}\,\boldsymbol{j}_{\rm free} + \frac{1}{c}\,\frac{\partial\boldsymbol{D}}{\partial t}\quad . \tag{6.73}$$

One calls $\boldsymbol{E}$ the *electric field* and $\boldsymbol{D}$ the *(electric) displacement field*. Garg (2012) calls $\boldsymbol{B}$ the magnetic field and $\boldsymbol{H}$ the *magnetizing field*, but other texts (*e.g.*, Landau and Lifshitz vol. 8) call $\boldsymbol{H}$ the magnetic field and $\boldsymbol{B}$ the *magnetic induction field*.

6.5.2 Clausius-Mossotti relation

In isotropic systems, we write $\boldsymbol{D} = \epsilon\boldsymbol{E}$ and $\boldsymbol{B} = \mu\boldsymbol{H}$. How to connect the electric permittivity (dielectric constant) ϵ and the magnetic permeability μ to microscopic quantities such the atomic polarizability α which relates the atomic or molecular dipole moment $\boldsymbol{d} = \alpha\mathbf{e}$ to the microscopic local electric field $\mathbf{e}$? We begin by writing $\mathbf{e} = \mathbf{e}_{\rm near} + \mathbf{e}_{\rm far}$ as a sum of contributions from nearby atoms and those which are far away. The distinction is that the far field is such that it is equal, to high precision, to its local average, *i.e.* $\mathbf{e}_{\rm far}(\boldsymbol{r}) = \boldsymbol{E}_{\rm far}(\boldsymbol{r}) = \int d^3r'\,f(\boldsymbol{r}-\boldsymbol{r}')\,\mathbf{e}_{\rm far}(\boldsymbol{r}')$, provided $r > R$. We expect R should be on the order of a hundred lattice spacings. Thus, $\boldsymbol{E}_{\rm far}(\boldsymbol{r})$ is the field that would exist at a point $\boldsymbol{r}$ if one were to remove all the material from a radius R about this point. The corresponding macroscopic near field, $\boldsymbol{E}_{\rm near}$, is that due to a

[7] There is a fourth term, $\boldsymbol{j}_{\rm conv}$, due to convection currents, given by $\boldsymbol{j}_{\rm conv} = (\varrho_{\rm free} + \varrho_{\rm pol})\boldsymbol{u}$, where $\boldsymbol{u}$ is the velocity of the convective flow. Convection currents arise in liquids and gases, but not in solids.

uniformly polarized sphere of dipole density $\boldsymbol{P}$, which from elementary electrostatics is given by $\boldsymbol{E}_{\rm near} = -\frac{4\pi}{3}\boldsymbol{P}$. Since $\boldsymbol{E} = \boldsymbol{E}_{\rm near} + \boldsymbol{E}_{\rm far}$, we have that

$$\boldsymbol{E}_{\rm far} = \boldsymbol{E} + \tfrac{4\pi}{3}\boldsymbol{P} = \frac{\epsilon+2}{3}\,\boldsymbol{E} \tag{6.74}$$

and therefore

$$\mathbf{e} = \boldsymbol{E}_{\rm far} + \mathbf{e}_{\rm near} = \frac{\epsilon+2}{3}\,\boldsymbol{E} + \mathbf{e}_{\rm near} \quad . \tag{6.75}$$

In an isotropic system such as a liquid, or in a crystal with cubic symmetry, one has that $\mathbf{e}_{\rm near}(\boldsymbol{r}) = 0$ if $\boldsymbol{r}$ is an atomic position about which the symmetry is manifested. We then have

$$\boldsymbol{P} = \frac{\epsilon-1}{4\pi}\,\boldsymbol{E} = n\alpha\mathbf{e} = \left(\frac{\epsilon+2}{3}\right)n\alpha\,\boldsymbol{E} \quad , \tag{6.76}$$

and therefore

$$\frac{\epsilon-1}{\epsilon+2} = \frac{4\pi n\alpha}{3} \quad , \tag{6.77}$$

a result known as the *Clausius-Mossotti relation*. Solving for ϵ,

$$\epsilon = 1 + \frac{4\pi n\alpha}{1 - \frac{4\pi}{3}n\alpha} \quad . \tag{6.78}$$

If $|n\alpha| \ll 1$, which is typical, then we have

$$\epsilon \approx 1 + 4\pi n\alpha \quad . \tag{6.79}$$

6.5.3 Theory of atomic polarizability

Consider an atom with Z valence electrons. A crude classical model of atomic polarizability by writing $\boldsymbol{F} = m\boldsymbol{a}$ for each valence electron, *viz.*

$$m_{\rm e}\ddot{\boldsymbol{r}} = -k\boldsymbol{r} - e\,\mathbf{e}(t) \quad , \tag{6.80}$$

where $\boldsymbol{r}$ is the displacement from equilibrium and $\mathbf{e}$ is the microscopic electric field. The dipole moment is then $\boldsymbol{d} = Ze\,\boldsymbol{r}$. If all quantities are taken to vary as $e^{-i\omega t}$, we obtain $\hat{\boldsymbol{d}}(\omega) = \alpha(\omega)\,\hat{\mathbf{e}}(\omega)$, with

$$\alpha(\omega) = \frac{Ze^2}{m_{\rm e}(\omega_0^2 - \omega^2)} \quad , \tag{6.81}$$

where we've defined $k \equiv m_{\rm e}\omega_0^2$. Quantum mechanically, we expect $\hbar\omega_0$ to be on the order of an atomic transition energy, which is typically on the order of electron volts, and since $\hbar = 6.58 \times 10^{-16}\,{\rm eV}\cdot{\rm s}$, this corresponds to frequencies on the order of $10^{16}\,{\rm Hz}$. For $\omega \ll \omega_0$, we have $\alpha = Ze^2/m_{\rm e}\omega_0^2$, which is frequency-independent. This is valid up to frequencies $\nu_0 = \omega_0/2\pi$ which are in the ultraviolet regime.

In polar crystals, the unit cell consists of positively and negatively charged ions. Examples include III-V and II-VI semiconductors, for example. For the sake of simplicity, we analyze the case of one positive and one negative ion per cell, with charges $\pm q$, respectively. The equations of motion for the ionic vibrations are

$$\begin{aligned} M_+ \ddot{u}_+^\alpha(\boldsymbol{R}) &= -\sum_{\boldsymbol{R}'}\sum_\beta \Big[\Phi_{++}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\,u_+^\beta(\boldsymbol{R}') + \Phi_{+-}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\,u_-^\beta(\boldsymbol{R}')\Big] + q\,\mathrm{e}^\alpha \\ M_- \ddot{u}_-^\alpha(\boldsymbol{R}) &= -\sum_{\boldsymbol{R}'}\sum_\beta \Big[\Phi_{-+}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\,u_+^\beta(\boldsymbol{R}') + \Phi_{--}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}')\,u_-^\beta(\boldsymbol{R}')\Big] - q\,\mathrm{e}^\alpha \quad , \end{aligned} \tag{6.82}$$

where

$$\Phi_{\eta\eta'}^{\alpha\beta}(\boldsymbol{R}-\boldsymbol{R}') = \frac{\partial^2 U}{\partial u_\eta^\alpha(\boldsymbol{R})\,u_{\eta'}^\beta(\boldsymbol{R}')} \tag{6.83}$$

are force constants for the lattice potential. Consider the $\boldsymbol{k}=0$ mode, where $\boldsymbol{u}_\pm(\boldsymbol{R})$ is independent of $\boldsymbol{R}$, *i.e.* all unit cell motions are in phase. Subtracting the second of the above equations from the first, we obtain

$$\begin{aligned} \ddot{u}_+^\alpha - \ddot{u}_-^\alpha = &-\frac{1}{M_+}\sum_{\boldsymbol{R}}\Phi_{++}^{\alpha\beta}(\boldsymbol{R})\,u_+^\beta - \frac{1}{M_+}\sum_{\boldsymbol{R}}\Phi_{+-}^{\alpha\beta}(\boldsymbol{R})\,u_-^\beta \\ &+ \frac{1}{M_-}\sum_{\boldsymbol{R}}\Phi_{-+}^{\alpha\beta}(\boldsymbol{R})\,u_+^\beta + \frac{1}{M_-}\sum_{\boldsymbol{R}}\Phi_{--}^{\alpha\beta}(\boldsymbol{R})\,u_-^\beta + \frac{q\,\mathrm{e}^\alpha}{M_+} + \frac{q\,\mathrm{e}^\alpha}{M_-} \quad . \end{aligned} \tag{6.84}$$

However, the fact that there is no restoring force for a uniform translation of the crystal requires

$$\sum_{\boldsymbol{R}}\Big[\Phi_{++}^{\alpha\beta}(\boldsymbol{R}) + \Phi_{+-}^{\alpha\beta}(\boldsymbol{R})\Big] = \sum_{\boldsymbol{R}}\Big[\Phi_{-+}^{\alpha\beta}(\boldsymbol{R}) + \Phi_{--}^{\alpha\beta}(\boldsymbol{R})\Big] = 0 \quad , \tag{6.85}$$

and therefore, with $\boldsymbol{\delta} \equiv \boldsymbol{u}_+ - \boldsymbol{u}_-$, and assuming cubic symmetry,

$$\ddot{\delta}^\alpha = -\overbrace{\sum_{\boldsymbol{R}}\left(\frac{\Phi_{++}^{\alpha\beta}(\boldsymbol{R})}{M_+} + \frac{\Phi_{--}^{\alpha\beta}(\boldsymbol{R})}{M_-}\right)}^{\equiv\,\bar\omega^2\,\delta^{\alpha\beta}}\delta^\beta + \frac{q\,\mathrm{e}^\alpha}{M_-} \quad . \tag{6.86}$$

We may rewrite this as

$$\ddot{\boldsymbol{\delta}} = -\bar\omega^2\,\boldsymbol{\delta} + \frac{q\,\mathbf{e}}{M^*} \quad , \tag{6.87}$$

where $M^* = M_+M_-/(M_+ + M_-)$ is the reduced mass. Here $\bar\omega$ is the frequency of the $\boldsymbol{k}=0$ optical phonon. mode The polarization density is then $\boldsymbol{P} = q\boldsymbol{\delta}/v_0$, where v_0 is the unit cell volume. Solving for an oscillating electric field $\mathbf{e}\,e^{-i\omega t}$, we obtain

$$\boldsymbol{\delta}(t) = \frac{q\,\mathbf{e}}{M^*}\cdot\frac{e^{-i\omega t}}{\bar\omega^2-\omega^2} \quad . \tag{6.88}$$

We conclude that the *displacement polarizability* is

$$\alpha_{\rm disp} = \frac{q^2}{M^*(\bar\omega^2 - \omega^2)} \quad . \tag{6.89}$$

Now in addition to the displacement polarizability, we also have to add in the individual atomic polarizabilities of the positive and negative ions, hence our final result is

$$\alpha(\omega) = \alpha_+ + \alpha_- + \alpha_{\rm disp}(\omega) \quad . \tag{6.90}$$

Thus, from the Clausius-Mossotti relation Eqn. 6.77, we have

$$\frac{\epsilon(\omega) - 1}{\epsilon(\omega) + 2} = \frac{4\pi}{3v_0}\left(\alpha_+ + \alpha_- + \frac{q^2}{M^*(\bar\omega^2 - \omega^2)}\right) \quad , \tag{6.91}$$

provided $\omega \ll \omega_0$. We then have, at $\omega = 0$,

$$\frac{\epsilon_0 - 1}{\epsilon_0 + 2} = \frac{4\pi}{3v}\left(\alpha_+ + \alpha_- + \frac{q^2}{M^*\bar\omega^2}\right) \quad , \tag{6.92}$$

while for $\bar\omega \ll \omega \ll \omega_0$, which we call 'infinite' frequency for our purposes,

$$\frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} = \frac{4\pi}{3v}\left(\alpha_+ + \alpha_-\right) \quad . \tag{6.93}$$

These equations allow us to write

$$\epsilon(\omega) = \epsilon_\infty \cdot \frac{\omega^2 - \omega_{\rm L}^2}{\omega^2 - \omega_{\rm T}^2} \quad , \tag{6.94}$$

where

$$\omega_{\rm T} = \left(\frac{\epsilon_\infty + 2}{\epsilon_0 + 2}\right)^{1/2}\bar\omega \qquad , \qquad \omega_{\rm L} = \left(\epsilon_0/\epsilon_\infty\right)^{1/2}\omega_{\rm T} \quad . \tag{6.95}$$

Note that $\epsilon(\omega_{\rm L}) = 0$ and $\epsilon(\omega_{\rm T}) = \infty$, and that

$$\left(\frac{\omega_{\rm L}}{\omega_{\rm T}}\right)^2 = \frac{\epsilon_0}{\epsilon_\infty} > 1 \quad , \tag{6.96}$$

a result known as the *Lyddane-Sachs-Teller relation*. The behavior of $\epsilon(\omega)$ is sketched in the left panel of Fig. 6.22.

6.5.4 Electromagnetic waves in a polar crystal

Consider now the propagation of electromagnetic waves in a polar crystal. Assuming the absence of free charges, we have $\boldsymbol{\nabla}\cdot\boldsymbol{D} = 0$ and $\boldsymbol{\nabla}\times\boldsymbol{E} = -c^{-1}\partial_t\boldsymbol{B}$ governing the macroscopic fields. We will ignore the $c^{-1}\partial_t\boldsymbol{B}$ term and justify this later on. A plane wave solution with $\boldsymbol{E}$ and $\boldsymbol{D}$ both proportional to $\exp(i\boldsymbol{k}\cdot\boldsymbol{r})$ thereby requires $\boldsymbol{k}\cdot\boldsymbol{D} = \boldsymbol{k}\times\boldsymbol{E} = 0$, which has no nontrivial solutions if $\boldsymbol{D} = \epsilon(\omega)\,\boldsymbol{E}$. *I.e.* either $\boldsymbol{E} = 0$ or $\boldsymbol{D} = 0$. Thus we have the following two possibilities:

- *longitudinal mode* : $\boldsymbol{E} \parallel \boldsymbol{k}$ with $\epsilon = 0$, hence $\omega = \omega_{\rm L}$ and $\boldsymbol{D} = 0$.
- *transverse mode* : $\boldsymbol{D} \perp \boldsymbol{k}$ with $\epsilon = \infty$, hence $\omega = \omega_{\rm T}$ and $\boldsymbol{E} = 0$.

These conclusions hold valid in the $\boldsymbol{k} \to 0$ limit. To find the dispersion for general $\boldsymbol{k}$, we need to solve Eqns. 6.82 under general conditions, *i.e.* not assuming all the unit cells are in phase. This yields a dispersion as shown in the right panel of Fig. 6.22.

Why were we allowed to drop the $c\,\partial_t\boldsymbol{B}$ term in Faraday's equation? This is because it is negligible in the limit $ck \gg \omega$. Since optical frequencies are on the order of that of zone edge phonons, this means we must satisfy $k \gg (\pi/a)\cdot(s/c)$, where a is the lattice spacing, s is the acoustic phonon velocity, and c the speed of light. Since $s/c \sim 10^{-5} - 10^{-4}$, we are in good shape provided k is not extremely close to the zone center. Finally, since the reflectivity is

$$R(\omega) = \left|\frac{\sqrt{\epsilon(\omega)}-1}{\sqrt{\epsilon(\omega)}+1}\right|^2 \quad , \tag{6.97}$$

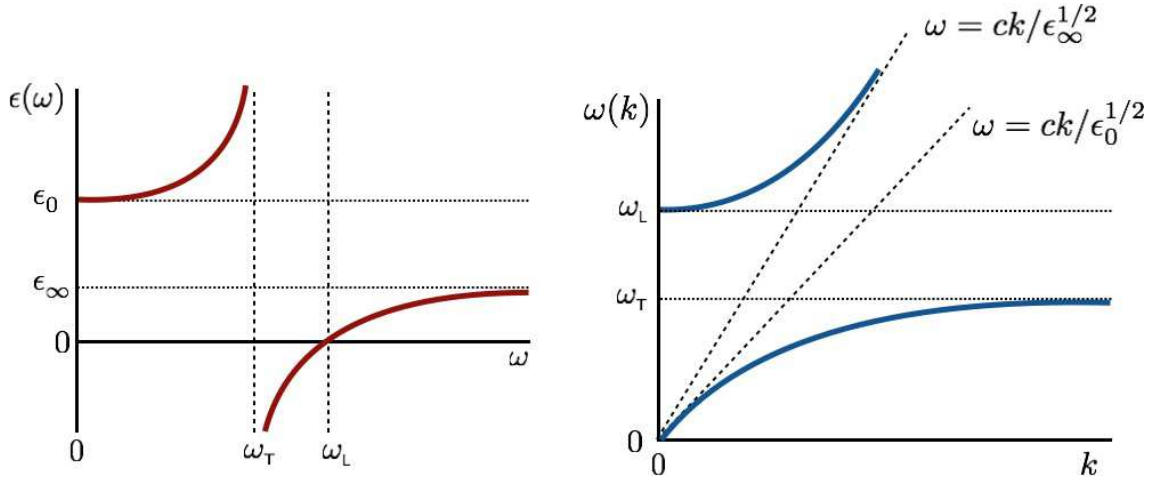


Figure 6.22: Left: dielectric function $\epsilon(\omega)$ in a polar crystal. Right: solution to the equation $\omega = ck/\epsilon^{1/2}(\omega)$ showing the polariton dispersion branches $\omega_\pm(k)$.

for $\omega \in [\omega_T, \omega_L]$ the dielectric function $\epsilon(\omega)$ is purely imaginary and thus the crystal is purely reflecting.

Nota bene : The "charge" q is poorly defined, since it is spread out in a continuous distribution rather than a Dirac delta function. Thus, Eqn. 6.94 and the LST relation are much more useful than Eqn. 6.91, since we can always measure ϵ_0 and ϵ_∞.

Chapter 7

Mesoscopia

7.1 Introduction

Current nanofabrication technology affords us the remarkable opportunity to study condensed matter systems on an unprecedented small scale. For example, small electron boxes known as *quantum dots* have been fabricated, with characteristic size ranging from 10 nm to 1 μm; the smallest quantum dots can hold as few as one single electron, while larger dots can hold thousands. In systems such as these, one can probe discrete energy level spectra associated with quantization in a finite volume. Oftentimes systems are so small that Bloch's theorem and the theoretical apparatus of Boltzmann transport are of dubious utility.

7.2 The Landauer Formula

Consider a disordered one-dimensional wire connected on each end to reservoirs at fixed chemical potential $\mu_{\rm L}$ and $\mu_{\rm R}$. For the moment, let us consider only a single spin species, or imagine that the spins are completely polarized. Suppose further that $\mu_{\rm L} > \mu_{\rm R}$, so that a current I flows from the left reservoir (L) to the right reservoir (R). Next, consider a cross-sectional surface Σ just to the right of the disordered region. We calculate the current flowing past this surface as a sum over three terms:

$$I_\Sigma = -e \int d\varepsilon \, \mathcal{N}(\varepsilon) \, v(\varepsilon) \Big\{ T(\varepsilon) \, f(\varepsilon - \mu_{\rm L}) + R'(\varepsilon) \, f(\varepsilon - \mu_{\rm R}) - f(\varepsilon - \mu_{\rm R}) \Big\} \quad . \tag{7.1}$$

Here, $\mathcal{N}(\varepsilon)$ is the density of states in the leads per spin degree of freedom, and corresponding to motion in a given direction (right or left but not both); $v(\varepsilon)$ is the velocity, and $f(\varepsilon - \mu_{\rm L,R})$ are the respective Fermi distributions. $T(\varepsilon)$ is the *transmission probability* that electrons of energy ε emerging from the left reservoir will end up in right reservoir; $R'(\varepsilon)$ is the *reflection probability*

that electrons emerging from right reservoir will return to the right reservoir. The three terms on the right hand side of (7.1) correspond, respectively, to: (i) electrons emerging from L which make it through the wire and are deposited in R, (ii) electrons emerging from R which fail to 'swim upstream' to L and are instead reflected back into R, and (iii) all electrons emerging from reservoir R (note this contribution is of opposite sign). The transmission and reflection probabilities are obtained by solving for the quantum mechanical scattering due to the disordered region. If the incoming *flux amplitudes* from the left and right sides are i and i', respectively, and the outgoing flux amplitudes on those sides o' and o, linearity of the Schrödinger equation requires that

$$\begin{pmatrix} o' \\ o \end{pmatrix} = \mathcal{S} \begin{pmatrix} i \\ i' \end{pmatrix} \qquad ; \qquad \mathcal{S} = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix} \quad . \tag{7.2}$$

The matrix $\mathcal{S}$ is known as the *scattering matrix* (or $\mathcal{S}$-matrix, for short). The S-matrix elements r, t, *etc.* are reflection and transmission *amplitudes*. The reflection and transmission *probabilities* are given by

$$\begin{aligned} R &= |r|^2 \qquad & T' &= |t'|^2 \\ T &= |t|^2 \qquad & R' &= |r'|^2 \quad . \end{aligned} \tag{7.3}$$

Going back to (7.1), let us assume that we are close to equilibrium, so the difference $\mu_{\rm R} - \mu_{\rm L}$ in chemical potentials is slight. We may then expand

$$f(\varepsilon - \mu_{\rm R}) = f(\varepsilon - \mu_{\rm L}) + f'(\varepsilon - \mu_{\rm L})\,(\mu_{\rm L} - \mu_{\rm R}) + \ldots \tag{7.4}$$

and obtain the result

$$\begin{aligned} I &= e(\mu_{\rm R} - \mu_{\rm L}) \int d\varepsilon\, \mathcal{N}(\varepsilon)\, v(\varepsilon) \left(-\frac{\partial f^0}{\partial \varepsilon}\right) T(\varepsilon) \\ &= \frac{e}{h}(\mu_{\rm R} - \mu_{\rm L}) \int d\varepsilon \left(-\frac{\partial f^0}{\partial \varepsilon}\right) T(\varepsilon) \quad , \end{aligned} \tag{7.5}$$

valid to lowest order in $(\mu_{\rm R} - \mu_{\rm L})$. We have invoked here a very simple, very important result for the one-dimensional density of states. Considering only states moving in a definite direction (left or right) and with a definite spin polarization (up or down), we have

$$\mathcal{N}(\varepsilon)\, d\varepsilon = \frac{dk}{2\pi} \quad \Longrightarrow \quad \mathcal{N}(\varepsilon) = \frac{1}{2\pi}\frac{dk}{d\varepsilon} = \frac{1}{h v(\varepsilon)} \tag{7.6}$$

where $h = 2\pi\hbar$ is Planck's constant. Thus, there is a remarkable cancellation in the product $\mathcal{N}(\varepsilon)\, v(\varepsilon) = h^{-1}$. Working at $T = 0$, we therefore obtain

$$I = \frac{e}{h}(\mu_{\rm R} - \mu_{\rm L})\, T(\varepsilon_{\rm F}) \quad , \tag{7.7}$$

where $T(\varepsilon_{\rm F})$ is the transmission probability at the Fermi energy. The chemical potential varies with voltage according to $\mu(V) = \mu(0) - eV$, hence the *conductance* $G = I/V$ is found to be

$$\begin{aligned} G &= \frac{e^2}{h}\, T(\varepsilon_{\rm F}) \qquad \text{(per spin channel)} \\ &= \frac{2e^2}{h}\, T(\varepsilon_{\rm F}) \qquad \text{(spin degeneracy included)} \end{aligned} \tag{7.8}$$

The quantity h/e^2 is a conveniently measurable $25,813\,\Omega$.

We conclude that *conductance is transmission* - G is e^2/h times the transmission probability $T(\varepsilon_{\rm F})$ with which an electron at the Fermi level passes through the wire. This has a certain intuitive appeal, since clearly if $T(\varepsilon_{\rm F}) = 0$ we should expect $G = 0$. However, two obvious concerns should be addressed:

- The power dissipated should be $P = I^2R = V^2G$. Yet the scattering in the wire is assumed to be purely elastic. Hence no dissipation occurs within the wire at all, and the Poynting vector immediately outside the wire must vanish. What, then, is the source of the dissipation?
- For a perfect wire, $T(\varepsilon_{\rm F}) = 1$, and $G = e^2/h$ (per spin) is finite. Shouldn't a perfect (*i.e.* not disordered) wire have *zero* resistance, and hence *infinite* conductance?

The answer to the first of these riddles is simple – all the dissipation takes place in the R reservoir. When an electron makes it through the wire from L to R, it deposits its excess energy $\mu_{\rm L} - \mu_{\rm R}$ in the R reservoir. The mechanism by which this is done is not our concern – we only need assume that there is *some* inelastic process (*e.g.* electron-phonon scattering, electron-electron scattering, *etc.*) which acts to equilibrate the R reservoir.

The second riddle is a bit more subtle. One solution is to associate the resistance h/e^2 of a perfect wire with the *contact resistance* due to the leads. The intrinsic conductance of the wire $G_{\rm i}$ is determined by assuming the wire resistance and contact resistances are in series:

$$G^{-1} = G_{\rm i}^{-1} + \frac{h}{e^2} \quad \Longrightarrow \quad G_{\rm i} = \frac{e^2}{h}\,\frac{T(\varepsilon_{\rm F})}{1-T(\varepsilon_{\rm F})} = \frac{e^2}{h}\,\frac{T(\varepsilon_{\rm F})}{R(\varepsilon_{\rm F})} \quad , \tag{7.9}$$

where $G_{\rm i}$ is the intrinsic conductance of the wire, per spin channel. Now we see that when $T(\varepsilon_{\rm F}) \to 1$ the intrinsic conductance diverges: $G_{\rm i} \to \infty$. When $T \ll 1$, $G_{\rm i} \approx G = (e^2/h)T$. This result (7.9) is known as the *Landauer Formula*.

To derive this result in a more systematic way, let us assume that the disordered segment is connected to the left and right reservoirs by perfect leads, and that the leads are not in equilibrium at chemical potentials $\mu_{\rm L}$ and $\mu_{\rm R}$ but instead at $\tilde\mu_{\rm L}$ and $\tilde\mu_{\rm R}$. To determine $\tilde\mu_{\rm L}$ and $\tilde\mu_{\rm R}$, we

compute the number density (per spin channel) in the leads,

$$\begin{aligned} n_{\rm L} &= \int d\varepsilon\, \mathcal{N}(\varepsilon) \Big\{ \big[1 + R(\varepsilon)\big] f(\varepsilon - \mu_{\rm L}) + T'(\varepsilon)\, f(\varepsilon - \mu_{\rm R}) \Big\} \\ n_{\rm R} &= \int d\varepsilon\, \mathcal{N}(\varepsilon) \Big\{ T(\varepsilon) f(\varepsilon - \mu_{\rm L}) + \big[1 + R'(\varepsilon)\big]\, f(\varepsilon - \mu_{\rm R}) \Big\} \end{aligned} \tag{7.10}$$

and associate these densities with chemical potentials $\tilde\mu_{\rm L}$ and $\tilde\mu_{\rm R}$ according to

$$\begin{aligned} n_{\rm L} &= 2 \int d\varepsilon\, \mathcal{N}(\varepsilon)\, f(\varepsilon - \tilde\mu_{\rm L}) \\ n_{\rm R} &= 2 \int d\varepsilon\, \mathcal{N}(\varepsilon)\, f(\varepsilon - \tilde\mu_{\rm R}) \quad , \end{aligned} \tag{7.11}$$

where the factor of two accounts for both directions of motion. To lowest order, then, we obtain

$$\begin{aligned} 2(\mu_{\rm L} - \tilde\mu_{\rm L}) = (\mu_{\rm L} - \mu_{\rm R})\, T' \quad &\Longrightarrow \quad \tilde\mu_{\rm L} = \mu_{\rm L} + \tfrac{1}{2} T' \, (\mu_{\rm R} - \mu_{\rm L}) \\ 2(\mu_{\rm R} - \tilde\mu_{\rm R}) = (\mu_{\rm R} - \mu_{\rm L})\, T \quad &\Longrightarrow \quad \tilde\mu_{\rm R} = \mu_{\rm R} + \tfrac{1}{2} T \, (\mu_{\rm L} - \mu_{\rm R}) \end{aligned} \tag{7.12}$$

and therefore

$$\begin{aligned} (\tilde\mu_{\rm L} - \tilde\mu_{\rm R}) &= \big(1 - \tfrac{1}{2}T - \tfrac{1}{2}T'\big)\, (\mu_{\rm L} - \mu_{\rm R}) \\ &= (1 - T)\, (\mu_{\rm L} - \mu_{\rm R}) \quad , \end{aligned} \tag{7.13}$$

where the last equality follows from unitarity ($\mathcal{S}^\dagger\mathcal{S} = \mathcal{S}\mathcal{S}^\dagger = 1$). There are two experimental configurations to consider:

- *Two probe measurement* – Here the current leads are also used as voltage leads. The voltage difference is $\Delta V = (\mu_{\rm R} - \mu_{\rm L})/e$ and the measured conductance is given by the expression $G_{\rm 2-probe} = (e^2/h)\, T(\varepsilon_{\rm F})$.
- *Four probe measurement* – Separate leads are used for current and voltage probes. The observed voltage difference is $\Delta V = (\tilde\mu_{\rm R} - \tilde\mu_{\rm L})/e$ and the measured conductance is given by the expression $G_{\rm 4-probe} = (e^2/h)\, T(\varepsilon_{\rm F})/R(\varepsilon_{\rm F})$.

7.2.1 Example: potential step

Perhaps the simplest scattering problem is one-dimensional scattering from a potential step, $V(x) = V_0\, \Theta(x)$. The potential is piecewise constant, hence the wavefunction is piecewise a plane wave:

$$\begin{aligned} x < 0: \quad & \psi(x) = I\, e^{ikx} + O'\, e^{-ikx} \\ x > 0: \quad & \psi(x) = O\, e^{ik'x} + I'\, e^{-ik'x} \quad , \end{aligned} \tag{7.14}$$

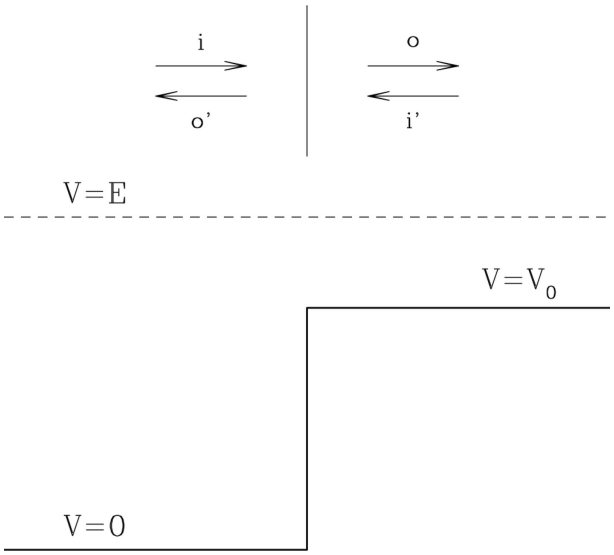


Figure 7.1: Scattering at a potential step.

with

$$E = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2 k'^2}{2m} + V_0 \quad . \tag{7.15}$$

The requirement that $\psi(x)$ and its derivative $\psi'(x)$ be continuous at $x = 0$ gives us two equations which relate the four wavefunction amplitudes:

$$\begin{aligned} I + O' &= O + I' \\ k(I - O') &= k'(O - I') \quad . \end{aligned} \tag{7.16}$$

As emphasized earlier, the $\mathcal{S}$-matrix acts on flux amplitudes. We have

$$\begin{pmatrix} i \\ o' \end{pmatrix} = \sqrt{v} \begin{pmatrix} I \\ O' \end{pmatrix} \quad , \quad \begin{pmatrix} o \\ i' \end{pmatrix} = \sqrt{v'} \begin{pmatrix} O \\ I' \end{pmatrix} \quad , \tag{7.17}$$

with $v = \hbar k/m$ and $v' = \hbar k'/m$. One easily finds the $\mathcal{S}$-matrix, defined in eqn. 7.2, is given by

$$\mathcal{S} = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix} = \begin{pmatrix} \frac{1-\epsilon}{1+\epsilon} & \frac{2\sqrt{\epsilon}}{1+\epsilon} \\ \frac{2\sqrt{\epsilon}}{1+\epsilon} & \frac{\epsilon-1}{1+\epsilon} \end{pmatrix} \quad , \tag{7.18}$$

where $\epsilon \equiv v'/v = k'/k = \sqrt{1 - \frac{V_0}{E}}$, where $E = \varepsilon_{\rm F}$ is the Fermi energy. The two- and four-terminal conductances are then given by

$$\begin{aligned} G_{2-\text{probe}} &= \frac{e^2}{h} |t|^2 = \frac{e^2}{h} \cdot \frac{4\epsilon}{(1+\epsilon)^2} \\ G_{4-\text{probe}} &= \frac{e^2}{h} \frac{|t|^2}{|r|^2} = \frac{e^2}{h} \cdot \frac{4\epsilon}{(1-\epsilon)^2} \quad . \end{aligned} \tag{7.19}$$

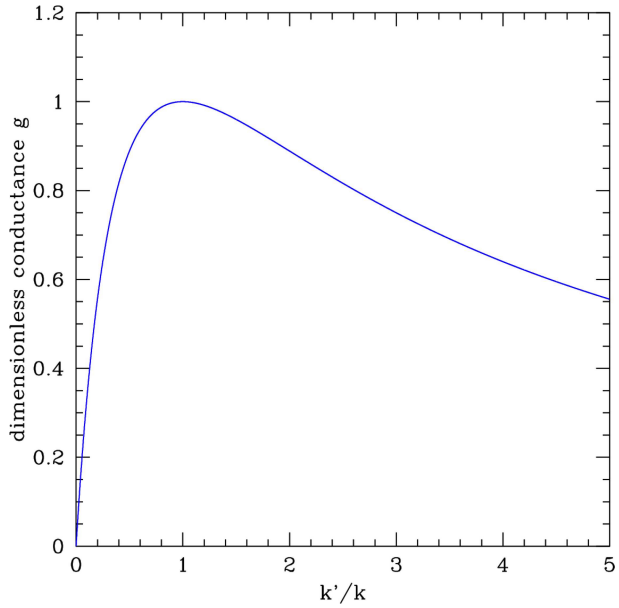


Figure 7.2: Dimensionless two-terminal conductance g *versus* k'/k for the potential step. The conductance is maximized when $k' = k$.

Both are maximized when the transmission probability $T = |t|^2 = 1$ is largest, which occurs for $\epsilon = 1$, *i.e.* $k' = k$.

7.3 Multichannel Systems

The single channel scenario described above is obtained as a limit of a more general multichannel case, in which there are transverse degrees of freedom (due *e.g.* to finite cross-sectional area of the wire) as well. We will identify the transverse states by labels i. Within the perfect leads, the longitudinal and transverse energies are decoupled, and we may write

$$\varepsilon = \varepsilon_{\perp i} + \varepsilon_{\parallel}(k) \quad , \tag{7.20}$$

where $\varepsilon_{\parallel}(k)$ is the one-dimensional dispersion due to motion along the wire (*e.g.* $\varepsilon_{\parallel}(k) = \hbar^2 k^2/2m^*$, $\varepsilon_{\parallel}(k) = -2t\cos ka$, *etc.*). k is the component of the wavevector along the axis of the wire. We assume that the transverse dimensions are finite, so fixing the Fermi energy $\varepsilon_{\rm F}$ in turn fixes the total number of transverse channels, $N_{\rm c}$, which contribute to the transport:

$$\begin{aligned} N_{\rm c}(\varepsilon) &= \sum_i \Theta(\varepsilon - \varepsilon_{\perp i}) \qquad \text{(continuum)} \\ &= \sum_i \Theta\big(2t - |\varepsilon - \varepsilon_{\perp i}|\,\big) \qquad \text{(tight binding)} \quad . \end{aligned} \tag{7.21}$$

Equivalently, an electron with energy ε in transverse state i has wavevector k_i which satisfies

$$\varepsilon_{\parallel}(k_i) = \varepsilon - \varepsilon_{\perp i} \quad . \tag{7.22}$$

$N_{\rm c}$ is the number of real positive roots of (7.22). Typically $N_{\rm c} \approx k_{\rm F}^{d-1} A$, where A is the cross-sectional area and $k_{\rm F}$ is the Fermi wavevector. The velocity v_i is

$$v_i(\varepsilon) = \frac{1}{\hbar}\frac{\partial\varepsilon}{\partial k}\bigg|_{k_i} = \frac{1}{\hbar}\frac{\partial\varepsilon_{\parallel}(k)}{\partial k}\bigg|_{k=k_i} \quad . \tag{7.23}$$

The density of states $\mathcal{N}_i(\varepsilon)$ (per unit spin, per direction) for electrons in the $i^{\rm th}$ transverse channel is

$$\mathcal{N}_i(\varepsilon) = \int\limits_{\hat{\Omega}} \frac{dk}{2\pi}\,\Theta\big(v(k)\big)\,\delta\big(\varepsilon - \varepsilon_{\perp i} - \varepsilon_{\parallel}(k)\big) = \frac{1}{2\pi}\frac{dk}{d\varepsilon_{\parallel}}\bigg|_{k=k_i} \quad , \tag{7.24}$$

so once again we have for the product $h\,v_i(\varepsilon)\,\mathcal{N}_i(\varepsilon) = 1$.

Consider now a section of disordered material connected to perfect leads on the left and right. The solution to the Schrödinger equation on either side of the disordered region is

$$\begin{aligned}
\psi_{\rm left}(\boldsymbol{x}_\perp, z) &= \sum_{j=1}^{N_{\rm c}^{\rm L}} \Big\{ I_j\, e^{+ik_j z} + O'_i\, e^{-ik_j z} \Big\}\, \varphi_j^{\rm L}(\boldsymbol{x}_\perp) \\
\psi_{\rm right}(\boldsymbol{x}_\perp, z) &= \sum_{a=1}^{N_{\rm c}^{\rm R}} \Big\{ O_a\, e^{+ik_a z} + I'_a\, e^{-ik_a z} \Big\}\, \varphi_a^{\rm R}(\boldsymbol{x}_\perp) \quad .
\end{aligned} \tag{7.25}$$

Here, we have assumed a general situation in which the number of transverse channels $N_{\rm c}^{\rm L,R}$ may differ between the left and right lead. The quantities $\{I_j, O'_j, O_a, I'_a\}$ are *wave function amplitudes*. The $\mathcal{S}$-matrix, on the other hand, acts on *flux amplitudes* $\{i_j, o'_j, o_a, i'_a\}$, which are related to the wavefunction amplitudes as follows:

$$\begin{aligned}
i_i &= v_i^{1/2}\, I_i \qquad o_a = v_a^{1/2}\, O_a \\
o'_i &= v_i^{1/2}\, O'_i \qquad i'_a = v_a^{1/2}\, I'_a \quad .
\end{aligned} \tag{7.26}$$

The $\mathcal{S}$-matrix is a $(N_{\rm c}^{\rm R} + N_{\rm c}^{\rm L}) \times (N_{\rm c}^{\rm R} + N_{\rm c}^{\rm L})$ matrix,

$$\mathcal{S} = \begin{pmatrix} r_{N_{\rm c}^{\rm L}\times N_{\rm c}^{\rm L}} & t'_{N_{\rm c}^{\rm L}\times N_{\rm c}^{\rm R}} \\ t_{N_{\rm c}^{\rm R}\times N_{\rm c}^{\rm L}} & r'_{N_{\rm c}^{\rm R}\times N_{\rm c}^{\rm R}} \end{pmatrix} \tag{7.27}$$

which relates outgoing and incoming flux amplitudes:

$$\begin{pmatrix} o' \\ o \end{pmatrix} = \overbrace{\begin{pmatrix} r & t' \\ t & r' \end{pmatrix}}^{\mathcal{S}} \begin{pmatrix} i \\ i' \end{pmatrix} \quad . \tag{7.28}$$

Unitarity of $\mathcal{S}$ means that $\mathcal{S}^\dagger\mathcal{S} = \mathcal{S}\mathcal{S}^\dagger = \mathbb{I}$, where

$$\mathcal{S} = \begin{pmatrix} r & t' \\ t & r' \end{pmatrix} \qquad \Longrightarrow \qquad \mathcal{S}^\dagger = \begin{pmatrix} r^\dagger & t^\dagger \\ t'^\dagger & r'^\dagger \end{pmatrix} \quad , \tag{7.29}$$

and hence unitarity says

$$\begin{aligned} r_{ik}\, r^*_{jk} + t'_{ic}\, t'^*_{jc} &= \delta_{ij} & \qquad r^*_{ki}\, r_{kj} + t^*_{ci}\, t_{cj} &= \delta_{ij} \\ t_{ak}\, t^*_{bk} + r'_{ac}\, r'^*_{bc} &= \delta_{ab} & \qquad t'^*_{ka}\, t'_{kb} + r'^*_{ca}\, r'_{cb} &= \delta_{ab} \\ r_{ik}\, t^*_{ak} + t'_{ic}\, r'^*_{ac} &= 0 & \qquad r^*_{ki}\, t'_{ka} + t^*_{ci}\, r'_{ca} &= 0 \quad , \end{aligned} \tag{7.30}$$

or, in matrix notation,

$$\begin{aligned} r\, r^\dagger + t'\, t'^\dagger &= \quad r^\dagger\, r + t^\dagger t \quad &&= \mathbb{I}_{N_{\rm c}^{\rm L}\times N_{\rm c}^{\rm L}} \\ t\, t^\dagger + r'\, r'^\dagger &= t'^\dagger\, t' + r'^\dagger\, r' &&= \mathbb{I}_{N_{\rm c}^{\rm R}\times N_{\rm c}^{\rm R}} \\ r\, t^\dagger + t'\, r'^\dagger &= r^\dagger\, t' + t^\dagger\, r' &&= \mathbb{O}_{N_{\rm c}^{\rm L}\times N_{\rm c}^{\rm R}} \\ t\, r^\dagger + r'\, t'^\dagger &= t'^\dagger\, r + r'^\dagger\, t &&= \mathbb{O}_{N_{\rm c}^{\rm R}\times N_{\rm c}^{\rm L}} \quad . \end{aligned} \tag{7.31}$$

We define the probabilities

$$R_i = \sum_{k=1}^{N_{\rm c}^{\rm L}} r_{ik}\, r^*_{ik} \quad , \quad T_a = \sum_{k=1}^{N_{\rm c}^{\rm L}} t_{ak}\, t^*_{ak} \quad , \quad T'_i = \sum_{c=1}^{N_{\rm c}^{\rm R}} t'_{ic}\, t'^*_{ic} \quad , \quad R'_a = \sum_{c=1}^{N_{\rm c}^{\rm R}} r'^*_{ac}\, r'_{ac} \quad , \tag{7.32}$$

for which it follows that

$$R_i + T'_i = 1 \qquad , \qquad R'_a + T_a = 1 \tag{7.33}$$

for all $i \in \{1, \ldots, N_{\rm c}^{\rm L}\}$ and $a \in \{1, \ldots, N_{\rm c}^{\rm R}\}$. Unitarity of the $\mathcal{S}$-matrix preserves particle flux:

$$|i|^2 - |i'|^2 = |o|^2 - |o'|^2 \, , \tag{7.34}$$

which is shorthand for

$$\sum_{j=1}^{N_{\rm c}^{\rm L}} |i_j|^2 + \sum_{a=1}^{N_{\rm c}^{\rm R}} |i'_a|^2 = \sum_{a=1}^{N_{\rm c}^{\rm R}} |o_a|^2 + \sum_{j=1}^{N_{\rm c}^{\rm L}} |o'_j|^2 \quad . \tag{7.35}$$

Onsager reciprocity demands that $\mathcal{S}(-\boldsymbol{H}) = \mathcal{S}^{\rm t}(\boldsymbol{H})$.

Let us now compute the current in the right lead flowing past the imaginary surface Σ

$$\begin{aligned} I_\Sigma &= -e \sum_{a=1}^{N_{\rm c}^{\rm R}} \int d\varepsilon\, \mathcal{N}_a(\varepsilon)\, v_a(\varepsilon) \left\{ \overbrace{\sum_{i=1}^{N_{\rm c}^{\rm L}} \big|t_{ai}(\varepsilon)\big|^2}^{T_a(\varepsilon)} \; f(\varepsilon - \mu_{\rm L}) + \left[\overbrace{\sum_{b=1}^{N_{\rm c}^{\rm R}} \big|r'_{ab}(\varepsilon)\big|^2}^{R'_a(\varepsilon)} - 1 \right] f(\varepsilon - \mu_{\rm R}) \right\} \\ &= \frac{e}{h}\,(\mu_{\rm R} - \mu_{\rm L}) \int d\varepsilon \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \sum_{a=1}^{N_{\rm c}^{\rm R}} T_a(\varepsilon) \quad . \end{aligned} \tag{7.36}$$

Thus, the result of a two-probe measurement would be

$$G_{2-\rm probe} = \frac{eI}{\mu_{\rm R} - \mu_{\rm L}} = \frac{e^2}{h} \int d\varepsilon \left(-\frac{\partial f^0}{\partial \varepsilon} \right) \sum_{a=1}^{N_{\rm c}^{\rm R}} T_a(\varepsilon) \quad . \tag{7.37}$$

At zero temperature, then,

$$G_{2-\rm probe} = \frac{e^2}{h} \ {\rm Tr}\, tt^\dagger \tag{7.38}$$

where

$${\rm Tr}\, tt^\dagger = {\rm Tr}\, t^\dagger t = \sum_{i=1}^{N_{\rm c}^{\rm L}} \sum_{a=1}^{N_{\rm c}^{\rm R}} \left| t_{ai} \right|^2 \quad . \tag{7.39}$$

To determine $G_{4-\rm probe}$, we must compute the effective chemical potentials $\tilde{\mu}_{\rm L}$ and $\tilde{\mu}_{\rm R}$ in the leads. We again do this by equating expressions for the electron number density. In the left lead,

$$\begin{aligned} n_{\rm L} &= \sum_{i=1}^{N_{\rm c}^{\rm L}} \int d\varepsilon \, \mathcal{N}_i(\varepsilon) \Big\{ \big[1 + R_i(\varepsilon) \big] f(\varepsilon - \mu_{\rm L}) + T_i'(\varepsilon) \, f(\varepsilon - \mu_{\rm R}) \Big\} \\ &= 2 \sum_i \int d\varepsilon \, \mathcal{N}_i(\varepsilon) \, f(\varepsilon - \tilde{\mu}_{\rm L}) \\ \Longrightarrow \quad \tilde{\mu}_{\rm L} &= \mu_{\rm L} - \tfrac{1}{2} \overline{T'} \, (\mu_{\rm L} - \mu_{\rm R}) \end{aligned} \tag{7.40}$$

where $\overline{T'}$ is a weighted average,

$$\overline{T}' \equiv \frac{\sum_i v_i^{-1} \, T_i'}{\sum_i v_i^{-1}} \quad . \tag{7.41}$$

Similarly, one obtains for the right lead,

$$\begin{aligned} n_{\rm R} &= \sum_{a=1}^{N_{\rm c}^{\rm R}} \int d\varepsilon \, \mathcal{N}_a(\varepsilon) \Big\{ T_a(\varepsilon) \, f(\varepsilon - \mu_{\rm L}) + \big[1 + R_a'(\varepsilon) \big] \, f(\varepsilon - \mu_{\rm R}) \Big\} \\ &= 2 \sum_a \int d\varepsilon \, \mathcal{N}_a(\varepsilon) \, f(\varepsilon - \tilde{\mu}_{\rm R}) \\ \Longrightarrow \quad \tilde{\mu}_{\rm R} &= \mu_{\rm R} + \tfrac{1}{2} \overline{T} \, (\mu_{\rm L} - \mu_{\rm R}) \end{aligned} \tag{7.42}$$

where

$$\overline{T} \equiv \frac{\sum_a v_a^{-1} \, T_a}{\sum_a v_a^{-1}} \quad . \tag{7.43}$$

(We have assumed zero temperature throughout.) The difference in lead chemical potentials is thus

$$(\tilde{\mu}_{\rm L} - \tilde{\mu}_{\rm R}) = \Big(1 - \tfrac{1}{2} \overline{T} - \tfrac{1}{2} \overline{T'} \Big) \cdot (\mu_{\rm L} - \mu_{\rm R}) \quad . \tag{7.44}$$

Hence, we obtain the 4-probe conductance,

$$G_{4-\rm probe} = \frac{e^2}{h}\,\frac{\sum_a T_a}{1 - \frac{1}{2}\Big(\sum_i T_i'\,v_i^{-1}\Big/\sum_i v_i^{-1}\Big) - \frac{1}{2}\Big(\sum_a T_a\,v_a^{-1}\Big/\sum_a v_a^{-1}\Big)} \tag{7.45}$$

7.3.1 Transfer matrices: the Pichard formula

The transfer matrix $\mathcal{S}$ acts on incoming flux amplitudes to give outgoing flux amplitudes. This linear relation may be recast as one which instead relates flux amplitudes in the right lead to those in the left lead, *i.e.*

$$\begin{pmatrix} o' \\ o \end{pmatrix} = \overbrace{\begin{pmatrix} r & t' \\ t & r' \end{pmatrix}}^{\mathcal{S}} \begin{pmatrix} i \\ i' \end{pmatrix} \qquad \Longrightarrow \qquad \begin{pmatrix} o \\ i' \end{pmatrix} = \overbrace{\begin{pmatrix} \mathcal{M}_{11} & \mathcal{M}_{12} \\ \mathcal{M}_{21} & \mathcal{M}_{22} \end{pmatrix}}^{\mathcal{M}} \begin{pmatrix} i \\ o' \end{pmatrix} \quad . \tag{7.46}$$

$\mathcal{M}$ is known as the *transfer matrix*. Note that each of the blocks of $\mathcal{M}$ is of dimension $N_{\rm c}^{\rm R} \times N_{\rm c}^{\rm L}$, and $\mathcal{M}$ itself is a rectangular $2N_{\rm c}^{\rm R} \times 2N_{\rm c}^{\rm L}$ matrix. The individual blocks of $\mathcal{M}$ are readily determined:

$$\begin{aligned} o' = r\,i + t'\,i' \qquad &\Longrightarrow \qquad i' = -t'^{-1}\,r\,i + t'^{-1}\,o' \\ o = t\,i + r'\,i' \qquad &\Longrightarrow \qquad o = (t - r'\,t'^{-1}\,r)\,i + r'\,t'^{-1}\,o' \quad , \end{aligned} \tag{7.47}$$

so we conclude

$$\mathcal{M}_{11} = t^{\dagger -1} \quad , \quad \mathcal{M}_{12} = r'\,t'^{-1} \quad , \quad \mathcal{M}_{21} = -t'^{-1}\,r \quad , \quad \mathcal{M}_{22} = t'^{-1} \quad . \tag{7.48}$$

WARNING: *None of this makes any sense if* $N_{\rm c}^{\rm L} \neq N_{\rm c}^{\rm R}$ *!* The reason is that it is problematic to take the inverse of a rectangular matrix such as t or t', as was blithely done above in Eqns. 7.47. We therefore must assume $N_{\rm c}^{\rm R} = N_{\rm c}^{\rm L} = N_{\rm c}$, and that the scatterers are separated by identical perfect regions. Practically, this imposes no limitations at all, since the width of the perfect regions can be taken to be arbitrarily small.

EXERCISE: Show that $\mathcal{M}_{11} = t - r'\,t'^{-1}\,r = t^{\dagger -1}$.

The virtue of transfer matrices is that they are *multiplicative*. Consider, for example, two disordered regions connected by a region of perfect conductor. The outgoing flux o from the first region becomes the incoming flux i for the second, as depicted in fig. 7.3. Thus, if $\mathcal{M}_1$ is the transfer matrix for scatterer #1, and $\mathcal{M}_2$ is the transfer matrix for scatterer #2, the transfer matrix for the two scatterers in succession is $\mathcal{M} = \mathcal{M}_2\mathcal{M}_1$:

$$\begin{pmatrix} o_2 \\ i_2' \end{pmatrix} = \begin{pmatrix} \mathcal{M}_2^{11} & \mathcal{M}_2^{12} \\ \mathcal{M}_2^{21} & \mathcal{M}_2^{22} \end{pmatrix} \begin{pmatrix} i_2 \\ o_2' \end{pmatrix} = \overbrace{\begin{pmatrix} \mathcal{M}_2^{11} & \mathcal{M}_2^{12} \\ \mathcal{M}_2^{21} & \mathcal{M}_2^{22} \end{pmatrix} \begin{pmatrix} \mathcal{M}_1^{11} & \mathcal{M}_1^{12} \\ \mathcal{M}_1^{21} & \mathcal{M}_1^{22} \end{pmatrix}}^{\mathcal{M} = \mathcal{M}_2\,\mathcal{M}_1} \begin{pmatrix} i_1 \\ o_1' \end{pmatrix} \quad . \tag{7.49}$$

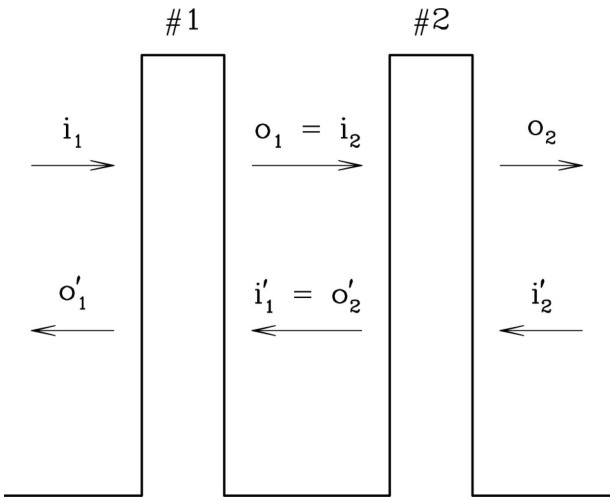


Figure 7.3: Two quantum scatterers in series. The right side data for scatterer #1 become the left side data for scatterer #2.

Clearly, then, if we have many scatterers in succession, this result generalizes to

$$\mathcal{M} = \mathcal{M}_N \, \mathcal{M}_{N-1} \cdots \mathcal{M}_1 \quad . \tag{7.50}$$

Unitarity of the $\mathcal{S}$-matrix means that the transfer matrix is *pseudo-unitary* in that it satisfies

$$\mathcal{M}^\dagger \, \Sigma \, \mathcal{M} = \Sigma \qquad \text{where} \quad \Sigma = \begin{pmatrix} \mathbb{I}_{N_c \times N_c} & \mathbb{O}_{N_c \times N_c} \\ \mathbb{O}_{N_c \times N_c} & -\mathbb{I}_{N_c \times N_c} \end{pmatrix} \quad . \tag{7.51}$$

This, in turn, implies conservation of the pseudo-norm,

$$|o|^2 - |i'|^2 = |i|^2 - |o'|^2 \, , \tag{7.52}$$

which is simply a restatement of (7.34).

We now assert that

$$\Big[\mathcal{M}^\dagger \mathcal{M} + (\mathcal{M}^\dagger \mathcal{M})^{-1} + 2 \cdot \mathbb{I}\Big]^{-1} = \tfrac{1}{4} \begin{pmatrix} t^\dagger t & 0 \\ 0 & t' t'^\dagger \end{pmatrix} \quad . \tag{7.53}$$

This result is in fact easily derived once one notes that

$$\mathcal{M}^{-1} = \Sigma \, \mathcal{M}^\dagger \, \Sigma = \begin{pmatrix} \mathcal{M}_{11}^\dagger & -\mathcal{M}_{21}^\dagger \\ -\mathcal{M}_{12}^\dagger & \mathcal{M}_{22}^\dagger \end{pmatrix} \quad . \tag{7.54}$$

EXERCISE: Verify eqn. (7.53).

The 2-probe conductance (per spin channel) may now be written in terms of the transfer matrix as

$$G_{2-\text{probe}} = \frac{2e^2}{h} \, \mathsf{Tr} \Big[\mathcal{M}^\dagger \mathcal{M} + (\mathcal{M}^\dagger \mathcal{M})^{-1} + 2 \cdot \mathbb{I}\Big]^{-1} \tag{7.55}$$

This is known as the *Pichard Formula*.

7.3.2 Discussion of the Pichard formula

It is convenient to work in an eigenbasis of the Hermitian matrix $\mathcal{M}^\dagger\mathcal{M}$. The eigenvalues of $\mathcal{M}^\dagger\mathcal{M}$ are roots of the characteristic polynomial

$$p(\lambda) = \det\left(\lambda - \mathcal{M}^\dagger\mathcal{M}\right) \quad . \tag{7.56}$$

Owing to the pseudo-unitarity of $\mathcal{M}$, we have

$$\begin{aligned} p(\lambda) &= \det\bigl(\lambda - \mathcal{M}^\dagger\mathcal{M}\bigr) \\ &= \det\bigl(\lambda - \Sigma\,\mathcal{M}^{-1}\,\Sigma\cdot\Sigma\,\mathcal{M}^{\dagger-1}\,\Sigma\bigr) \\ &= \det\bigl(\lambda - \Sigma\,\mathcal{M}^{-1}\,\mathcal{M}^{\dagger-1}\,\Sigma\bigr) \\ &= \lambda^{2N_{\rm c}}\det\bigl(\lambda^{-1} - \mathcal{M}^\dagger\mathcal{M}\bigr)/\det\bigl(\mathcal{M}^\dagger\mathcal{M}\bigr) \quad , \end{aligned} \tag{7.57}$$

from which we conclude that $p(\lambda) = 0$ implies $p(\lambda^{-1}) = 0$, and the eigenvalues of $\mathcal{M}^\dagger\mathcal{M}$ come in (λ, λ^{-1}) pairs. We can therefore write Pichard's formula as

$$G_{2-\rm probe} = \frac{e^2}{h}\sum_{i=1}^{N_{\rm c}} \frac{4}{\lambda_i + \lambda_i^{-1} + 2} \quad , \tag{7.58}$$

where without loss of generality we assume $\lambda_i \geqslant 1$ for each $i \in \{1, \ldots, N_{\rm c}\}$. We define the $i^{\rm th}$ *localization length* ξ_i through

$$\lambda_i \equiv \exp\left(\frac{2L}{\xi_i}\right) \qquad \Longrightarrow \qquad \xi_i = \frac{2L}{\ln\lambda_i} \quad , \tag{7.59}$$

where L is the length of the disordered region. We now have

$$G_{2-\rm probe} = \frac{e^2}{h}\sum_{i=1}^{N_{\rm c}} \frac{2}{1 + \cosh(2L/\xi_i)} \tag{7.60}$$

If $N_{\rm c}$ is finite, then as $L \to \infty$ the $\{\xi_i\}$ converge to definite values, for a wide range of distributions $P(\mathcal{M}_n)$ for the individual scatterer transfer matrices. This follows from a version of the central limit theorem as applied to nonabelian multiplicative noise (*i.e.* products of random matrices), known as *Oseledec's theorem*. We may choose to order the eigenvalues such that $\lambda_1 < \lambda_2 < \cdots < \lambda_{N_{\rm c}}$, and hence $\xi_1 > \xi_2 > \cdots > \xi_{N_{\rm c}}$. In the $L \to \infty$ limit, then, the conductance is dominated by the largest localization length, and

$$G(L) \simeq \frac{4e^2}{h}\,e^{-2L/\xi_1} \qquad (N_{\rm c}\ \text{finite},\ L \to \infty) \quad . \tag{7.61}$$

(We have dropped the label '2-probe' on G.) The quantity $\xi \equiv \xi_1$ is called the *localization length*, and it is dependent on the (Fermi) energy: $\xi = \xi(\varepsilon)$.

Suppose now that L is finite, and furthermore that $\xi_1 > 2L > \xi_{N_c}$. Channels for which $2L \ll \xi_j$ give $\cosh(2L/\xi_j) \approx 1$, and therefore contribute a quantum of conductance e^2/h to G. These channels are called *open*. Conversely, when $2L \gg \xi_j$, we have $\cosh(2L/\xi_j) \sim \frac{1}{2}\exp(2L/\xi_j) \gg 1$, and these *closed* channels each contribute $\Delta G_j = (2e^2/h)e^{-2L/\xi_j}$ to the conductance, a negligible amount. Thus,

$$G(L) \simeq \frac{e^2}{h}\, N_{\rm c}^{\rm open} \qquad , \qquad N_{\rm c}^{\rm open} \equiv \sum_{j=1}^{N_{\rm c}} \Theta(\xi_j - 2L) \quad . \tag{7.62}$$

Of course, $N_{\rm c}^{\rm closed} = N_{\rm c} - N_{\rm c}^{\rm open}$, although there is no precise definition for open *vs.* closed for channels with $\xi_j \sim 2L$. This discussion naturally leads us to the following classification scheme:

- When $L > \xi_1$, the system is in the *localized regime*. The conductance vanishes exponentially with L according to $G(L) \approx (4e^2/h)\exp(-2L/\xi)$, where $\xi(\varepsilon) = \xi_1(\varepsilon)$ is the *localization length*. In the localized regime, there are no open channels: $N_{\rm c}^{\rm open} = 0$.
- When $N_{\rm c}^{\rm open} = \ell\, N_{\rm c}/L$, where ℓ is the *elastic scattering length*, one is in the *Ohmic regime*. In the Ohmic regime, for a d-dimensional system of length L and ($(d-1)$-dimensional) cross-sectional area A,

  $$G_{\rm Ohmic} \approx \frac{e^2}{h}\frac{\ell}{L}\, k_{\rm F}^{d-1} A = \frac{e^2}{h}\, k_{\rm F}^{d-1}\, \ell \cdot \frac{A}{L} \quad . \tag{7.63}$$

  Note that G is proportional to the cross sectional area A and inversely proportional to the length L, which is the proper Ohmic behavior: $G = \sigma\, A/L$, where

  $$\sigma \approx \frac{e^2}{h}\, k_{\rm F}^{d-1}\, \ell \tag{7.64}$$

  is the conductivity.
- When $L < \xi_{N_c}$, all the channels are open: $N_{\rm c}^{\rm open} = N_{\rm c}$. The conductance is

  $$G(L) = \frac{e^2}{h}\, N_{\rm c} \approx \frac{e^2}{h}\, k_{\rm F}^{d-1}\, A \quad . \tag{7.65}$$

  This is the *ballistic regime*.

If we keep $N_{\rm c} \propto (k_{\rm F} L)^{d-1}$, then for $L \to \infty$ Oseledec's theorem does not apply, because the transfer matrix is ∞-dimensional. If $\xi_1(\varepsilon)$ nonetheless remains finite, then $G(L) \approx (4e^2/h)\exp(-L/\xi) \to 0$ and the system is in the localized regime. If, on the other hand, $\xi_1(\varepsilon)$ diverges as $L \to \infty$ such that $\exp(L/\xi_1)$ is finite, then $G > 0$ and the system is a conductor.

If we define $\nu_i \equiv \ln \lambda_i$, the dimensionless conductance $g = (h/e^2)\, G$ is given by

$$g = 2\int_0^\infty d\nu\, \frac{\sigma(\nu)}{1+\cosh\nu} \quad , \tag{7.66}$$

where

$$\sigma(\nu) = \sum_{i=1}^{N_{\rm c}} \delta(\nu - \nu_i) \tag{7.67}$$

is the density of ν values. This distribution is normalized so that $\int_0^\infty d\nu\, \sigma(\nu) = N_{\rm c}$. Spectral properties of the $\{\nu_i\}$ thus determine the statistics of the conductance. For example, averaging over disorder realizations gives

$$\langle g \rangle = 2 \int\limits_0^\infty d\nu\, \frac{\big\langle \sigma(\nu) \big\rangle}{1 + \cosh\nu} \quad . \tag{7.68}$$

The average of g^2, though, depends on the two-point correlation function, *viz.*

$$\langle g^2 \rangle = 4 \int\limits_0^\infty d\nu \int\limits_0^\infty d\nu'\, \frac{\big\langle \sigma(\nu)\, \sigma(\nu') \big\rangle}{(1 + \cosh\nu)(1 + \cosh\nu')} \quad . \tag{7.69}$$

7.3.3 Two quantum resistors in series

Let us consider the case of two scatterers in series. For simplicity, we will assume that $N_{\rm c} = 1$, in which case the transfer matrix for a single scatterer may be written as

$$\mathcal{M} = \begin{pmatrix} 1/t^* & -r^*/t^* \\ -r/t' & 1/t' \end{pmatrix} \quad . \tag{7.70}$$

A pristine segment of wire of length L has a diagonal transfer matrix

$$\mathcal{N} = \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \quad , \tag{7.71}$$

where $\beta = kL$. Thus, the composite transfer matrix for two scatterers joined by a length L of pristine wire is $\mathcal{M} = \mathcal{M}_2 \mathcal{N} \mathcal{M}_1$, *i.e.*

$$\mathcal{M} = \begin{pmatrix} 1/t_2^* & -r_2^*/t_2^* \\ -r_2/t_2' & 1/t_2' \end{pmatrix} \begin{pmatrix} e^{i\beta} & 0 \\ 0 & e^{-i\beta} \end{pmatrix} \begin{pmatrix} 1/t_1^* & -r_1^*/t_1^* \\ -r_1/t_1' & 1/t_1' \end{pmatrix} \quad . \tag{7.72}$$

In fact, the inclusion of the transfer matrix $\mathcal{N}$ is redundant; the phases $e^{\pm i\beta}$ can be completely absorbed via a redefinition of $\{t_1, t_1', r_1, r_1'\}$.

Extracting the upper left element of $\mathcal{M}$ gives

$$\frac{1}{t^*} = \frac{e^{i\beta} - e^{-i\beta}\, r_1'^* r_2^*}{t_1^* t_2^*} \quad , \tag{7.73}$$

hence the transmission coefficient T for the composite system is

$$T = \frac{T_1 T_2}{1 + R_1 R_2 - 2\sqrt{R_1 R_2}\,\cos\delta} \tag{7.74}$$

where $\delta = 2\beta + \arg(r_1' r_2)$. The dimensionless Landauer resistance is then

$$\begin{aligned}\mathcal{R} = \frac{R}{T} &= \frac{R_1 + R_2 - 2\sqrt{R_1 R_2}\,\cos\delta}{T_1 T_2} \\ &= \mathcal{R}_1 + \mathcal{R}_2 + 2\,\mathcal{R}_1\mathcal{R}_2 - 2\sqrt{\mathcal{R}_1\mathcal{R}_2(1+\mathcal{R}_1)(1+\mathcal{R}_2)}\,\cos\delta \quad .\end{aligned} \tag{7.75}$$

If we average over the random phase δ, we obtain

$$\langle \mathcal{R} \rangle = \mathcal{R}_1 + \mathcal{R}_2 + 2\,\mathcal{R}_1\mathcal{R}_2 \quad . \tag{7.76}$$

The first two terms correspond to Ohm's law. The final term is unfamiliar and leads to a divergence of resistivity as a function of length. To see this, imagine that that $\mathcal{R}_2 = \varrho\, dL$ is small, and solve (7.76) iteratively. We then obtain a differential equation for the dimensionless resistance $\mathcal{R}(L)$:

$$d\mathcal{R} = (1 + 2\mathcal{R})\,\varrho\, dL \qquad \Longrightarrow \qquad \mathcal{R}(L) = \tfrac{1}{2}\big(e^{2\varrho L} - 1\big) \quad . \tag{7.77}$$

In fact, the distribution $P_L(\mathcal{R})$ is extremely broad, and it is more appropriate to average the quantity $\ln(1+\mathcal{R})$. Using

$$\int_0^{2\pi} \frac{d\delta}{2\pi}\, \ln(a - b\cos\delta) = \ln\left(\tfrac{1}{2}a + \tfrac{1}{2}\sqrt{a^2 - b^2}\right) \tag{7.78}$$

with

$$\begin{aligned} a &= 1 + \mathcal{R}_1 + \mathcal{R}_2 + 2\,\mathcal{R}_1\,\mathcal{R}_2 \\ b &= 2\sqrt{\mathcal{R}_1\mathcal{R}_2(1+\mathcal{R}_1)(1+\mathcal{R}_2)} \quad , \end{aligned} \tag{7.79}$$

we obtain the result

$$\langle \ln(1+\mathcal{R}) \rangle = \ln(1+\mathcal{R}_1) + \ln(1+\mathcal{R}_2) \quad . \tag{7.80}$$

We define the quantity

$$x(L) \equiv \ln\big\{1 + \mathcal{R}(L)\big\} \quad , \tag{7.81}$$

and we observe

$$\begin{aligned} \big\langle x(L) \big\rangle &= \varrho\, L \\ \big\langle e^{x(L)} \big\rangle &= \tfrac{1}{2}\big(e^{2\varrho L} + 1\big) \quad . \end{aligned} \tag{7.82}$$

Note that $\langle e^x \rangle \neq e^{\langle x \rangle}$. The quantity $x(L)$ is an appropriately *self-averaging* quantity in that its root mean square fluctuations are small compared to its average, *i.e.* it obeys the central limit theorem. On the other hand, $\mathcal{R}(L)$ is *not* self-averaging, *i.e.* it is not normally distributed.

Abelian multiplicative random processes

Let $p(x)$ be a distribution on the nonnegative real numbers, normalized according to

$$\int_0^\infty dx\, p(x) = 1 \quad , \tag{7.83}$$

and define

$$X \equiv \prod_{i=1}^N x_i \qquad , \qquad Y \equiv \ln X = \sum_{i=1}^N \ln x_i \quad . \tag{7.84}$$

The distribution for Y is

$$\begin{aligned} P_N(Y) &= \int_0^\infty dx_1 \int_0^\infty dx_2 \cdots \int_0^\infty dx_N\, p(x_1)\, p(x_2) \cdots p(x_N)\, \delta\left(Y - \sum_{i=1}^N \ln x_i\right) \\ &= \int_{-\infty}^\infty \frac{d\omega}{2\pi}\, e^{i\omega Y} \left\{ \int_0^\infty dx\, p(x)\, e^{-i\omega \ln x} \right\}^N \\ &= \int_{-\infty}^\infty \frac{d\omega}{2\pi}\, e^{i\omega Y} \left[1 - i\omega \langle \ln x \rangle - \tfrac{1}{2}\omega^2 \langle \ln^2 x \rangle + \mathcal{O}(\omega^3) \right]^N \\ &= \int_{-\infty}^\infty \frac{d\omega}{2\pi}\, e^{i\omega (Y - N\langle \ln x \rangle)}\, e^{-\frac{1}{2} N\omega^2 \left(\langle \ln^2 x \rangle - \langle \ln x \rangle^2 \right) + \mathcal{O}(\omega^3)} \\ &= \frac{1}{\sqrt{2\pi N \sigma^2}}\, e^{-(Y - N\mu)^2 / 2N\sigma^2} \cdot \Big\{ 1 + \mathcal{O}(N^{-1}) \Big\} \quad , \end{aligned} \tag{7.85}$$

with

$$\mu = \langle \ln x \rangle \qquad , \qquad \sigma^2 = \langle \ln^2 x \rangle - \langle \ln x \rangle^2 \tag{7.86}$$

and

$$\big\langle f(x) \big\rangle \equiv \int_0^\infty dx\, p(x)\, f(x) \quad . \tag{7.87}$$

Thus, Y is normally distributed with mean $\langle Y \rangle = N\mu$ and standard deviation $\langle (Y - N\mu)^2 \rangle = N\sigma^2$. This is typical for extensive self-averaging quantities: the average is proportional to the size N of the system, and the root mean square fluctuations are proportional to $\sqrt{N}$. Since $\lim_{N\to\infty} Y_{\rm rms}/\langle Y \rangle \sim \sigma/\sqrt{N}\mu \to 0$, we have that

$$P_{N\to\infty}(Y) \simeq \delta\big(Y - N\mu\big) \quad . \tag{7.88}$$

This is the central limit theorem (CLT) at work. The quantity Y is a sum of independent random variables: $Y = \sum_i \ln x_i$, and is therefore normally distributed with a mean $\bar{Y} = N\mu$ and standard deviation $\sqrt{N}\sigma$, as guaranteed by the CLT. On the other hand, $X = \exp(Y)$ is *not* normally distributed. Indeed, one readily computes the moments of X to be

$$\langle X^k \rangle = e^{kN\mu}\, e^{Nk^2/2\sigma^2} \quad , \tag{7.89}$$

hence

$$\frac{\langle X^k \rangle}{\langle X \rangle^k} = e^{Nk(k-1)/2\sigma^2} \quad , \tag{7.90}$$

which increases exponentially with N. In particular, one finds

$$\frac{\sqrt{\langle X^2 \rangle - \langle X \rangle^2}}{\langle X \rangle} = \left(e^{N/\sigma^2} - 1 \right)^{1/2} \quad . \tag{7.91}$$

The multiplication of random transfer matrices is a more difficult problem to analyze, owing to its essential nonabelian nature. However, as we have seen in our analysis of series quantum resistors, a similar situation pertains: it is the logarithm $\ln(1 + \mathcal{R})$, and not the dimensionless resistance $\mathcal{R}$ itself, which is an appropriate self-averaging quantity.

7.3.4 Two quantum resistors in parallel

The case of parallel quantum resistors is more difficult than that of series resistors. The reason for this is that the conduction path for parallel resistances is multiply connected, *i.e.* electrons can get from start to finish by traveling through either resistor #1 or resistor #2.

Consider electrons with wavevector $k > 0$ moving along a line. The wavefunction is

$$\psi(x) = I\, e^{ikx} + O'\, e^{-ikx} \quad , \tag{7.92}$$

hence the transfer matrix $\mathcal{M}$ for a length L of pristine wire is

$$\mathcal{M}(L) = \begin{pmatrix} e^{ikL} & 0 \\ 0 & e^{-ikL} \end{pmatrix} \quad . \tag{7.93}$$

Now let's bend our wire of length L into a ring. We therefore identify the points $x = 0$ and $x = L = 2\pi R$, where R is the radius. In order for the wavefunction to be single-valued we must have

$$\Big[\mathcal{M}(L) - \mathbb{I} \Big] \begin{pmatrix} I \\ O' \end{pmatrix} = 0 \quad , \tag{7.94}$$

and in order to have a nontrivial solution (*i.e.* I and O' not both zero), we must demand $\det(\mathcal{M} - \mathbb{I}) = 0$, which says $\cos kL = 1$, *i.e.* $k = 2\pi n/L$ with integer n. The energy is then quantized: $\varepsilon_n = \varepsilon_{\scriptscriptstyle \parallel}(k = 2\pi n/L)$.

Next, consider the influence of a vector potential on the transfer matrix. Let us assume the vector potential A along the direction of motion is nonzero over an interval from $x = 0$ to $x = d$. The Hamiltonian is given by the *Peierls substitution*,

$$\mathcal{H} = \varepsilon_{\parallel}\Big(-i\partial_x + \tfrac{e}{\hbar c}A(x)\Big) \quad . \tag{7.95}$$

Note that we can write

$$\begin{aligned} \mathcal{H} &= \Lambda^\dagger(x)\,\varepsilon_{\parallel}(-i\partial_x)\,\Lambda(x) \\ \Lambda(x) &= \exp\left\{\frac{ie}{\hbar c}\int\limits_0^x dx'\,A(x')\right\} \quad . \end{aligned} \tag{7.96}$$

Hence the solutions $\psi(x)$ to $\mathcal{H}\psi = \varepsilon\psi$ are given by

$$\psi(x) = I\,\Lambda^\dagger(x)\,e^{ikx} + O'\,\Lambda^\dagger(x)\,e^{-ikx} \quad . \tag{7.97}$$

The transfer matrix for a segment of length d is then

$$\mathcal{M}(d,A) = \begin{pmatrix} e^{ikd}\,e^{-i\gamma} & 0 \\ 0 & e^{-ikd}\,e^{-i\gamma} \end{pmatrix} \tag{7.98}$$

with

$$\gamma = \frac{e}{\hbar c}\int\limits_0^d dx\,A(x) \quad . \tag{7.99}$$

We are free to choose any gauge we like for $A(x)$. The only constraint is that the gauge-invariant content, which is encoded in he magnetic fluxes through every closed loop $\mathcal{C}$,

$$\Phi_{\mathcal{C}} = \oint\limits_{\mathcal{C}} \boldsymbol{A}\cdot d\boldsymbol{l} \quad , \tag{7.100}$$

must be preserved. On a ring, there is one flux Φ to speak of, and we define the dimensionless flux $\phi = e\Phi/\hbar c = 2\pi\Phi/\phi_0$, where $\phi_0 = hc/e = 4.137\times10^{-7}\,\mathrm{G}\cdot\mathrm{cm}^2$ is the Dirac flux quantum. In a field of $B = 1\,\mathrm{kG}$, a single Dirac quantum is enclosed by a ring of radius $R = 0.11\,\mu\mathrm{m}$. It is convenient to choose a gauge in which A vanishes everywhere along our loop except for a vanishingly small region, in which all the accrued vector potential piles up in a δ-function of strength Φ. The transfer matrix for this infinitesimal region is then

$$\mathcal{M}(\phi) = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix} = e^{i\phi}\cdot\mathbb{I} \quad . \tag{7.101}$$

If $k > 0$ corresponds to clockwise motion around the ring, then the phase accrued is $-\gamma$, which explains the sign of ϕ in the above equation.

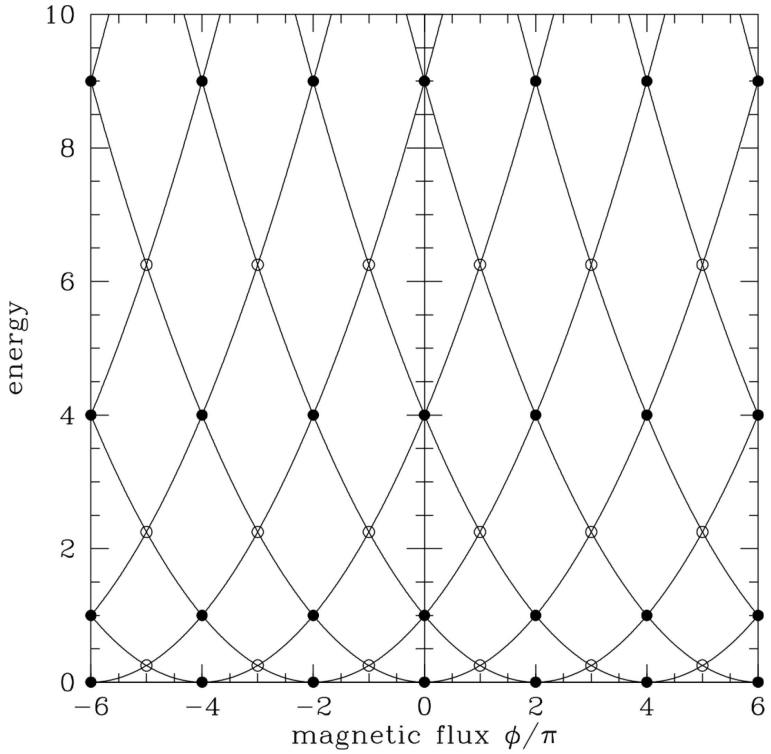


Figure 7.4: Energy *versus* dimensionless magnetic flux for free electrons on a ring. The degeneracies are lifted in the presence of a crystalline potential.

For a pristine ring, then, combining the two transfer matrices gives

$$\mathcal{M} = \begin{pmatrix} e^{ikL} & 0 \\ 0 & e^{-ikL} \end{pmatrix} \begin{pmatrix} e^{i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix} = \begin{pmatrix} e^{ikL}\, e^{i\phi} & 0 \\ 0 & e^{-ikL}\, e^{i\phi} \end{pmatrix} \quad , \tag{7.102}$$

and thus $\det(\mathcal{M} - \mathbb{I}) = 0$ gives the solutions,

$$\begin{aligned} kL &= 2\pi n - \phi \qquad \text{(right-movers)} \\ kL &= 2\pi n + \phi \qquad \text{(left-movers)} \quad . \end{aligned} \tag{7.103}$$

Note that different n values are allowed for right- and left-moving branches since by assumption $k > 0$. We can simplify matters if we simply write $\psi(x) = A\, e^{ikx}$ with k unrestricted in sign, in which case $k = (2\pi n - \phi)/L$ with n chosen from the entire set of integers. The allowed energies for free electrons are then

$$\varepsilon_n(\phi) = \frac{2\pi^2\hbar^2}{mL^2} \cdot \big(n - \frac{\phi}{2\pi}\big)^2 \quad , \tag{7.104}$$

which are plotted in fig. 7.4.

Now let us add in some scatterers. This problem was first considered in a beautiful paper by Büttiker, Imry, and Azbel, *Phys. Rev. A* **30**, 1982 (1984). Consider the ring geometry depicted

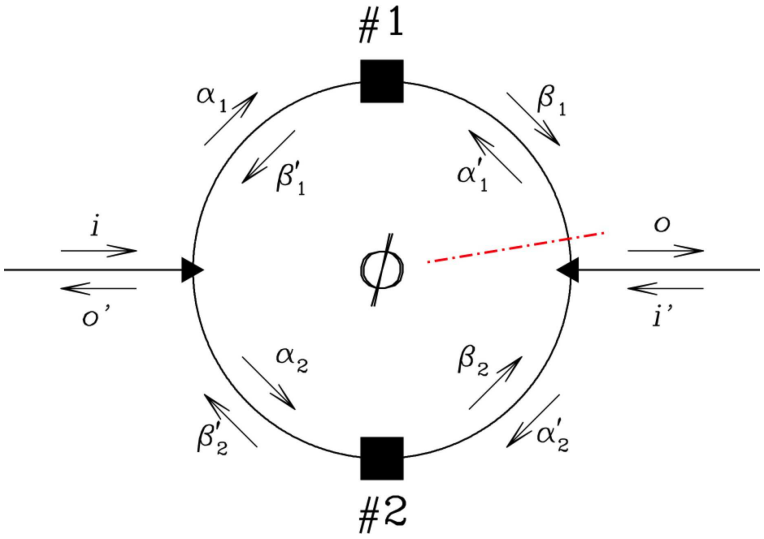


Figure 7.5: Scattering problem for a ring enclosing a flux Φ. The square and triangular blocks represent scattering regions and are describes by 2×2 and 3×3 $\mathcal{S}$-matrices, respectively. The dot-dash line represents a cut across which the phase information due to the enclosed flux is accrued discontinuously.

in fig. 7.5. We want to compute the $\mathcal{S}$-matrix for the ring. We now know how to describe the individual quantum resistors #1 and #2 in terms of $\mathcal{S}$-matrices (or, equivalently, $\mathcal{M}$-matrices). Assuming there is no magnetic field penetrating the wire (or that the wire itself is infinitesimally thin), we have $\mathcal{S} = \mathcal{S}^{\rm t}$ for each scatterer. In this case, we have $t = t' = \sqrt{T}\, e^{i\alpha}$. We know $|r|^2 = |r'|^2 = 1 - |t|^2$, but in general r and r' may have different phases. The most general 2×2 transfer matrix, under conditions of time-reversal symmetry, depends on three parameters, which may be taken to be the overall transmission probability T and two phases:

$$\mathcal{M}(T, \alpha, \beta) = \frac{1}{\sqrt{T}} \begin{pmatrix} e^{i\alpha} & \sqrt{1-T}\, e^{i\beta} \\ \sqrt{1-T}\, e^{-i\beta} & e^{-i\alpha} \end{pmatrix} \quad . \tag{7.105}$$

We can include the effect of free-particle propagation in the transfer matrix $\mathcal{M}$ by multiplying $\mathcal{M}$ on the left and the right by a free propagation transfer matrix of the form

$$\mathcal{N} = \begin{pmatrix} e^{i\theta/4} & 0 \\ 0 & e^{-i\theta/4} \end{pmatrix} \quad , \tag{7.106}$$

where $\theta = kL = 2\pi kR$ is the phase accrued by a particle of wavevector k freely propagating once around the ring. $\mathcal{N}$ is the transfer matrix corresponding to one quarter turn around the ring. One easily finds

$$\mathcal{M} \to \mathcal{N}\, \mathcal{M}(T, \alpha, \beta)\, \mathcal{N} = \mathcal{M}(T, \alpha + \tfrac{1}{2}\theta, \beta). \tag{7.107}$$

For pedagogical reasons, we will explicitly account for the phases due to free propagation, and

write

$$\begin{pmatrix}\beta_1\\ \alpha_1'\end{pmatrix} = \mathcal{N}\,\mathcal{M}_1\,\mathcal{N}\begin{pmatrix}\alpha_1\\ \beta_1'\end{pmatrix} \quad , \qquad \begin{pmatrix}\beta_2'\\ \alpha_2\end{pmatrix} = \mathcal{N}\,\widetilde{\mathcal{M}}_2\,\mathcal{N}\begin{pmatrix}\alpha_2'\\ \beta_2\end{pmatrix} \quad , \tag{7.108}$$

where $\widetilde{\mathcal{M}}_2$ is the transfer matrix for scatterer #2 going from right to left.

EXERCISE: Show that the right-to-left transfer matrix $\widetilde{\mathcal{M}}$ is related to the left-to-right transfer matrix $\mathcal{M}$ according to

$$\widetilde{\mathcal{M}} = \Lambda\,\Sigma\,\mathcal{M}^\dagger\,\Sigma\,\Lambda \quad , \tag{7.109}$$

where

$$\Sigma = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \quad , \qquad \Lambda = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \quad . \tag{7.110}$$

We now have to model the connections between the ring and the leads, which lie at the confluence of three segments. Accordingly, these regions are described by 3×3 $\mathcal{S}$-matrices. The constraints $\mathcal{S}=\mathcal{S}^\dagger$ (unitarity) and $\mathcal{S}=\mathcal{S}^{\mathrm{t}}$ (time-reversal symmetry) reduce the number of independent real parameters in $\mathcal{S}$ from 18 to 5. Further assuming that the scattering is symmetric with respect to the ring branches brings this number down to 3, and finally assuming $\mathcal{S}$ is real reduces the dimension of the space of allowed $\mathcal{S}$-matrices to one. Under these conditions, the most general 3×3 $\mathcal{S}$-matrix may be written

$$\begin{pmatrix}-(a+b) & \sqrt{\epsilon} & \sqrt{\epsilon}\\ \sqrt{\epsilon} & a & b\\ \sqrt{\epsilon} & b & a\end{pmatrix} \tag{7.111}$$

where

$$\begin{aligned}(a+b)^2+2\epsilon &= 1\\ a^2+b^2+\epsilon &= 1 \quad .\end{aligned} \tag{7.112}$$

The parameter ϵ, which may be taken as a measure of the coupling between the ring and the leads ($\epsilon=0$ means ring and leads are decouped) is restricted to the range $0\leqslant\epsilon\leqslant\frac{1}{2}$. There are four solutions for each allowed value of ϵ:

$$a = \pm\tfrac{1}{2}\big(\sqrt{1-2\epsilon}-1\big) \quad , \quad b = \pm\tfrac{1}{2}\big(\sqrt{1-2\epsilon}+1\big) \tag{7.113}$$

and

$$a = \pm\tfrac{1}{2}\big(\sqrt{1-2\epsilon}+1\big) \quad , \quad b = \pm\tfrac{1}{2}\big(\sqrt{1-2\epsilon}-1\big) \quad . \tag{7.114}$$

We choose the first pair, since it corresponds to the case $|b|=1$ when $\epsilon=0$, *i.e.* perfect transmission through the junction. We choose the top sign in (7.113).

We therefore have at the left contact,

$$\begin{pmatrix}o'\\ \alpha_2\\ \alpha_1\end{pmatrix} = \begin{pmatrix}-(a_{\rm L}+b_{\rm L}) & \sqrt{\epsilon}_{\rm L} & \sqrt{\epsilon}_{\rm L}\\ \sqrt{\epsilon}_{\rm L} & a_{\rm L} & b_{\rm L}\\ \sqrt{\epsilon}_{\rm L} & b_{\rm L} & a_{\rm L}\end{pmatrix}\begin{pmatrix}i\\ \beta_2'\\ \beta_1'\end{pmatrix} \quad , \tag{7.115}$$

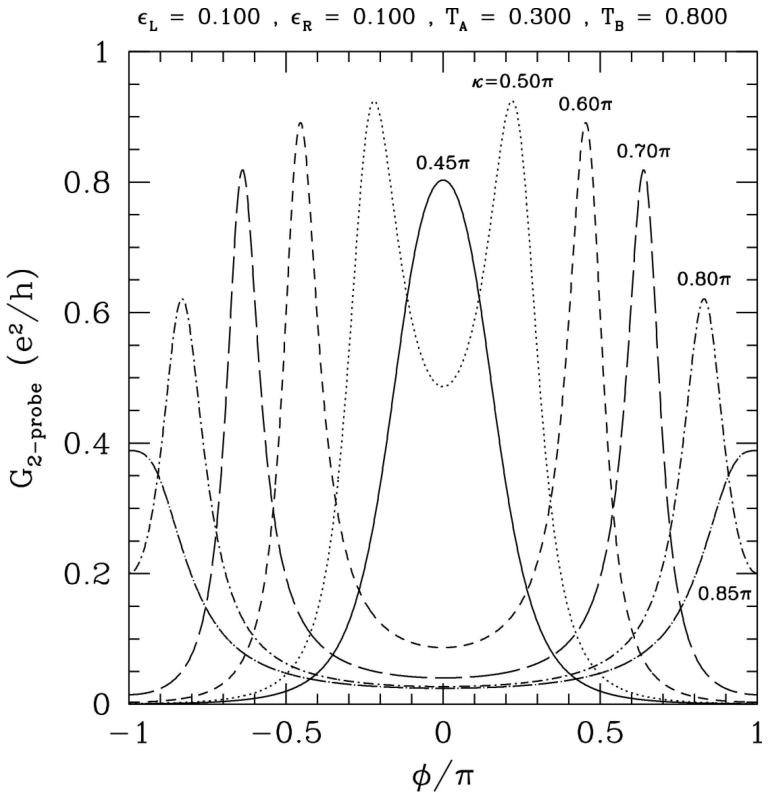


Figure 7.6: Two-probe conductance $G(\phi, \kappa)$ of a model ring with two scatterers. The enclosed magnetic flux is $\phi\hbar c/e$, and $\kappa = 2\pi kR$ (ring radius R). G *vs.* ϕ curves for various values of κ are shown.

and at the right contact

$$\begin{pmatrix} o \\ \widetilde{\alpha}'_1 \\ \alpha'_2 \end{pmatrix} = \begin{pmatrix} -(a_{\rm R} + b_{\rm R}) & \sqrt{\epsilon}_{\rm R} & \sqrt{\epsilon}_{\rm R} \\ \sqrt{\epsilon}_{\rm R} & a_{\rm R} & b_{\rm R} \\ \sqrt{\epsilon}_{\rm R} & b_{\rm R} & a_{\rm R} \end{pmatrix} \begin{pmatrix} i' \\ \widetilde{\beta}_1 \\ \beta_2 \end{pmatrix} \quad , \tag{7.116}$$

where accounting for the vector potential gives us

$$\widetilde{\beta}_1 = e^{i\phi}\,\beta_1 \quad , \quad \widetilde{\alpha}'_1 = e^{i\phi}\,\alpha'_1 \quad . \tag{7.117}$$

We set $i = 1$ and $i' = 0$, so that the transmission and reflection amplitudes are obtained from $t = o$ and $r = o'$.

From (7.115), we can derive the relation

$$\begin{pmatrix} \alpha_1 \\ \beta'_1 \end{pmatrix} = \overbrace{\frac{1}{b_{\rm L}} \begin{pmatrix} b_{\rm L}^2 - a_{\rm L}^2 & a_{\rm L} \\ -a_{\rm L} & 1 \end{pmatrix}}^{\mathcal{Q}_{\rm L}} \begin{pmatrix} \beta'_2 \\ \alpha_2 \end{pmatrix} + \frac{\sqrt{\epsilon_{\rm L}}}{b_{\rm L}} \begin{pmatrix} b_{\rm L} - a_{\rm L} \\ -1 \end{pmatrix} \quad . \tag{7.118}$$

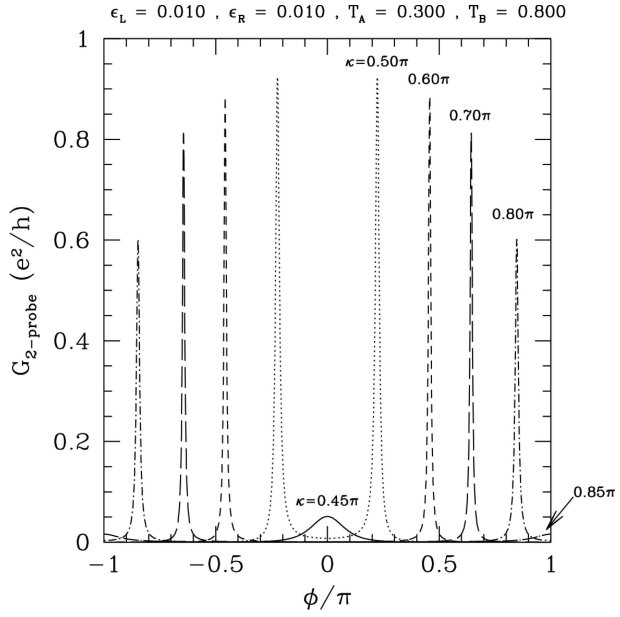


Figure 7.7: Two-probe conductance $G(\phi, \kappa)$ of a model ring with two scatterers. The enclosed magnetic flux is $\phi\hbar c/e$, and $\kappa = 2\pi kR$ (ring radius R). G *vs.* ϕ curves for various values of κ are shown. The coupling between leads and ring is one tenth as great as in fig. 7.6, and accordingly the resonances are much narrower. Note that the resonances at $\kappa = 0.45\pi$ and $\kappa = 0.85\pi$ are almost completely suppressed.

Similarly, from (7.116), we have

$$\begin{pmatrix} \alpha_2' \\ \beta_2 \end{pmatrix} = \overbrace{\frac{1}{b_{\rm R}} \begin{pmatrix} b_{\rm R}^2 - a_{\rm R}^2 & a_{\rm R} \\ -a_{\rm R} & 1 \end{pmatrix}}^{\mathcal{Q}_{\rm R}} e^{i\phi} \begin{pmatrix} \beta_1 \\ \alpha_1' \end{pmatrix} \quad . \tag{7.119}$$

The matrices $\mathcal{Q}_{\rm L}$ and $\mathcal{Q}_{\rm R}$ resemble transfer matrices. However, they are not pseudo-unitary: $\mathcal{Q}^\dagger \Sigma \mathcal{Q} \neq \Sigma$. This is because some of the flux can leak out along the leads. Indeed, when $\epsilon = 0$, we have $b = 1$ and $a = 0$, hence $\mathcal{Q} = 1$, which *is* pseudo-unitary (*i.e.* flux preserving). Combining these results with those in (7.108), we obtain the solution

$$\begin{pmatrix} \alpha_1 \\ \beta_1' \end{pmatrix} = \left\{ \mathbb{I} - e^{i\phi}\, \mathcal{Q}_{\rm L}\, \mathcal{N}\, \widetilde{\mathcal{M}}_2\, \mathcal{N}\, \mathcal{Q}_{\rm R}\, \mathcal{N}\, \mathcal{M}_1\, \mathcal{N} \right\}^{-1} \frac{\sqrt{\epsilon_{\rm L}}}{b_{\rm L}} \begin{pmatrix} b_{\rm L} - a_{\rm L} \\ -1 \end{pmatrix} \quad . \tag{7.120}$$

From this result, using (7.108), all the flux amplitudes can be obtained.

We can define the effective ring transfer matrix $\mathcal{P}$ as

$$\mathcal{P} \equiv e^{i\phi}\, \mathcal{Q}_{\rm L}\, \mathcal{N}\, \widetilde{\mathcal{M}}_2\, \mathcal{N}\, \mathcal{Q}_{\rm R}\, \mathcal{N}\, \mathcal{M}_1\, \mathcal{N} \quad , \tag{7.121}$$

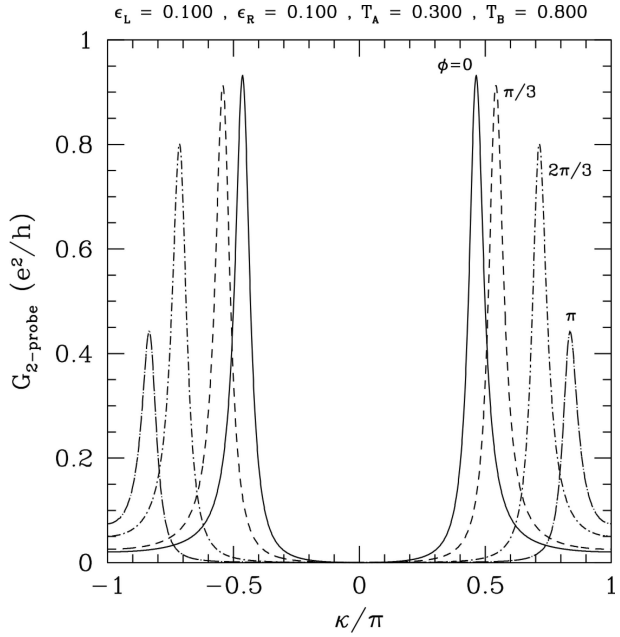


Figure 7.8: Two-probe conductance $G(\phi, \kappa)$ of a model ring with two scatterers. The enclosed magnetic flux is $\phi\hbar c/e$, and $\kappa = 2\pi kR$ (ring radius R). G *vs.* κ curves for various values of ϕ are shown.

which has the following simple interpretation. Reading from right to left, we first move $\frac{1}{4}$-turn clockwise around the ring ($\mathcal{N}$). Then we encounter scatterer #1 ($\mathcal{M}_1$). After another quarter turn ($\mathcal{N}$), we encounter the right T-junction ($\mathcal{Q}_{\rm R}$). Then it's yet another quarter turn ($\mathcal{N}$) until scatterer #2 ($\mathcal{M}_2$), and one last quarter turn ($\mathcal{N}$) brings us to the left T-junction ($\mathcal{Q}_{\rm L}$), by which point we have completed one revolution. As the transfer matrix acts on both right-moving and left-moving flux amplitudes, it accounts for both clockwise as well as counterclockwise motion around the ring. The quantity

$$\bigl\{\mathbb{I} - \mathcal{P}\bigr\}^{-1} = 1 + \mathcal{P} + \mathcal{P}^2 + \mathcal{P}^3 + \ldots \quad , \tag{7.122}$$

then sums up over all possible integer windings around the ring. In order to properly account for the effects of the ring, an infinite number of terms must be considered; these may be resummed into the matrix inverse in (7.121). The situation is analogous to what happens when an electromagnetic wave reflects off a thin dielectric slab. At the top interface, the wave can reflect. However, it can also refract, entering the slab, where it may undergo an arbitrary number of internal reflections before exiting.

The transmission coefficient t is just the outgoing flux amplitude: $t = o$. We have from

(7.116,7.108) that

$$\begin{aligned} t &= \sqrt{\epsilon_{\rm R}}\,\bigl(\beta_1\,e^{i\phi}+\beta_2\bigr) \\ &= \frac{\sqrt{\epsilon_{\rm R}}}{b_{\rm R}}\,\begin{pmatrix} b_{\rm R}-a_{\rm R} & 1\end{pmatrix}\begin{pmatrix}\beta_1\\ \alpha_1'\end{pmatrix} \\ &= \frac{\sqrt{\epsilon_{\rm L}\,\epsilon_{\rm R}}}{b_{\rm L}\,b_{\rm R}}\,\bigl(\mathcal{Y}_{12}^{-1}+\mathcal{Y}_{22}^{-1}-\mathcal{Y}_{11}^{-1}-\mathcal{Y}_{21}^{-1}\bigr) \end{aligned} \tag{7.123}$$

where

$$\mathcal{Y} = \mathcal{Q}_{\rm L}\,\mathcal{N}\,\widetilde{\mathcal{M}}_2\,\mathcal{N}\,\mathcal{Q}_{\rm R} - e^{-i\phi}\,\mathcal{N}^{-1}\mathcal{M}_1^{-1}\mathcal{N}^{-1} \quad . \tag{7.124}$$

It is straightforward to numerically implement the above calculation. Sample results are shown in figs. 7.6 and 7.8.

7.4 Universal Conductance Fluctuations in Dirty Metals

The conductance of a disordered metal is a function of the strength and location of the individual scatterers. We now ask, how does the conductance fluctuate when the position or strength of a scatterer or a group of scatterers is changed. From the experimental point of view, this seems a strange question to ask, since we generally do not have direct control over the position of individual scatterers within a bulk system. However, we can imagine changing some external parameter, such as the magnetic field B or the chemical potential μ (via the density n). Using computer modeling, we can even 'live the dream' of altering the position of a single scatterer to investigate its effect on the overall conductance. Naïvely, we would expect there to be very little difference in the conductance if we were to, say, vary the position of a single scatterer by a distance ℓ, or if we were to change the magnetic field by $\Delta B = \phi_0/A$, where A is the cross sectional area of the system. Remarkably, though, what is found both experimentally and numerically is that the conductance exhibits fluctuations with varying field B, chemical potential μ, or impurity configuration (in computer models). The root-mean-square magnitude of these fluctuations for a given sample is the same as that between different samples, and is on the order $\delta G \sim e^2/h$. These *universal conductance fluctuations* (UCF) are independent on the degree of disorder, the sample size, the spatial dimensions, so long as the inelastic mean free path (or *phase breaking length*) satisfies $L_\phi > L$, *i.e.* the system is mesoscopic.

Theoretically the phenomenon of UCF has a firm basis in diagrammatic perturbation theory. Here we shall content ourselves with understanding the phenomenon on a more qualitative level, following the beautiful discussion of P. A. Lee in *Physica* **140A**, 169 (1986). We begin with the multichannel Landauer formula,

$$G = \frac{e^2}{h}\,{\rm Tr}\,tt^\dagger = \frac{e^2}{h}\sum_{a,j=1}^{N_{\rm c}} |t_{aj}|^2 \quad . \tag{7.125}$$

The transmission amplitudes t_{aj} can be represented as a quantum mechanical sum over paths γ,

$$t_{aj} = \sum_\gamma \mathcal{A}_{aj}(\gamma) \quad , \tag{7.126}$$

where $\mathcal{A}_{aj}(\gamma)$ is the probability amplitude for Feynman path γ to connect channels j and a. The sum is over all such Feynman paths, and ultimately we must project onto a subspace of definite energy – this is, in Lee's own words, a 'heuristic argument'. Now assume that the $\mathcal{A}_{aj}(\gamma)$ are independent complex random variables. The fluctuations in $|t_{aj}|^2$ are computed from

$$\begin{aligned} \big\langle |t_{aj}|^4 \big\rangle &= \sum_{\substack{\gamma_1,\gamma_2\\ \gamma_3,\gamma_4}} \big\langle \mathcal{A}_{aj}(\gamma_1)\, \mathcal{A}^*_{aj}(\gamma_2)\, \mathcal{A}_{aj}(\gamma_3)\, \mathcal{A}^*_{aj}(\gamma_4) \big\rangle \\ &= 2 \Big\langle \sum_\gamma \big| \mathcal{A}_{aj}(\gamma) \big|^2 \Big\rangle^2 + \Big\langle \sum_\gamma \big| \mathcal{A}_{aj}(\gamma) \big|^4 \Big\rangle \\ &= 2 \big\langle |t_{aj}|^2 \big\rangle^2 \cdot \Big\{ 1 + \mathcal{O}(M^{-1}) \Big\} \quad , \end{aligned} \tag{7.127}$$

where M is the (extremely large) number of paths in the sum over γ. As the $\mathcal{O}(M^{-1})$ term is utterly negligible, we conclude

$$\frac{\big\langle |t_{aj}|^4 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle^2}{\big\langle |t_{aj}|^2 \big\rangle^2} = 1 \quad . \tag{7.128}$$

Now since

$$\begin{aligned} \mathrm{var}(G) &= \langle G^2 \rangle - \langle G \rangle^2 \\ &= \frac{e^4}{h^2} \sum_{\substack{a,a'\\ j,j'}} \Big\{ \big\langle |t_{aj}|^2 \, |t_{a'j'}|^2 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle \big\langle |t_{aj}|^2 \big\rangle \Big\} \quad , \end{aligned} \tag{7.129}$$

we also need to know about the correlation between $|t_{aj}|^2$ and $|t_{a'j'}|^2$. The simplest assumption is to assume they are uncorrelated unless $a = a'$ and $j = j'$, *i.e.*

$$\big\langle |t_{aj}|^2 \, |t_{a'j'}|^2 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle \big\langle |t_{aj}|^2 \big\rangle = \Big(\big\langle |t_{aj}|^4 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle^2 \Big) \delta_{aa'}\, \delta_{jj'} \quad , \tag{7.130}$$

in which case the conductance is a given by a sum of $N_{\rm c}^2$ independent real random variables, each of which has a standard deviation equal to its mean, *i.e.* equal to $\big\langle |t_{aj}|^2 \big\rangle$. According to the central limit theorem, then, the rms fluctuations of G are given by

$$\Delta G = \sqrt{\mathrm{var}(G)} = \frac{e^2}{h} N_{\rm c} \, \big\langle |t_{aj}|^2 \big\rangle \quad . \tag{7.131}$$

Further assuming that we are in the Ohmic regime where $G = \sigma \, L^{d-2}$, with $\sigma \approx (e^2/h)\, k_{\rm F}^{d-1} \ell$, and $N_{\rm c} \approx (k_{\rm F} L)^{d-1}$, we finally conclude

$$\big\langle |t_{aj}|^2 \big\rangle \approx \frac{1}{N_{\rm c}} \cdot \frac{\ell}{L} \quad \Longrightarrow \quad \Delta G \approx \frac{e^2}{h} \frac{\ell}{L} \quad . \tag{7.132}$$

This result is much smaller than the correct value of $\Delta G \sim (e^2/h)$.

To reiterate the argument in terms of the dimensionless conductance g,

$$g = \sum_{a,j} |t_{aj}|^2 \simeq N_{\rm c} \cdot \frac{\ell}{L} \quad \Longrightarrow \quad \big\langle |t_{aj}|^2 \big\rangle = \frac{g}{N_{\rm c}^2} \quad , \tag{7.133}$$

and thus we might expect

$$\begin{aligned} {\rm var}(g) &= \sum_{\substack{a,a' \\ j,j'}} \Big\{ \big\langle |t_{aj}|^2 \, |t_{a'j'}|^2 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle \big\langle |t_{aj}|^2 \big\rangle \Big\} \\ &\approx \sum_{aj} \Big\{ \big\langle |t_{aj}|^4 \big\rangle - \big\langle |t_{aj}|^2 \big\rangle^2 \Big\} \\ &= \sum_{aj} \big\langle |t_{aj}|^2 \big\rangle^2 \approx N_{\rm c}^2 \cdot \Big(\frac{g}{N_{\rm c}^2} \Big)^2 = \frac{g^2}{N_{\rm c}^2} \quad (???) \quad , \end{aligned} \tag{7.134}$$

and therefore

$$\sqrt{{\rm var}(g)} \approx \frac{g}{N_{\rm c}} = \frac{\ell}{L} \quad ({\rm WRONG!}) \quad . \tag{7.135}$$

What went wrong? The problem lies in the assumption that the contributions $\mathcal{A}_{aj}(\gamma)$ are independent for different paths γ. The reason is that in disordered systems there are certain preferred channels within the bulk along which the conduction paths run. Different paths γ will often coincide along these channels. A crude analogy: whether you're driving from La Jolla to Burbank, or from El Cajon to Malibu, eventually you're going to get on the 405 freeway – anyone driving from the San Diego area to the Los Angeles area must necessarily travel along one of a handful of high-volume paths. The same is not true of reflection, though! Those same two hypothetical drivers executing local out-and-back trips from home will in general travel along completely different, hence uncorrelated, routes. Accordingly, let us compute ${\rm var}(N_{\rm c} - g)$, which is identical to ${\rm var}(g)$, but is given in terms of a sum over reflection coefficients. We will see that making the same assumptions as we did in the case of the transmission coefficients produces the desired result. We need only provide a sketch of the argument:

$$N_{\rm c} - g = \sum_{i,j} |r_{ij}|^2 \simeq N_{\rm c} \cdot \Big(1 - \frac{\ell}{L} \Big) \quad \Longrightarrow \quad \big\langle |r_{ij}|^2 \big\rangle = \frac{N_{\rm c} - g}{N_{\rm c}^2} \quad , \tag{7.136}$$

so

$$\begin{aligned} {\rm var}(N_{\rm c} - g) &= \sum_{\substack{i,i' \\ j,j'}} \Big\{ \big\langle |r_{ij}|^2 \, |r_{i'j'}|^2 \big\rangle - \big\langle |r_{ij}|^2 \big\rangle \big\langle |r_{i'j'}|^2 \big\rangle \Big\} \\ &\approx \sum_{ij} \Big\{ \big\langle |r_{ij}|^4 \big\rangle - \big\langle |r_{ij}|^2 \big\rangle^2 \Big\} \\ &= \sum_{ij} \big\langle |r_{ij}|^2 \big\rangle^2 \approx N_{\rm c}^2 \cdot \Big(\frac{N_{\rm c} - g}{N_{\rm c}^2} \Big)^2 = \Big(1 - \frac{\ell}{L} \Big)^2 \quad , \end{aligned} \tag{7.137}$$

and therefore

$$\sqrt{\mathrm{var}(N_{\rm c}-g)} = \sqrt{\mathrm{var}(g)} = \left(1-\frac{\ell}{L}\right) \quad . \tag{7.138}$$

The assumption of uncorrelated *reflection* paths is not as problematic as that of uncorrelated *transmission* paths. Again, this is due to the existence of preferred internal channels within the bulk, along which transmission occurs. In reflection, though there is no need to move along identical segments.

There is another bonus to thinking about reflection *versus* transmission. Let's express the reflection probability as a sum over paths, *viz.*

$$|r_{ij}|^2 = \sum_{\gamma,\gamma'} \mathcal{A}_{ij}(\gamma)\,\mathcal{A}^*_{ij}(\gamma') \quad . \tag{7.139}$$

Each path γ will have a time-reversed mate $\gamma^{\rm T}$ for which, in the absence of external magnetic fields,

$$\mathcal{A}_{ij}(\gamma) = \mathcal{A}_{ij}(\gamma^{\rm T}) \quad . \tag{7.140}$$

This is because the action functional,

$$S[\boldsymbol{r}(t)] = \int\limits_{t_1}^{t_2} dt\,\left\{\tfrac{1}{2}m\dot{\boldsymbol{r}}^2 - V(\boldsymbol{r}) - \frac{e}{c}\boldsymbol{A}(\boldsymbol{r})\cdot\dot{\boldsymbol{r}}\right\} \tag{7.141}$$

satisfies

$$S[\boldsymbol{r}(t)] = S[\boldsymbol{r}(-t)] \qquad \text{if } \boldsymbol{B}=0 \quad . \tag{7.142}$$

There is, therefore, an extra *negative* contribution to the conductance G arising from *phase coherence of time-reversed paths*. In the presence of an external magnetic field, the path γ and its time-reversed mate $\gamma^{\rm T}$ have a relative phase $\eta = 4\pi\Phi_\gamma/\phi_0$, where Φ_γ is the magnetic flux enclosed by the path γ. A magnetic field, then, tends to destroy the phase coherence between time-reversed paths, and hence we expect a *positive magnetoconductance* (*i.e.* negative magnetoresistance) in mesoscopic disordered metals.

Conductance fluctuations in metallic rings

The conductance of a ring must be periodic under $\Phi \to \Phi + n\phi_0$ for any integer n – rings with flux differing by an integer number of Dirac quanta are gauge-equivalent, provided no magnetic field penetrates the ring itself. The conductance as a function of the enclosed flux Φ must be of the form

$$G(\Phi) = G_{\rm cl} + \sum_{m=1}^{\infty} G_m \cos\left(\frac{2\pi m\Phi}{\phi_0} + \alpha_m\right) \tag{7.143}$$

where $G_{\rm cl}$ is the classical (Boltzmann) conductance of the ring. The second harmonic $G_{m=2}$ is usually detectable and is in many cases much larger than the $m=1$ term. The origin of

the $m = 2$ term, which is periodic under $\Phi \to \Phi + \frac{1}{2}\phi_0$, lies in the interference between time-reversed paths of winding number ± 1. The $m = 1$ fundamental is easily suppressed, *e.g.* by placing several rings in series.

7.4.1 Weak localization

A more rigorous discussion of enhanced backscattering was first discussed by Altshuler, Aronov, and Spivak (AAS) in 1981. AAS showed that there are corrections to Boltzmann transport of the form $\sigma = \sigma_0 + \delta\sigma$, where $\sigma_0 = ne^2\tau/m^*$ is the Drude conductivity and (including a factor of 2 for spin),

$$\delta\sigma = -\frac{2e^2}{h}\,\ell^2 \cdot \frac{1}{V}\int d^d r\, \mathcal{C}(\boldsymbol{r},\boldsymbol{r}) \quad , \tag{7.144}$$

where ℓ is the elastic mean free path and $\mathcal{C}(\boldsymbol{r},\boldsymbol{r}')$ is the *Cooperon propagator*, which satisfies

$$\left\{-\ell^2\Big(\boldsymbol{\nabla} + \frac{2ie}{\hbar c}\,\boldsymbol{A}(\boldsymbol{r})\Big)^2 + \frac{\tau}{\tau_\phi}\right\}\mathcal{C}(\boldsymbol{r},\boldsymbol{r}') = \delta(\boldsymbol{r}-\boldsymbol{r}') \quad , \tag{7.145}$$

where τ_ϕ is the inelastic collision time, $\tau_\phi = L_\phi^2/D$, where $D = v_{\rm F}\ell = \ell^2/\tau$ is the diffusion constant. The linear differential operator

$$\mathcal{L} = -\ell^2\Big(\boldsymbol{\nabla} + \frac{2ie}{\hbar c}\,\boldsymbol{A}(\boldsymbol{r})\Big)^2 \tag{7.146}$$

bears a strong resemblance to the Hamiltonian of a particle of charge $e^* = 2e$ in an external magnetic field $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$. Expanding in eigenfunctions of L, we obtain the solution

$$\begin{aligned}\Big(\mathcal{L} + \frac{\tau}{\tau_\phi}\Big)\,\mathcal{C}(\boldsymbol{r},\boldsymbol{r}') &= \delta(\boldsymbol{r}-\boldsymbol{r}') \\ \mathcal{L}\,\psi_\alpha(\boldsymbol{r}) &= \lambda_\alpha\psi_\alpha(\boldsymbol{r}) \\ \mathcal{C}(\boldsymbol{r},\boldsymbol{r}') &= \sum_\alpha \frac{\psi_\alpha(\boldsymbol{r})\,\psi_\alpha^*(\boldsymbol{r}')}{\lambda_\alpha + \frac{\tau}{\tau_\phi}} \quad ,\end{aligned} \tag{7.147}$$

which resembles a Green's function from quantum mechanics, where the energy parameter is identified with $-\tau/\tau_\phi$. There is accordingly a path integral representation for $\mathcal{C}(\boldsymbol{r},\boldsymbol{r}')$:

$$\mathcal{C}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \int ds\, e^{-is\tau/\tau_\phi}\int\limits_{\substack{\boldsymbol{r}(0)=\boldsymbol{r}_1\\ \boldsymbol{r}(s)=\boldsymbol{r}_2}} \mathcal{D}\boldsymbol{r}[u]\,\exp\left\{i\int_0^s du\left[\frac{1}{4\ell^2}\Big(\frac{\partial\boldsymbol{r}}{\partial u}\Big)^2 - \frac{2e}{\hbar c}\boldsymbol{A}(\boldsymbol{r})\cdot\frac{\partial\boldsymbol{r}}{\partial u}\right]\right\} \quad . \tag{7.148}$$

Notice that there is no potential term $V(\boldsymbol{r})$ in the action of (7.148). The effect of the static random potential here is to provide a 'step length' ℓ for the propagator. According to the AAS result

(7.144), the corrections to the conductivity involve paths which begin and end at the same point $\boldsymbol{r}$ in space. The charge $e^* = 2e$ which appears in the Cooperon action arises from adding up contributions due to time-reversed paths, as we saw earlier.

Let's try to compute $\delta\sigma$, which is called the *weak localization* correction to the conductivity. In the absence of an external field, the eigenvalues of $\mathcal{L}$ are simply $\boldsymbol{k}^2\ell^2$, where $k_\mu = 2\pi n_\mu/L$. We then have

$$\mathcal{C}(\boldsymbol{r},\boldsymbol{r}) = \int \frac{d^dk}{(2\pi)^d}\,\frac{1}{\boldsymbol{k}^2\ell^2 + \frac{\tau}{\tau_\phi}} \quad . \tag{7.149}$$

A cutoff Λ is needed in dimensions $d \geqslant 2$ in order to render the integral convergent in the ultraviolet. This cutoff is the inverse step size for diffusion: $\Lambda \sim \ell^{-1}$. This gives, for $\tau \ll \tau_\phi$,

$$\delta\sigma_d \sim \frac{e^2}{h}\cdot\begin{cases} -\ell^{-1} & (d=3) \\ -\ln(\tau_\phi/\tau) & (d=2) \\ -\ell\sqrt{\tau_\phi/\tau} & (d=1) \end{cases} \quad . \tag{7.150}$$

Let us now compute the magnetoconductance in $d = 2$. In the presence of a uniform magnetic field, $\mathcal{L}$ has evenly spaced eigenvalues

$$\lambda_n = (n+\tfrac{1}{2})\left(\frac{2\ell}{\ell_B}\right)^2 \quad , \tag{7.151}$$

where $\ell_B = \sqrt{\hbar c/eB}$ is the magnetic length (for a charge $q = -e$ electron). Each of these Landau levels has a macroscopic degeneracy of $2eBA/hc = A/\pi\ell_B^2$, Thus,

$$\begin{aligned}\delta\sigma(B) &= -\frac{2e^2}{h}\,\frac{\ell^2}{\pi\ell_B^2}\sum_{n=0}^{\ell_B^2/4\ell^2}\frac{1}{(n+\frac{1}{2})\frac{4\ell^2}{\ell_B^2}+\frac{\tau}{\tau_\phi}} \\ &= -\frac{e^2}{2\pi h}\sum_{n=0}^{\ell_B^2/4\ell^2}\frac{1}{n+\frac{1}{2}+\frac{\ell_B^2}{L_\phi^2}} \quad ,\end{aligned} \tag{7.152}$$

where we have invoked $\tau/\tau_\phi = \ell^2/L_\phi^2$. The magnetoconductance is then

$$\delta\sigma(B) - \delta\sigma(0) = -\frac{1}{2\pi}\,\frac{e^2}{h}\left\{\Psi\Big(\frac{1}{2}+\frac{\ell_B^2}{4\ell^2}\Big) - \Psi\Big(\frac{1}{2}+\frac{\ell_B^2}{4L_\phi^2}\Big)\right\} \quad , \tag{7.153}$$

where

$$\Psi(z) = \frac{1}{\Gamma(z)}\,\frac{d\Gamma(z)}{dz} = \ln z + \frac{1}{2z} - \frac{1}{12z^2} + \ldots \tag{7.154}$$

is the digamma function. If the field is weak, so that $\ell_B \gg \ell$, then

$$\delta\sigma(B) - \delta\sigma(0) = +\frac{1}{6\pi}\,\frac{e^2}{h}\left(\frac{L_\phi}{\ell_B}\right)^2 \quad , \tag{7.155}$$

which is positive, as previously discussed. The magnetic field suppresses phase coherence between time-reversed paths, and thereby promotes diffusion by suppressing the resonant backscattering contributions to $\delta\sigma$. At large values of the field, the behavior is logarithmic. Generally, we can write

$$\delta\sigma(B) - \delta\sigma(0) = \frac{1}{2\pi}\,\frac{e^2}{h}\,f\Big(\frac{B}{B_\phi}\Big) \quad , \tag{7.156}$$

where

$$\frac{2\pi B_\phi}{B} \equiv \frac{\ell_B^2}{4L_\phi^2} \quad \Longrightarrow \quad B_\phi = \frac{\phi_0}{8\pi L_\phi^2} \tag{7.157}$$

and

$$f(x) = \ln x + \Psi\big(\tfrac{1}{2} + \tfrac{1}{x}\big) = \begin{cases} \frac{x^2}{24} & \text{as } x \to 0 \\ \ln x - 1.96351\ldots & \text{as } x \to \infty \end{cases} \quad . \tag{7.158}$$

7.5 Anderson Localization

In 1958, P. W. Anderson proposed that static disorder could lead to localization of electronic eigenstates in a solid. Until this time, it was generally believed that disorder gave rise to an elastic scattering length ℓ and a diffusion constant $D = v_{\rm F}\ell$. The diffusion constant is related to the electrical conductivity through the Einstein relation: $\sigma = \frac{1}{2}e^2 D(\varepsilon_{\rm F})\mathcal{N}(\varepsilon_{\rm F})$. If the states at the Fermi level are *localized*, then $D(\varepsilon_{\rm F}) = 0$.

Anderson considered an electron propagating in a random potential:

$$\mathcal{H} = -\frac{\hbar^2}{2m}\boldsymbol{\nabla}^2 + V(\boldsymbol{r}) \quad , \tag{7.159}$$

where $V(\boldsymbol{r})$ is chosen from an ensemble of random functions. Physically, $V(\boldsymbol{r})$ is bounded and smooth, although often theorists often study uncorrelated 'white noise' potentials where the ensemble is described by the distribution functional

$$P[V(\boldsymbol{r})] = \exp\Big\{-\frac{1}{2\gamma}\int d^d\!r\, V^2(\boldsymbol{r})\Big\} \quad , \tag{7.160}$$

for which

$$\big\langle V(\boldsymbol{r})\,V(\boldsymbol{r}')\big\rangle = \gamma\,\delta(\boldsymbol{r}-\boldsymbol{r}') \quad . \tag{7.161}$$

A tight binding version of this model would resemble

$$\mathcal{H} = -t\sum_{\langle ij\rangle}\big(c_i^\dagger\,c_j + c_j^\dagger\,c_i\big) + \sum_i \varepsilon_i\,c_i^\dagger\,c_i \quad , \tag{7.162}$$

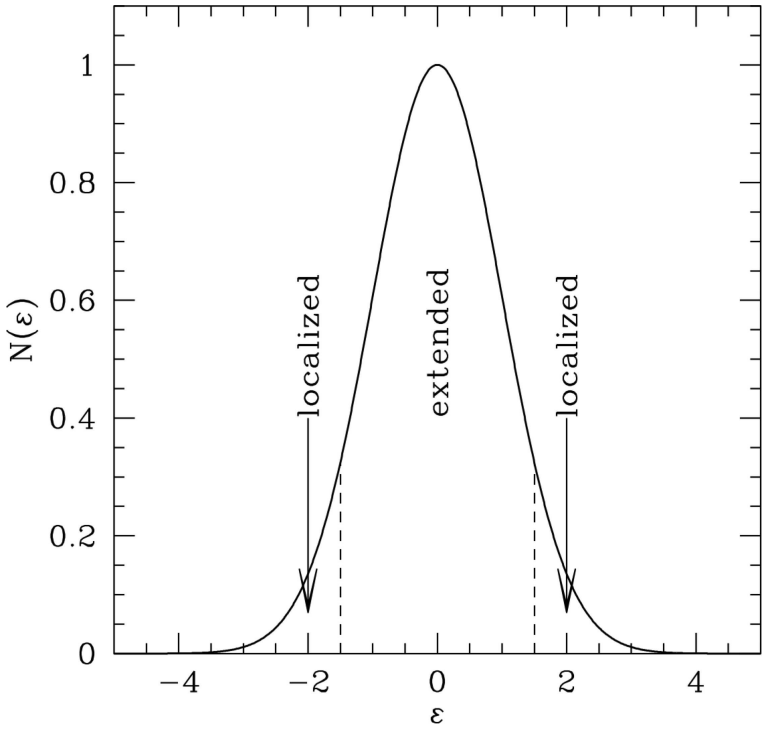


Figure 7.9: Schematic picture of density of states in a disordered system showing mobility edges (dashed lines) and localized states in band tails.

where the single site energies $\{\varepsilon_i\}$ are independently distributed according to some function $p(\varepsilon)$. The first term can be diagonalized by Fourier transform:

$$\mathcal{H}_0 = -t \sum_{\langle ij \rangle} \left(c_i^\dagger \, c_j + c_j^\dagger \, c_i \right) = -zt \sum_{\boldsymbol{k}} \gamma_{\boldsymbol{k}} \, c_{\boldsymbol{k}}^\dagger \, c_{\boldsymbol{k}} \quad , \tag{7.163}$$

where $\gamma_{\boldsymbol{k}} = z^{-1} \sum_{\boldsymbol{\delta}} e^{i\boldsymbol{k}\cdot\boldsymbol{\delta}}$, where δ is a nearest neighbor direct lattice vector and z is the lattice coordination number; the bandwidth is $2zt$. What happens when we add the random potential term to (7.163)? Suppose the width of the distribution $p(\varepsilon)$ is W, *e.g.* $p(\varepsilon) = W^{-1} \, \Theta(\frac{1}{4}W^2 - \varepsilon^2)$, with $W \ll zt$. We expect that the band edges shift from $\pm zt$ to $\pm(zt + \frac{1}{2}W)$. The density of states must vanish for $|\varepsilon| > zt + \frac{1}{2}W$; the regions in the vicinity of $\pm(zt + \frac{1}{2}W)$ are known as *Lifshitz tails*.

Aside from the formation of Lifshitz tails, the density of states doesn't change much. What does change is the character of the eigenfunctions. Suppose we can find a region of contiguous sites all of which have energies $\varepsilon_i \approx \frac{1}{2}W$. Then we could form an approximate eigenstate by concentrating the wavefunction in this region of sites, setting its phase to be constant throughout. This is an example of a *localized state*. We can think of such a state as a particle in a box – the electron binds itself to local fluctuations in the potential. Outside this region, the wavefunction decays, typically exponentially. Scattering states are then extended states, and are associated

with 'average' configurations of the $\{\varepsilon_i\}$. The typical spatial extent of the localized states is given by the localization length $\xi(\varepsilon)$. The localization length diverges at the *mobility edges* as

$$\xi(\varepsilon) \sim |\varepsilon - \varepsilon_{\rm c}|^{-\nu} \quad . \tag{7.164}$$

There is no signature of the mobility edge in the density of states itself – $\mathcal{N}(\varepsilon)$ is completely smooth through the mobility edge $\varepsilon_{\rm c}$.

7.5.1 Characterization of localized and extended states

One way of characterizing the localization properties of quantum mechanical eigenstates (of $\mathcal{H}$) is to compute the *participation ratio* Q. The inverse participation ratio for the state α is given by

$$\begin{aligned} Q_\alpha^{-1} &= \sum_i \big|\psi_\alpha(i)\big|^4 \Big/ \Big(\sum_i \big|\psi_\alpha(i)\big|^2 \Big)^2 \qquad \text{(discrete)} \\ &= \int d^dx \, \big|\psi_\alpha(\boldsymbol{x})\big|^4 \Big/ \Big(\int d^dx \, \big|\psi_\alpha(\boldsymbol{x})\big|^2 \Big)^2 \qquad \text{(continuous)} \end{aligned} \tag{7.165}$$

Consider the discrete case. If ψ_α is localized on a single site, then we have $\sum_i \big|\psi_\alpha(i)\big|^k = 1$ for all k, *i.e.* $Q_\alpha = 1$. But if ψ_α is spread evenly over N sites, then $Q_\alpha^{-1} = N/N^2 = N^{-1}$, and $Q_\alpha = N$. Hence, Q_α tells us approximately how many states $\big|\, i \,\big\rangle$ *participate* in the state $\big|\, \psi_\alpha \,\big\rangle$. The dependence of Q_α on the system size (linear dimension) L can be used as a diagnostic:

$$\begin{aligned} \big|\, \psi_\alpha \,\big\rangle \quad \text{localized} \quad &\Longrightarrow \quad Q_\alpha \propto L^0 \\ \big|\, \psi_\alpha \,\big\rangle \quad \text{extended} \quad &\Longrightarrow \quad Q_\alpha \propto L^\beta \quad , \end{aligned} \tag{7.166}$$

where $\beta > 0$.

Another way to distinguish extended from localized states is to examine their sensitivity to boundary conditions:

$$\psi(x_1\,,\,\ldots\,,\,x_\mu + L\,,\,\ldots\,,\,x_d) = e^{i\theta_\mu}\,\psi(x_1\,,\,\ldots\,,\,x_\mu\,,\,\ldots\,,\,x_d) \quad , \tag{7.167}$$

where $\mu \in \{1, \ldots, d\}$. For periodic boundary conditions $\theta_\mu = 0$, while antiperiodic boundary conditions have $\theta_\mu = \pi$. For plane wave states, this changes the allowed wavevectors, so that $k_\mu \to k_\mu + \pi/L$. Thus, for an extended state,

$$\delta\varepsilon_{\rm ext} = \varepsilon^{\rm pbc} - \varepsilon^{\rm apbc} \simeq \frac{\partial\varepsilon}{\partial k}\,\delta k \propto L^{-1} \quad . \tag{7.168}$$

For a localized state,

$$\delta\varepsilon_{\rm loc} \propto e^{-L/\xi(\varepsilon)} \quad . \tag{7.169}$$

One defines the dimensionless *Thouless number*

$$\mathrm{Th}(\varepsilon, L) = \left|\varepsilon^{\mathrm{pbc}} - \varepsilon^{\mathrm{apbc}}\right| \cdot \mathcal{N}(\varepsilon) \quad . \tag{7.170}$$

In the vicinity of a mobility edge, a scaling hypothesis suggests

$$\mathrm{Th}(\varepsilon, L) = f\big(L/\xi(\varepsilon)\big) \quad , \tag{7.171}$$

where $f(x)$ is a *universal scaling function*.

As the Fermi level passes through the mobility edge into a region of localized states, the conductivity vanishes as

$$\sigma(\varepsilon_{\mathsf{F}}) \sim |\varepsilon_{\mathsf{F}} - \varepsilon_{\mathrm{c}}|^s \quad , \tag{7.172}$$

where $s > 0$. Since the density is a continuous function of ε_{F}, this can also be turned into a statement about the behavior of $\sigma(n)$:

$$\sigma(n) \sim (n - n_{\mathrm{c}})^s \, \Theta(n - n_{\mathrm{c}}) \quad . \tag{7.173}$$

7.5.2 Numerical studies of the localization transition

Pioneering work in numerical studies of the localization transition was performed by MacKinnon and Kramer in the early 1980's. They computed the localization length $\xi_M(W/t, E)$ for systems of dimension $M^{d-1} \times N$, from the formula

$$\xi_M^{-1}\Big(\frac{W}{t}, E\Big) = -\lim_{N\to\infty} \frac{1}{2N} \Big\langle \ln \sum_{i,j=1}^{M^{d-1}} \big|G_{1i,Nj}(E)\big|^2 \Big\rangle \quad , \tag{7.174}$$

where $G(E) = (E + i\epsilon - \mathcal{H})^{-1}$ is the Green's function, and i, j label transverse sites. The average $\langle \cdots \rangle$ is over disorder configurations. It is computationally very convenient to compute the localization length in this manner, rather than from exact diagonalization, because the Green's function can be computed recursively. For details, see A. MacKinnon and B. Kramer, *Phys. Rev. Lett.* **47**, 1546 (1981). Note also that it is $\langle |G|^2 \rangle$ which is computed, rather than $\langle G \rangle$. The reason for this is that the Green's function itself carries a complex phase which when averaged over disorder configurations results in a decay of $\langle G_{\boldsymbol{R},\boldsymbol{R}'} \rangle$ on the scale of the elastic mean free path ℓ.

MacKinnon and Kramer computed the localization length by employing the *Ansatz* of *finite size scaling*. They assumed that

$$\xi_M\Big(\frac{W}{t}, E\Big) = M f\bigg(\xi_\infty\Big(\frac{W}{t}, E\Big)\Big/M\bigg) \quad , \tag{7.175}$$

where $f(x)$ is a universal scaling function which depends only on the dimension d. MK examined the band center, at $E = 0$; for $d > 2$ this is the last region to localize as W/t is increased from zero. A mock-up of typical raw data is shown in fig. 7.10.

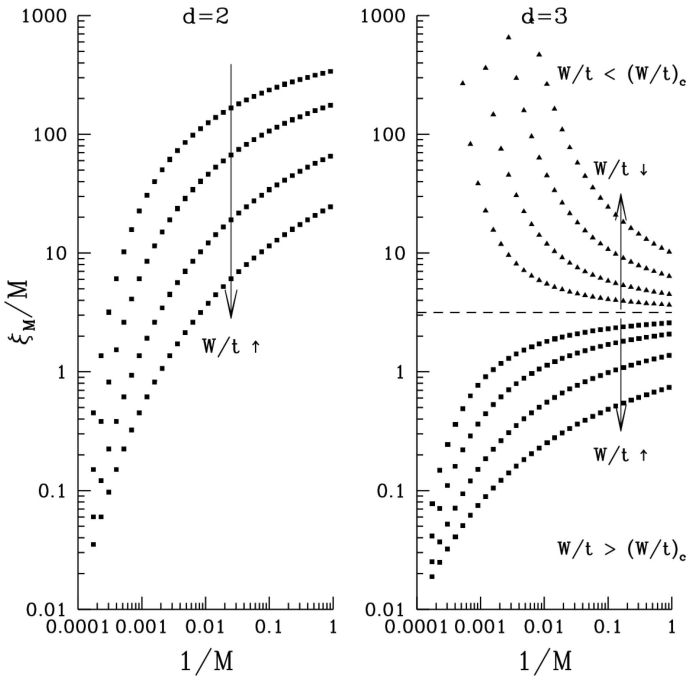


Figure 7.10: Mock-up of typical raw data from numerical study of localization length for $d = 2$ (left panel) and $d = 3$ (right panel) systems. For $d = 2$, $\xi_\infty(W/t)$ is finite and monotonically decreasing with increasing W/t. For $d = 3$, there is a critical value of W/t. For $W/t < (W/t)_c$, $\xi_M(W/t)$ diverges as $M \to \infty$; this is the extended phase. For $W/t > (W/t)_c$, $\xi_M(W/t)$ remains finite as $M \to \infty$; this is the localized phase.

In the $d = 2$ case, all states are localized. Accordingly, $\xi_M/M \to 0$ as $M \to \infty$, and $\xi_M(W/t)$ decreases with increasing W/t. In the $d = 3$ case, states at the band center are extended for weak disorder. As W/t increases, $\xi_M(W/t)$ decreases, but with $\xi_\infty(W/t)$ still divergent. At the critical point, $(W/t)_c$, this behavior changes. The band center states localize, and $\xi_\infty(W/t)$ is finite for $W/t > (W/t)_c$. If one rescales and plots ξ_M/M *versus* ξ_∞/M, the scaling function $f(x)$ is revealed. This is shown in fig. 7.11, which is from the paper by MacKinnon and Kramer. Note that there is only one branch to the scaling function for $d = 2$, but two branches for $d = 3$. MacKinnon and Kramer found $(W/t)_c \simeq 16.5$ for a disordered tight binding model on a simple cubic lattice.

7.5.3 Scaling theory of localization

In the metallic limit, the dimensionless conductance of a L^d hypercube is given by the Ohmic result

$$g(L) = \frac{h\sigma}{e^2}\, L^{d-2} \quad , \tag{7.176}$$

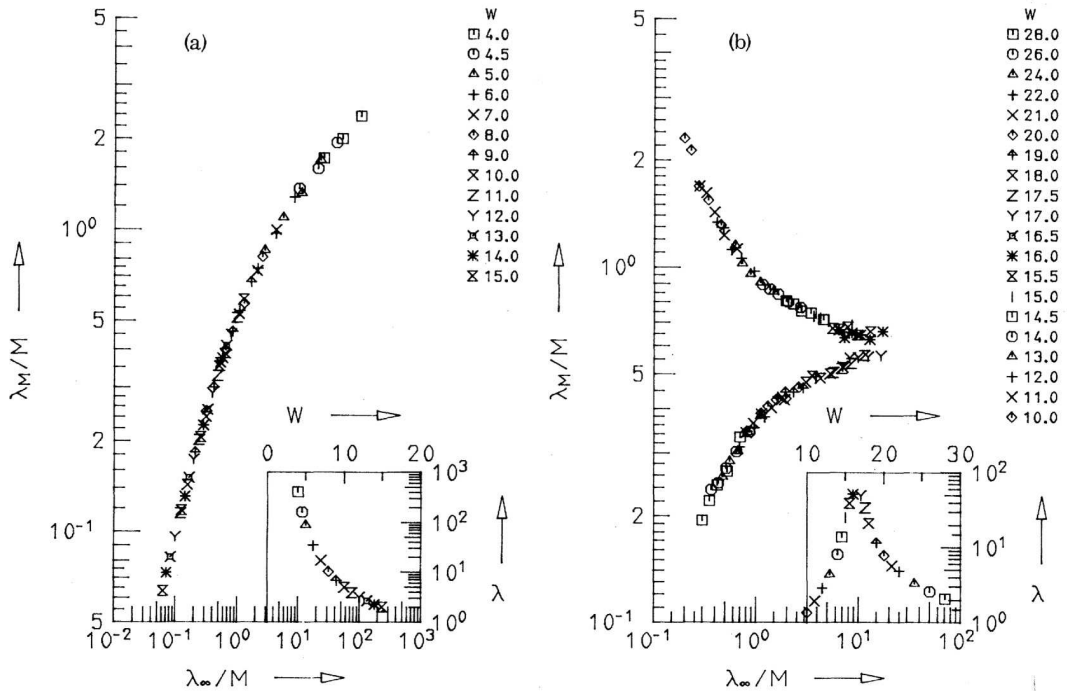


Figure 7.11: Scaling function λ_M/M *versus* λ_∞/M for the localization length λ_M of a system of thickness M for (a) $d = 2$, and (b) $d = 3$. Insets show the scaling parameter λ_∞ as a function of the disorder W. From A. MacKinnon and B. Kramer, *Phys. Rev. Lett.* **47**, 1546 (1981).

whereas in the localized limit we have, from Pichard's formula,

$$g(L) = 4\, e^{-2L/\xi} \quad . \tag{7.177}$$

It is instructive to consider the function,

$$\beta(g) \equiv \frac{d \ln g}{d \ln L} \quad , \tag{7.178}$$

which describes the change of g when we vary the size of the system. We now know the limiting values of $\beta(g)$ for small and large g:

$$\begin{aligned} \text{metallic } (g \gg 1) \quad &\Longrightarrow \quad \beta(g) = d - 2 \\ \text{localized } (g \ll 1) \quad &\Longrightarrow \quad \beta(g) = -\frac{2L}{\xi} = \ln g + \text{const.} \end{aligned} \tag{7.179}$$

If we assume that $\beta(g)$ is a smooth monotonic function of g, we arrive at the picture in fig. 7.12. Note that in $d = 1$, we can compute $\beta(g)$ exactly, using the Landauer formula,

$$g = \frac{T}{1 - T} \quad , \tag{7.180}$$

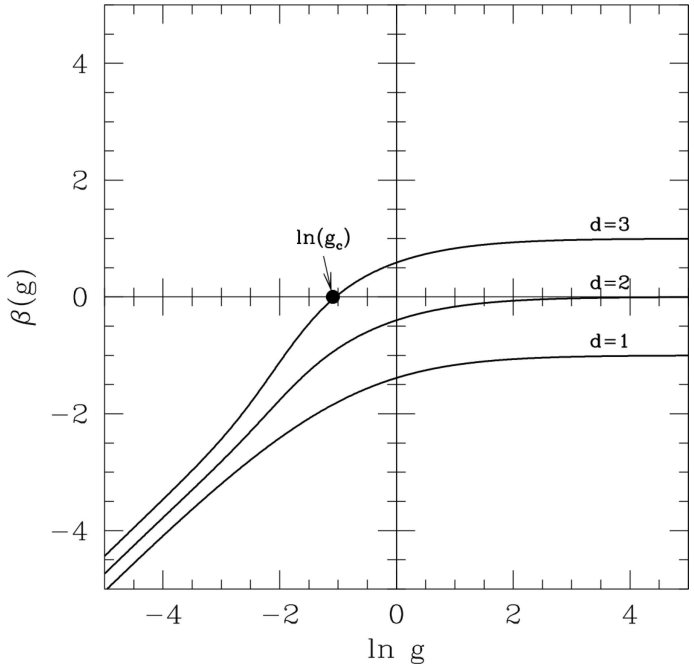


Figure 7.12: Sketch of the β-function for the localization problem for $d = 1, 2, 3$. A critical point exists at $g = g_c$ for the $d = 3$ case.

where $T \propto \exp(-L/\xi)$. From this, we obtain

$$\beta_{d=1}(g) = -(1+g)\,\ln(1+g^{-1}) \quad . \tag{7.181}$$

It should be stressed that the very existence of a β-function is hardly clear. If it does exist, it says that the conductance of a system of size L is uniquely determined by its conductance at some other length scale, typically chosen to be microscopic, *e.g.* $L_0 = \ell$. Integrating the β-function, we obtain an integral equation to be solved implicitly for $g(L)$:

$$\ln\left(\frac{L}{L_0}\right) = \int\limits_{\ln g(L_0)}^{\ln g(L)} \frac{d\ln g}{\beta(g)} \quad . \tag{7.182}$$

A priori it seems more likely, though, that as L is increased the changes to the conductance may depend on more than g alone. *E.g.* the differential change dg might depend on the entire distribution function $P(W)$ for the disorder.

Integrating the β-function: $d = 3$

We know $\beta(g \to 0) \simeq \ln g$ and $\beta(g \to \infty) = 1$, hence by the intermediate value theorem there is at least one point were $\beta(g)$ vanishes. Whenever g satisfies $\beta(g) = 0$, the conductance g is *scale*

invariant – it does not change with increasing (or decreasing) system size L. We will assume the situation is reflected by the sketch of fig. 7.12, and that there is one such point, $g_{\rm c}$.

We now apply (7.182). Not knowing the precise form of $\beta(g)$, we approximate it piecewise:

$$\beta(g) \simeq \begin{cases} 1 & \text{if } g \geqslant g_+ \\ \alpha\,\ln(g/g_{\rm c}) & \text{if } g_- < g < g_+ \\ \ln g & \text{if } g < g_- \quad , \end{cases} \tag{7.183}$$

where $\alpha = g_{\rm c}\,\beta'(g_{\rm c})$. We determine g_+ and g_- by continuity:

$$\begin{aligned} \ln g_+ &= \ln g_{\rm c} + \frac{1}{\alpha} \\ \ln g_- &= \frac{\alpha}{\alpha - 1}\,\ln g_{\rm c} \quad . \end{aligned} \tag{7.184}$$

Now suppose we start with $g_0 = g_{\rm c} + \delta g$, where $|\delta g| \ll 1$. We integrate out to $g = g_+$ and then from g_+ to $g \gg 1$:

$$\begin{aligned} \ln\left(\frac{L_+}{L_0}\right) &= \int\limits_{\ln g_0}^{\ln g_+} \frac{d\ln g}{\alpha\,\ln(g/g_{\rm c})} = \frac{1}{\alpha}\ln\left(\frac{\ln(g_+/g_{\rm c})}{\ln(g_0/g_{\rm c})}\right) \\ \ln\left(\frac{L}{L_+}\right) &= \int\limits_{\ln g_+}^{\ln g} d\ln g = \ln(g/g_+) \quad , \end{aligned} \tag{7.185}$$

which together imply

$$g(L) = A_+\,g_{\rm c}\cdot\frac{L}{L_0}\cdot(g_0 - g_{\rm c})^{1/\alpha} \quad , \tag{7.186}$$

where $A_+ = (e\alpha/g_{\rm c})^{1/\alpha}$. The conductivity is

$$\sigma = \frac{e^2}{h}\cdot\frac{g}{L} = \frac{e^2}{h}\cdot\frac{A_+\,g_{\rm c}}{L_0}\cdot(g_0 - g_{\rm c})^{1/\alpha} \quad . \tag{7.187}$$

If instead we start with $g_0 = g_{\rm c} - \delta g$ and integrate out to large negative values of $\ln g$, then

$$\begin{aligned} \ln\left(\frac{L_-}{L_0}\right) &= \int\limits_{\ln g_0}^{\ln g_-} \frac{d\ln g}{\alpha\,\ln(g/g_{\rm c})} = \frac{1}{\alpha}\ln\left(\frac{\ln(g_{\rm c}/g_-)}{\ln(g_{\rm c}/g_0)}\right) \\ \ln\left(\frac{L}{L_-}\right) &= \int\limits_{\ln g_-}^{\ln g} \frac{d\ln g}{\ln g} = \ln\left(\frac{\ln g}{\ln g_-}\right) \quad , \end{aligned} \tag{7.188}$$

which says

$$g(L) = e^{-2L/\xi}$$
$$\xi = \frac{2L_0}{A_-} \cdot (g_{\rm c} - g_0)^{-1/\alpha} \quad , \tag{7.189}$$

with

$$A_- = \frac{\ln(1/g_-)}{g_{\rm c} \cdot \left[\ln(g_{\rm c}/g_-)\right]^{1/\alpha}} \quad . \tag{7.190}$$

On the metallic side of the transition, when $g_0 > g_{\rm c}$, we can identify a localization length through

$$g \equiv g_{\rm c}/\xi \quad , \tag{7.191}$$

which says

$$\xi = \frac{L_0}{A_+} (g_0 - g_{\rm c})^{-1/\alpha} \quad . \tag{7.192}$$

Finally, since g_0 is determined by the value of the Fermi energy $\varepsilon_{\rm F}$. we can define the *critical energy*, or *mobility edge* $\varepsilon_{\rm c}$, through

$$g(L_0, \varepsilon_{\rm c}) = g_{\rm c} \quad , \tag{7.193}$$

in which case

$$\delta g \equiv g(L_0, \varepsilon_{\rm F}) - g_{\rm c} = \left.\frac{\partial g(L_0, \varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=\varepsilon_{\rm c}} \cdot \left(\varepsilon_{\rm F} - \varepsilon_{\rm c}\right) \quad . \tag{7.194}$$

Thus, $\delta g \propto \delta\varepsilon \equiv (\varepsilon_{\rm F} - \varepsilon_{\rm c})$.

Integrating the β-function: $d = 2$

In two dimensions, there is no fixed point. In the Ohmic limit $g \gg 1$, we have

$$\beta(g) = -\frac{c}{g} + \mathcal{O}(g^{-2}) \quad , \tag{7.195}$$

where c is a constant. Thus,

$$\ln\left(\frac{L}{L_0}\right) = \int\limits_{\ln g_0}^{\ln g} \frac{d\ln g}{\beta(g)} = -\frac{g - g_0}{c} + \ldots \tag{7.196}$$

and

$$g(L) = k_{\rm F}\ell - c\,\ln(L/\ell) \quad , \tag{7.197}$$

where we have used the Drude result $g = k_{\rm F}\ell$, valid for $L_0 = \ell$. We now see that the localization length ξ is the value of L for which the correction term is on the same order as g_0: $\xi = \ell \exp(k_{\rm F}\ell/c)$. A first principles treatment yields $c = \frac{2}{\pi}$. The metallic regime in $d = 2$ is often called the *weak localization* regime.

$2+\epsilon$ dimensions

At or below $d=2$ dimensions, there is no mobility edge and all eigenstates are localized. $d=2$ *is the lower critical dimension for the localization transition.* Consider now the problem in $d=2+\epsilon$ dimensions. One has

$$\beta(g) = \epsilon - \frac{c}{g} + \mathcal{O}(g^{-2}) \quad . \tag{7.198}$$

The critical conductance lies at $g_{\rm c} = c/\epsilon$. For $\epsilon \to 0^+$, this is large enough that higher order terms in the expansion of the β-function can safely be ignored in the metallic limit. An analysis similar to that for $d=3$ now yields

$$\begin{aligned} g > g_{\rm c} \quad &\Longrightarrow \quad g(L) = \frac{h}{e^2}\,\sigma\,L^{\epsilon} \\ g < g_{\rm c} \quad &\Longrightarrow \quad g(L) = e^{-2L/\xi} \quad , \end{aligned} \tag{7.199}$$

with

$$\begin{aligned} \sigma &= \frac{e^2}{h}\,L_0^{\epsilon}\cdot(g_0 - g_{\rm c}) \\ \xi &= \frac{2L_0}{A_-}\cdot(g_{\rm c} - g_0) \quad . \end{aligned} \tag{7.200}$$

Note that $\alpha = g_{\rm c}\,\beta'(g_{\rm c}) = +c/g_{\rm c} = \epsilon$. We thus obtain

$$\xi(\varepsilon) \propto |\varepsilon - \varepsilon_{\rm c}|^{-\nu} \tag{7.201}$$

with $\nu = 1 + \mathcal{O}(\epsilon)$. Close to the transition on the metallic side, the conductivity vanishes as

$$\sigma(\varepsilon) \propto |\varepsilon - \varepsilon_{\rm c}|^{s} \quad . \tag{7.202}$$

The relation $s = (d-2)\nu$, which follows from the above treatment, may be used to relate the localization length and conductivity critical exponents. (In $d=3$, MacKinnon and Kramer obtained $\nu = s \simeq 1.2$.)

7.5.4 Finite temperature

In the metallic regime, one obtains from the scaling theory,

$$\begin{aligned} \sigma_{d=3}(L) &= \frac{e^2}{h}\cdot\left\{\frac{2k_{\rm F}^2\ell}{3\pi} - \frac{2}{\pi^2}\left(\frac{1}{\ell} - \frac{1}{L}\right)\right\} \\ \sigma_{d=2}(L) &= \frac{e^2}{h}\cdot\left\{k_{\rm F}\ell - \frac{2}{\pi}\ln\left(\frac{L}{\ell}\right)\right\} \\ \sigma_{d=1}(L) &= \frac{e^2}{h}\cdot\Big\{4\ell - 2(L-\ell)\Big\} \quad . \end{aligned} \tag{7.203}$$

Clearly the $d = 1$ result must break down for even microscopic $L \gtrsim 3\ell$. The above results are computed using the β-function

$$\beta(g) = (d-2) - \frac{c_d}{g} + \mathcal{O}(g^{-2}) \quad , \tag{7.204}$$

where the coefficients c_d are computed from perturbation theory in the disorder.

At finite temperature, the cutoff becomes $\min(L, L_\phi)$, where $L_\phi = \sqrt{D\tau_\phi}$ is the inelastic scattering length and $D = v_{\rm F}\ell$ is the diffusion constant. Suppose that $\tau_\phi(T) \propto T^{-p}$ as $T \to 0$, so that $L_\phi = a\,(T/T_0)^{-p/2}$, where T_0 is some characteristic temperature (*e.g.* the Debye temperature, if the inelastic mechanism is electron-phonon scattering). Then, for $L_\phi > L$,

$$\begin{aligned}
\sigma_{d=3}(T) &= \sigma^{\rm B}_{d=3} - \frac{2}{\pi^2}\,\frac{e^2}{h}\left\{\frac{1}{\ell} - \frac{1}{a}\left(\frac{T}{T_0}\right)^{p/2}\right\} \\
\sigma_{d=2}(T) &= \sigma^{\rm B}_{d=2} - \frac{2}{\pi}\,\frac{e^2}{h}\left\{\ln\left(\frac{a}{\ell}\right) - \frac{p}{2}\ln\left(\frac{T}{T_0}\right)\right\} \\
\sigma_{d=1}(T) &= \sigma^{\rm B}_{d=1} - 2\,\frac{e^2}{h}\left\{a\left(\frac{T_0}{T}\right)^{p/2} - \ell\right\} \quad ,
\end{aligned} \tag{7.205}$$

where $\sigma^{\rm B}_d$ is the Boltzmann conductivity. Note that $\sigma(T)$ decreases with decreasing temperature, unlike the classic low T result for metals, where $\rho(T) = \rho_0 + AT^2$. *I.e.* usually $\rho(T)$ increases as T increases due to a concomitant decrease in transport scattering time τ. Weak localization physics, though, has the opposite effect, as the enhanced backscattering is suppressed as T increases and L_ϕ decreases. The result is that $\rho(T)$ starts to decrease as T is lowered from high temperatures, but turns around at low T and starts increasing again. This behavior was first observed in 1979 by Dolan and Osheroff, who studied thin metallic PdAu films, observing a logarithmic increase in $\rho_{d=2}(T)$ at the lowest temperatures.

Chapter 8

Hartree-Fock and Density Functional Theories

8.1 Second Quantization

8.1.1 Basis states and creation/annihilation operators

Second quantization is a convenient scheme to label basis states of a many particle quantum system. We are ultimately interested in solutions of the many-body Schrödinger equation,

$$\hat{H}\Psi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = E\,\Psi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) \tag{8.1}$$

where the Hamiltonian is

$$\begin{aligned}\hat{H} &= \sum_{i=1}^{N}\left(-\frac{\hbar^2}{2m}\boldsymbol{\nabla}_i^2 + v_{\text{ext}}(\boldsymbol{x}_i)\right) + \sum_{j<k}^{N} u(\boldsymbol{x}_j-\boldsymbol{x}_k) \\ &\equiv \hat{T}+\hat{V}+\hat{U} \quad ,\end{aligned} \tag{8.2}$$

where $\hat{T}$ is the kinetic energy, $\hat{V}$ the one-body potential energy, and $\hat{U}$ the two-body potential energy. To the coordinate labels $\{\boldsymbol{x}_1,\ldots\boldsymbol{x}_N\}$ we may also append labels for internal degrees of freedom, such as spin polarization, denoted $\{\sigma_1,\ldots,\sigma_N\}$. Since $\left[\hat{H},\pi\right]=0$ for all permutations $\pi\in\mathcal{S}_N$, the many-body wavefunctions may be chosen to transform according to irreducible representations of the symmetric group $\mathcal{S}_N$. Thus, for any $\pi\in\mathcal{S}_N$,

$$\Psi\big(\boldsymbol{x}_{\pi(1)},\ldots,\boldsymbol{x}_{\pi(N)}\big) = \left\{\begin{matrix}1\\ \mathsf{sgn}(\pi)\end{matrix}\right\}\Psi(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) \quad , \tag{8.3}$$

where the upper choice is for Bose-Einstein statistics and the lower sign for Fermi-Dirac statistics. Here $\boldsymbol{x}_j$ may include not only the spatial coordinates of particle j, but its internal quantum number(s) as well, such as the spin polarization σ_j.

A convenient basis for the many body states is obtained from the single-particle eigenstates $\{|\,\alpha\,\rangle\}$ of some single-particle Hamiltonian $\hat{H}_0$, with $\langle\,\boldsymbol{x}\,|\,\alpha\,\rangle = \varphi_\alpha(\boldsymbol{x})$ and $\hat{H}_0\,|\,\alpha\,\rangle = \varepsilon_\alpha\,|\,\alpha\,\rangle$. The basis may be taken as orthonormal, *i.e.* $\langle\,\alpha\,|\,\alpha'\,\rangle = \delta_{\alpha\alpha'}$. Now define

$$\Psi_{\{\alpha_1,\ldots,\alpha_N\}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = \frac{1}{\sqrt{N!\prod_\alpha n_\alpha!}}\sum_{\pi\in\mathcal{S}_N}\left\{\begin{matrix}1\\ \mathsf{sgn}(\pi)\end{matrix}\right\}\varphi_{\alpha_{\pi(1)}}(\boldsymbol{x}_1)\cdots\varphi_{\alpha_{\pi(N)}}(\boldsymbol{x}_N) \quad . \tag{8.4}$$

These states form a basis for the N-particle Hilbert space. Here n_α is the number of times the index α appears among the set $\{\alpha_1,\ldots,\alpha_N\}$. For BE statistics, $n_\alpha \in \{0,1,2,\ldots\}$, whereas for FD statistics, $n_\alpha \in \{0,1\}$. Note that the above states are normalized[1]:

$$\begin{aligned}\int d^dx_1\cdots\int d^dx_N\,\left|\Psi_{\{\alpha_1,\ldots,\alpha_N\}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)\right|^2 &= \frac{1}{N!\prod_\alpha n_\alpha!}\sum_{\pi,\mu\in\mathcal{S}_N}\left\{\begin{matrix}1\\ \mathsf{sgn}(\pi\mu)\end{matrix}\right\}\prod_{j=1}^N\int d^dx_j\;\varphi^*_{\alpha_{\pi(j)}}(\boldsymbol{x}_j)\,\varphi_{\alpha_{\mu(j)}}(\boldsymbol{x}_j)\\ &= \frac{1}{\prod_\alpha n_\alpha!}\sum_{\pi\in\mathcal{S}_N}\prod_{j=1}^N\delta_{\alpha_j,\alpha_{\pi(j)}} = 1 \quad .\end{aligned} \tag{8.5}$$

Note that

$$\begin{aligned}\sum_{\pi\in\mathcal{S}_N}\varphi_{\alpha_{\pi(1)}}(\boldsymbol{x}_1)\cdots\varphi_{\alpha_{\pi(N)}}(\boldsymbol{x}_N) &\equiv \mathsf{per}\big\{\varphi_{\alpha_i}(\boldsymbol{x}_j)\big\}\\ \sum_{\pi\in\mathcal{S}_N}\mathsf{sgn}(\pi)\,\varphi_{\alpha_{\pi(1)}}(\boldsymbol{x}_1)\cdots\varphi_{\alpha_{\pi(N)}}(\boldsymbol{x}_N) &\equiv \det\big\{\varphi_{\alpha_i}(\boldsymbol{x}_j)\big\} \quad ,\end{aligned} \tag{8.6}$$

which stand for *permanent* and *determinant*, respectively. We may now write

$$\Psi_{\{\alpha_1,\ldots,\alpha_N\}}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = \big\langle\,\boldsymbol{x}_1\,,\,\ldots\,,\,\boldsymbol{x}_N\,\big|\,\alpha_1\cdots\alpha_N\,\big\rangle \quad , \tag{8.7}$$

where

$$|\,\alpha_1\cdots\alpha_N\,\rangle = \frac{1}{\sqrt{N!\prod_\alpha n_\alpha!}}\sum_{\pi\in\mathcal{S}_N}\left\{\begin{matrix}1\\ \mathsf{sgn}(\pi)\end{matrix}\right\}|\,\alpha_{\pi(1)}\,\rangle\otimes|\,\alpha_{\pi(2)}\,\rangle\otimes\cdots\otimes|\,\alpha_{\pi(N)}\,\rangle \quad . \tag{8.8}$$

Note that $|\,\alpha_{\pi(1)}\cdots\alpha_{\pi(N)}\,\rangle = (\pm1)^\pi\,|\,\alpha_1\cdots\alpha_N\,\rangle$, where by $(\pm1)^\pi$ we mean 1 in the case of BE statistics and $\mathsf{sgn}(\pi)$ in the case of FD statistics.

We may express $|\,\alpha_1\cdots\alpha_N\,\rangle$ as a product of creation operators acting on a vacuum $|\,0\,\rangle$ in *Fock space*. For bosons,

$$|\,\alpha_1\cdots\alpha_N\,\rangle = \prod_\alpha\frac{(b^\dagger_\alpha)^{n_\alpha}}{\sqrt{n_\alpha!}}\,|\,0\,\rangle \equiv |\,\{n_\alpha\}\,\rangle \quad , \tag{8.9}$$

[1] In the normalization integrals, each $\int d^dx$ implicitly includes a sum $\sum_\zeta$ over any internal indices that may be present.

where n_α is the multiplicity of the label α in the set $\{\alpha_1, \ldots, \alpha_N\}$. Consequently, $N = \sum_\alpha n_\alpha$. The Hermitian conjugate of the creation operator $b_\alpha^\dagger$ is the annihilation operator b_α. The relations among these operators are

$$\big[b_\alpha\,,\,b_\beta\big] = 0 \qquad , \qquad \big[b_\alpha^\dagger\,,\,b_\beta^\dagger\big] = 0 \qquad , \qquad \big[b_\alpha\,,\,b_\beta^\dagger\big] = \delta_{\alpha\beta} \quad , \tag{8.10}$$

where $[\,\bullet\,,\bullet\,]$ is the commutator.

For fermions,

$$|\,\alpha_1 \cdots \alpha_N\,\rangle = \prod_\alpha (c_\alpha^\dagger)^{n_\alpha}\,|\,0\,\rangle = c_{\alpha_1}^\dagger c_{\alpha_2}^\dagger \cdots c_{\alpha_N}^\dagger\,|\,0\,\rangle \equiv |\,\{n_\alpha\}\,\rangle \quad , \tag{8.11}$$

where each $n_\alpha \in \{0, 1\}$ and where $\sum_\alpha n_\alpha = N$. Thus $n_\alpha = 1$ for each $\alpha \in \{\alpha_1, \ldots, \alpha_N\}$ and $n_\alpha = 0$ otherwise. We also implicitly assume a canonical, though arbitrary, ordering of the single particle labels α. The fermionic creation and annihilation operators satisfy the relations

$$\big\{c_\alpha\,,\,c_\beta\big\} = 0 \qquad , \qquad \big\{c_\alpha^\dagger\,,\,c_\beta^\dagger\big\} = 0 \qquad , \qquad \big\{c_\alpha\,,\,c_\beta^\dagger\big\} = \delta_{\alpha\beta} \quad , \tag{8.12}$$

where $\{\bullet\,,\bullet\}$ is the anticommutator. Because the fermion creation operators all anticommute, we have

$$c_{\alpha_{\pi(1)}}^\dagger c_{\alpha_{\pi(2)}}^\dagger \cdots c_{\alpha_{\pi(N)}}^\dagger\,|\,0\,\rangle = \mathsf{sgn}(\pi)\,|\,\{n_\alpha\}\,\rangle \quad , \tag{8.13}$$

for any $\pi \in \mathcal{S}_N$.

We may also define the operators

$$b(\boldsymbol{x}) = \sum_\alpha \varphi_\alpha(\boldsymbol{x})\,b_\alpha \qquad , \qquad c(\boldsymbol{x}) = \sum_\alpha \varphi_\alpha(\boldsymbol{x})\,c_\alpha \quad , \tag{8.14}$$

which satisfy

$$\big[b(\boldsymbol{x})\,,\,b^\dagger(\boldsymbol{x}')\big] = \delta(\boldsymbol{x} - \boldsymbol{x}') \qquad , \qquad \big\{c(\boldsymbol{x})\,,\,c^\dagger(\boldsymbol{x}')\big\} = \delta(\boldsymbol{x} - \boldsymbol{x}') \quad . \tag{8.15}$$

In cases where there are internal (*e.g.*, spin) degrees of freedom, the above relations become

$$\big[b_m(\boldsymbol{x})\,,\,b_{m'}^\dagger(\boldsymbol{x}')\big] = \delta(\boldsymbol{x} - \boldsymbol{x}')\,\delta_{mm'} \qquad , \qquad \big\{c_m(\boldsymbol{x})\,,\,c_{m'}^\dagger(\boldsymbol{x}')\big\} = \delta(\boldsymbol{x} - \boldsymbol{x}')\,\delta_{mm'} \quad . \tag{8.16}$$

In other words,

$$b_{\alpha,m} = \int d^dx\;\varphi_\alpha^*(\boldsymbol{x}, m)\,b_m(\boldsymbol{x}) \qquad , \qquad c_{\alpha,m} = \int d^dx\;\varphi_\alpha^*(\boldsymbol{x}, m)\,c_m(\boldsymbol{x}) \quad . \tag{8.17}$$

Note the difference between the many-body states

$$\begin{aligned} |\,\boldsymbol{x}_1 \cdots \boldsymbol{x}_N\,\rangle &\equiv \psi^\dagger(\boldsymbol{x}_1) \cdots \psi^\dagger(\boldsymbol{x}_N)\,|\,0\,\rangle \\ &= \frac{1}{\sqrt{N!}} \sum_{\pi \in \mathcal{S}_N} \left\{ \begin{matrix} 1 \\ \mathsf{sgn}(\pi) \end{matrix} \right\} |\,\boldsymbol{x}_{\pi(1)}\,,\,\ldots\,,\,\boldsymbol{x}_{\pi(N)}\,\rangle \end{aligned} \tag{8.18}$$

and

$$|\,\boldsymbol{x}_1\,,\,\ldots\,,\,\boldsymbol{x}_N\,\rangle = |\,\boldsymbol{x}_1\,\rangle \otimes \cdots \otimes |\,\boldsymbol{x}_N\,\rangle \quad . \tag{8.19}$$

In particular,

$$\big\langle\, \boldsymbol{x}_1\,,\,\ldots\,,\,\boldsymbol{x}_N \,\big|\, \alpha_1 \cdots \alpha_N \,\big\rangle = \Psi_{\{\alpha_1\,,\,\ldots\,,\,\alpha_N\}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \tag{8.20}$$

but

$$\big\langle\, \boldsymbol{x}_1 \cdots \boldsymbol{x}_N \,\big|\, \alpha_1 \cdots \alpha_N \,\big\rangle = \sqrt{N!}\;\Psi_{\{\alpha_1\,,\,\ldots\,,\,\alpha_N\}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \quad . \tag{8.21}$$

A general normalized N-body wavefunction $\Psi_{m_1,\ldots,m_N}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ may be expressed in second quantized notation as

$$|\,\Psi\,\rangle = \frac{1}{\sqrt{N!}} \sum_{m_1} \cdots \sum_{m_N} \int d^d x_1 \cdots \int d^d x_N\; \Psi_{m_1,\ldots,m_N}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\; \psi^\dagger_{m_1}(\boldsymbol{x}_1) \cdots \psi^\dagger_{m_N}(\boldsymbol{x}_N)\,|\,0\,\rangle \tag{8.22}$$

where $\psi^\dagger_m(\boldsymbol{x})$ is the bosonic or fermionic creation operator for particles of internal index m at position $\boldsymbol{x}$. Dropping for the moment the internal indices for the sake of simplicity, note that

$$\begin{aligned}\langle\, 0\,|\,\psi(\boldsymbol{y}_2)\,\psi(\boldsymbol{y}_1)\,\psi^\dagger(\boldsymbol{x}_1)\,\psi^\dagger(\boldsymbol{x}_2)\,|\,0\,\rangle &= \langle\, 0\,|\,\psi(\boldsymbol{y}_2)\,\psi^\dagger(\boldsymbol{x}_2)\,|\,0\,\rangle\,\delta(\boldsymbol{x}_1 - \boldsymbol{y}_1) \\ &\qquad \pm \langle\, 0\,|\,\psi(\boldsymbol{y}_2)\,\psi(\boldsymbol{x}_1)\,\psi^\dagger(\boldsymbol{y}_1)\,\psi^\dagger(\boldsymbol{x}_2)\,|\,0\,\rangle \\ &= \delta(\boldsymbol{x}_1 - \boldsymbol{y}_1)\,\delta(\boldsymbol{x}_2 - \boldsymbol{y}_2) \pm \delta(\boldsymbol{x}_1 - \boldsymbol{y}_2)\,\delta(\boldsymbol{x}_2 - \boldsymbol{y}_1) \quad .\end{aligned} \tag{8.23}$$

Reasoning thusly, we conclude that

$$\langle\, 0\,|\,\psi(\boldsymbol{y}_N)\cdots\psi(\boldsymbol{y}_1)\,\psi^\dagger(\boldsymbol{x}_1)\cdots\psi^\dagger(\boldsymbol{x}_N)\,|\,0\,\rangle = \sum_{\sigma\in\mathcal{S}_N} \left\{ \begin{matrix} 1 \\ \mathsf{sgn}(\pi) \end{matrix} \right\} \prod_{j=1}^N \delta(\boldsymbol{y}_j - \boldsymbol{x}_{\sigma(j)}) \quad . \tag{8.24}$$

Including the internal indices, then,

$$\langle\, 0\,|\,\psi(\boldsymbol{y}_N)_{n_N}\cdots\psi_{n_1}(\boldsymbol{y}_1)\,\psi^\dagger_{m_1}(\boldsymbol{x}_1)\cdots\psi^\dagger_{m_N}(\boldsymbol{x}_N)\,|\,0\,\rangle = \sum_{\sigma\in\mathcal{S}_N} \left\{ \begin{matrix} 1 \\ \mathsf{sgn}(\pi) \end{matrix} \right\} \prod_{j=1}^N \delta(\boldsymbol{y}_j - \boldsymbol{x}_{\sigma(j)})\,\delta_{n_j\,,\,m_{\sigma(j)}} \quad . \tag{8.25}$$

We then have

$$\langle\, \Phi\,|\,\Psi\,\rangle = \sum_{m_1} \cdots \sum_{m_N} \int d^d x_1 \cdots \int d^d x_N\; \Phi^*_{m_1,\ldots,m_N}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\; \Psi_{m_1,\ldots,m_N}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \tag{8.26}$$

Another useful thing to derive are expressions for the one- and two-body density matrices. Note that (with internal indices once again suppressed)

$$\begin{aligned}\psi(\boldsymbol{x})\,\psi^\dagger(\boldsymbol{x}_1)\cdots\psi^\dagger(\boldsymbol{x}_N) &= \delta(\boldsymbol{x} - \boldsymbol{x}_1)\,\psi^\dagger(\boldsymbol{x}_2)\cdots\psi^\dagger(\boldsymbol{x}_N) \pm \delta(\boldsymbol{x} - \boldsymbol{x}_2)\,\psi^\dagger(\boldsymbol{x}_1)\,\psi^\dagger(\boldsymbol{x}_3)\cdots\psi^\dagger(\boldsymbol{x}_N) \\ &\qquad + \ldots + (\pm 1)^{N-1}\,\delta(\boldsymbol{x} - \boldsymbol{x}_N)\,\psi^\dagger(\boldsymbol{x}_1)\cdots\psi^\dagger(\boldsymbol{x}_{N-1}) \quad ,\end{aligned} \tag{8.27}$$

from which we may derive the one-body density matrix,

$$\begin{aligned} n_1(\boldsymbol{x} \,|\, \boldsymbol{y}) &\equiv \langle\, \Psi \,|\, \psi^\dagger(\boldsymbol{y})\, \psi(\boldsymbol{x}) \,|\, \Psi \,\rangle \\ &= N \int d^d x_2 \cdots \int d^d x_N \; \Psi^*(\boldsymbol{y}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)\, \Psi(\boldsymbol{x}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \quad . \end{aligned} \tag{8.28}$$

Similarly, the two-body density matrix is given by

$$\begin{aligned} n_2(\boldsymbol{x}, \boldsymbol{x}' \,|\, \boldsymbol{y}, \boldsymbol{y}') &\equiv \langle\, \Psi \,|\, \psi^\dagger(\boldsymbol{y}')\, \psi^\dagger(\boldsymbol{y})\, \psi(\boldsymbol{x})\, \psi(\boldsymbol{x}') \,|\, \Psi \,\rangle \\ &= N(N-1) \int d^d x_3 \cdots \int d^d x_N \; \Psi^*(\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{x}_3, \ldots, \boldsymbol{x}_N)\, \Psi(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{x}_3, \ldots, \boldsymbol{x}_N) \quad . \end{aligned} \tag{8.29}$$

Note the sum rules upon integrating the diagonal elements, *viz.*

$$\begin{aligned} \int d^d x \; n_1(\boldsymbol{x} \,|\, \boldsymbol{x}) &= N \\ \int d^d x \int d^d x' \; n_2(\boldsymbol{x}, \boldsymbol{x}' \,|\, \boldsymbol{x}, \boldsymbol{x}') &= N(N-1) \quad . \end{aligned} \tag{8.30}$$

In a plane wave basis, we write

$$\psi_{\boldsymbol{k}} = \frac{1}{\sqrt{V}} \int d^d x \; \psi(\boldsymbol{x})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \tag{8.31}$$

and therefore

$$\begin{aligned} \langle\, \Psi \,|\, \psi^\dagger_{\boldsymbol{k}}\, \psi_{\boldsymbol{k}'} \,|\, \Psi \,\rangle &= \frac{1}{V} \int d^d x \int d^d r \; e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}}\, e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}}\, \langle\, \Psi \,|\, \psi^\dagger(\boldsymbol{x})\, \psi(\boldsymbol{x}+\boldsymbol{r}) \,|\, \Psi \,\rangle \\ &= \delta_{\boldsymbol{k},\boldsymbol{k}'} \int d^d r \; e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\, \langle\, \Psi \,|\, \psi^\dagger(\boldsymbol{0})\, \psi(\boldsymbol{r}) \,|\, \Psi \,\rangle \\ &= \delta_{\boldsymbol{k},\boldsymbol{k}'} \int d^d r \; e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \int d^d x_2 \cdots \int d^d x_N \Psi^*(\boldsymbol{0}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)\, \Psi(\boldsymbol{r}, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \quad . \end{aligned} \tag{8.32}$$

8.1.2 The second quantized Hamiltonian

Now consider the action of permutation-symmetric first quantized operators such as the kinetic energy $\hat{T} = -\frac{\hbar^2}{2m} \sum_{i=1}^N \boldsymbol{\nabla}_i^2 = \sum_{i=1}^N \hat{t}_i$ and the potential energy $\hat{U} = \sum_{i<j}^N u(\boldsymbol{x}_i - \boldsymbol{x}_j)$. For a one-

body operator such as $\hat{T}$, we have

$$\langle\, \alpha_1 \cdots \alpha_N \,|\, \hat{T} \,|\, \alpha'_1 \cdots \alpha'_N \,\rangle = \int d^d x_1 \cdots \int d^d x_N \left(\prod_\alpha n_\alpha ! \right)^{-1/2} \left(\prod_\alpha n'_\alpha ! \right)^{-1/2} \tag{8.33}$$

$$\times \sum_{\pi \in \mathcal{S}_N} (\pm 1)^\pi \varphi^*_{\alpha_{\pi(1)}}(\boldsymbol{x}_1) \cdots \varphi^*_{\alpha_{\pi(N)}}(\boldsymbol{x}_N) \sum_{k=1}^N \hat{t}_i \, \varphi_{\alpha'_{\pi(1)}}(\boldsymbol{x}_1) \cdots \varphi_{\alpha'_{\pi(N)}}(\boldsymbol{x}_N)$$

$$= \sum_{\pi \in \mathcal{S}_N} (\pm 1)^\pi \left(\prod_\alpha n_\alpha ! \, n'_\alpha ! \right)^{-1/2} \sum_{i=1}^N \prod_{\substack{j \\ (j \neq i)}} \delta_{\alpha_j, \alpha'_{\pi(j)}} \int d^d x_1 \, \varphi^*_{\alpha_i}(\boldsymbol{x}_1) \, \hat{t}_1 \, \varphi_{\alpha'_{\pi(i)}}(\boldsymbol{x}_1) \quad .$$

One may verify that any permutation-symmetric one-body operator such as $\hat{T}$ is faithfully represented by the second quantized expression,

$$\hat{T} = \sum_{\alpha,\beta} \langle\, \alpha \,|\, \hat{t} \,|\, \beta \,\rangle \, \psi^\dagger_\alpha \, \psi_\beta \quad , \tag{8.34}$$

where $\psi^\dagger_\alpha$ is $b^\dagger_\alpha$ or $c^\dagger_\alpha$ as the application determines, and

$$\langle\, \alpha \,|\, \hat{t} \,|\, \beta \,\rangle = \int d^d x \, \varphi^*_\alpha(\boldsymbol{x}) \, \hat{t}(\boldsymbol{\nabla}) \, \varphi_\beta(\boldsymbol{x}) \equiv \tau_{\alpha\beta} \quad . \tag{8.35}$$

Similarly,

$$\hat{V} = \sum_{\alpha,\beta} \langle\, \alpha \,|\, \hat{v}_{\text{ext}} \,|\, \beta \,\rangle \, \psi^\dagger_\alpha \, \psi_\beta \quad , \tag{8.36}$$

where

$$\langle\, \alpha \,|\, \hat{v}_{\text{ext}} \,|\, \beta \,\rangle = \int d^d x \, \varphi^*_\alpha(\boldsymbol{x}) \, \hat{v}_{\text{ext}}(\boldsymbol{x}) \, \varphi_\beta(\boldsymbol{x}) \equiv v^{\text{ext}}_{\alpha\beta} \quad . \tag{8.37}$$

Two-body operators such as $\hat{U}$ are represented as

$$\hat{U} = \tfrac{1}{2} \sum_{\alpha,\beta,\gamma,\delta} \langle\, \alpha\beta \,|\, \hat{u} \,|\, \gamma\delta \,\rangle \, \psi^\dagger_\alpha \, \psi^\dagger_\beta \, \psi_\delta \, \psi_\gamma \quad , \tag{8.38}$$

where

$$\langle\, \alpha\beta \,|\, \hat{u} \,|\, \gamma\delta \,\rangle = \int d^d x \int d^d x' \, \varphi^*_\alpha(\boldsymbol{x}) \, \varphi^*_\beta(\boldsymbol{x}') \, \hat{u}(\boldsymbol{x}, \boldsymbol{x}') \, \varphi_\delta(\boldsymbol{x}') \, \varphi_\gamma(\boldsymbol{x}) \equiv u_{\alpha\beta\gamma\delta} \quad . \tag{8.39}$$

The general form for an n-body operator is then

$$\hat{R} = \frac{1}{n!} \sum_{\substack{\alpha_1 \cdots \alpha_n \\ \beta_1 \cdots \beta_n}} \langle\, \alpha_1 \cdots \alpha_n \,|\, \hat{r} \,|\, \beta_1 \cdots \beta_n \,\rangle \, \psi^\dagger_{\alpha_n} \cdots \psi^\dagger_{\alpha_n} \, \psi_{\beta_n} \cdots \psi_{\beta_1} \tag{8.40}$$

where

$$\langle\, \alpha_1 \cdots \alpha_n \,|\, \hat{r} \,|\, \beta_1 \cdots \beta_n \,\rangle = \int d^d x_1 \cdots \int d^d x_n \, \varphi^*_{\alpha_1}(\boldsymbol{x}_1) \cdots \varphi^*_{\alpha_n}(\boldsymbol{x}_n) \, \hat{r}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \, \varphi_{\beta_n}(\boldsymbol{x}_n) \cdots \varphi_{\beta_1}(\boldsymbol{x}_1) \quad . \tag{8.41}$$

If the particles have no internal degrees of freedom, then the operators $\hat{r}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ are just functions of the spatial coordinates $\{\boldsymbol{x}_i\}$. If there are (discrete) internal degrees of freedom, then $\hat{r}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ also has operator content in the internal Hilbert space as well.

Finally, if the Hamiltonian is noninteracting, consisting solely of one-body operators, then

$$\hat{H} = \sum_\alpha \varepsilon_\alpha \, \psi^\dagger_\alpha \, \psi_\alpha \quad , \tag{8.42}$$

where $\{\varepsilon_\alpha\}$ is the spectrum of the single particle Hamiltonian.

8.2 Hartree-Fock Theory

Consider the interacting electron Hamiltonian

$$\hat{H} = \sum_{i=1}^N \left\{ -\frac{\hbar^2}{2m} \boldsymbol{\nabla}_i^2 + v(\boldsymbol{x}_i) \right\} + \sum_{i<j}^N u(\boldsymbol{x}_i - \boldsymbol{x}_j) \quad . \tag{8.43}$$

We now endeavor to construct the best possible single Slater determinant state,

$$\begin{aligned} \Psi(\boldsymbol{x}_1\,\sigma_1\,,\,\ldots\,,\,\boldsymbol{x}_N\,\sigma_N) &= \frac{1}{\sqrt{N!}}\,\mathcal{A}\left[\varphi_{\alpha_1}(\boldsymbol{x}_1,\sigma_1)\cdots\varphi_{\alpha_N}(\boldsymbol{x}_N,\sigma_N)\right] \\ &= \frac{1}{\sqrt{N!}}\sum_{\pi\in\mathcal{S}_N}\mathsf{sgn}(\pi)\prod_{i=1}^N\varphi_{\alpha_i}(\boldsymbol{x}_{\pi(i)}\,,\,\sigma_{\pi(i)}) \quad , \end{aligned} \tag{8.44}$$

where $\mathcal{A}$ is the antisymmetrizer, and $\varphi_\alpha(\boldsymbol{x},\sigma)$ is a single particle wavefunction. Typically we will take $\alpha = (j,\gamma)$ to be a composite label, and write

$$\varphi_\alpha(\boldsymbol{x},\sigma) = \varphi_j(\boldsymbol{x})\,\eta_\gamma(\sigma) \quad , \tag{8.45}$$

with $\eta_\gamma(\sigma) = \delta_{\sigma\gamma}$. In second-quantized notation, the wavefunction is given by

$$|\,\Psi\,\rangle = \prod_\alpha^{\mathsf{OCC}} c^\dagger_\alpha\,|\,0\,\rangle \quad . \tag{8.46}$$

The set OCC comprises the N distinct *occupied* orbitals.

The second-quantized Hamiltonian is

$$\hat{H} = \int d^d x\, \psi^\dagger_\sigma(\boldsymbol{x}) \left\{ -\frac{\hbar^2}{2m}\,\delta_{\sigma\tau}\boldsymbol{\nabla}^2 + v^{\sf ext}_{\sigma\tau}(\boldsymbol{x}) \right\} \psi_\tau(\boldsymbol{x}) \\ + \tfrac{1}{2}\int d^d x \int d^d x'\, \psi^\dagger_\sigma(\boldsymbol{x})\,\psi^\dagger_{\sigma'}(\boldsymbol{x}')\,u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}')\,\psi_{\tau'}(\boldsymbol{x}')\,\psi_\tau(\boldsymbol{x}) \quad , \tag{8.47}$$

where $\{\sigma,\tau,\sigma',\tau'\}$ are spin polarizations, and where the two-body interaction for spin-isotropic systems is written as

$$u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') = u^{\sf SCALAR}(\boldsymbol{x}-\boldsymbol{x}')\,\delta_{\sigma\tau}\,\delta_{\sigma'\tau'} + u^{\sf SPIN}(\boldsymbol{x}-\boldsymbol{x}')\,\boldsymbol{\sigma}_{\sigma\tau}\cdot\boldsymbol{\sigma}_{\sigma'\tau'} \quad . \tag{8.48}$$

Here $u^{\sf SCALAR}$ is the scalar component and $u^{\sf SPIN}$ the Heisenberg component of the two-body interaction, and $\boldsymbol{\sigma}$ are the Pauli matrices. Throughout we adopt the Einstein convention over summing over repeated indices.

In order to evaluate the expectation value $E = \langle\,\Psi\,|\,\hat{H}\,|\,\Psi\,\rangle$, we need the following:

$$\langle\Psi|\psi^\dagger_\sigma(\boldsymbol{x})\psi_\tau(\boldsymbol{y})|\Psi\rangle = \sum_\alpha^{\sf OCC}\varphi^*_\alpha(\boldsymbol{x},\sigma)\,\varphi_\alpha(\boldsymbol{y},\tau) \tag{8.49}$$

$$\langle\Psi|\psi^\dagger_\sigma(\boldsymbol{x})\psi^\dagger_{\sigma'}(\boldsymbol{x}')\psi_{\tau'}(\boldsymbol{x}')\psi_\tau(\boldsymbol{x})|\Psi\rangle = \sum_{\alpha,\beta}^{\sf OCC}\varphi^*_\alpha(\boldsymbol{x},\sigma)\,\varphi^*_\beta(\boldsymbol{x}',\sigma')\,\Big(\varphi_\alpha(\boldsymbol{x},\tau)\varphi_\beta(\boldsymbol{x}',\tau') - \varphi_\beta(\boldsymbol{x},\tau)\,\varphi_\alpha(\boldsymbol{x}',\tau')\Big) \quad .$$

This generalizes to

$$\langle\,\Psi\,|\,\psi^\dagger_{\sigma_1}(\boldsymbol{x}_1)\cdots\psi^\dagger_{\sigma_n}(\boldsymbol{x}_n)\,\psi_{\tau_n}(\boldsymbol{x}_n)\cdots\psi_{\tau_1}(\boldsymbol{x}_1)\,|\,\Psi\,\rangle = \sum_{\alpha_1\cdots\alpha_n}^{\sf OCC}\left(\prod_{i=1}^n\varphi^*_{\alpha_i}(\boldsymbol{x}_i,\sigma_i)\right)\det\big[\varphi_{\alpha_j}(\boldsymbol{x}_k,\tau_k)\big] \quad . \tag{8.50}$$

The RHS is necessarily zero if $n > N$ because there is then a linear dependence among the rows of the matrix $M_{j,k} = \det\big[\varphi_{\alpha_j}(\boldsymbol{x}_k,\sigma_k)\big]$.

We now have $E = T + V + U$, with

$$T = \sum_\alpha^{\sf OCC}\int d^d x\,\varphi^*_\alpha(\boldsymbol{x},\sigma)\,\hat{t}_{\sigma\tau}(\boldsymbol{\nabla})\,\varphi_\alpha(\boldsymbol{x},\tau) \qquad , \qquad V = \sum_\alpha^{\sf OCC}\int d^d x\,\varphi^*_\alpha(\boldsymbol{x},\sigma)\,\hat{v}^{\sf ext}_{\sigma\tau}(\boldsymbol{x})\,\varphi_\alpha(\boldsymbol{x},\tau) \tag{8.51}$$

and

$$U = \tfrac{1}{2}\sum_{\alpha,\beta}^{\sf OCC}\int d^d x\int d^d x' u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}')\,\varphi^*_\alpha(\boldsymbol{x},\sigma)\,\varphi^*_\beta(\boldsymbol{x}',\sigma')\,\Big(\varphi_\alpha(\boldsymbol{x},\tau)\,\varphi_\beta(\boldsymbol{x}',\tau') - \varphi_\beta(\boldsymbol{x},\tau)\,\varphi_\alpha(\boldsymbol{x}',\tau')\Big) \quad .$$

Now let's functionally vary with respect to the wavefunction $\varphi^*_\alpha(\boldsymbol{x},\sigma)$. We have

$$\frac{\delta T}{\delta\varphi^*_\alpha(\boldsymbol{x},\sigma)} = \hat{t}_{\sigma\tau}(\boldsymbol{\nabla})\,\varphi_\alpha(\boldsymbol{x},\tau) \qquad , \qquad \frac{\delta V}{\delta\varphi^*_\alpha(\boldsymbol{x},\sigma)} = \hat{v}^{\sf ext}_{\sigma\tau}(\boldsymbol{x})\,\varphi_\alpha(\boldsymbol{x},\tau) \tag{8.52}$$

and

$$\frac{\delta U}{\delta\varphi^*_\alpha(\boldsymbol{x},\sigma)} = \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') \sum_\beta^{\text{OCC}} \varphi^*_\beta(\boldsymbol{x}',\sigma') \Big(\varphi_\alpha(\boldsymbol{x},\tau)\,\varphi_\beta(\boldsymbol{x}',\tau') - \varphi_\beta(\boldsymbol{x},\tau)\,\varphi_\alpha(\boldsymbol{x}',\tau')\Big) \quad . \tag{8.53}$$

In order to maintain orthonormality of the single particle wavefunctions, *i.e.* $\langle\,\varphi_\alpha\,|\,\varphi_\beta\,\rangle = \delta_{\alpha\beta}$, we extremize not E but rather E^*, where

$$E^* = T + V + U - \sum_{\alpha,\beta} \Lambda_{\alpha\beta}\Big(\langle\,\varphi_\alpha\,|\,\varphi_\beta\,\rangle - \delta_{\alpha\beta}\Big) \quad , \tag{8.54}$$

where the $\{\Lambda_{\alpha\beta}\}$ are a set of Lagrange multipliers. The condition $\delta E^* = 0$ now yields

$$\begin{aligned}\Bigg\{-\frac{\hbar^2}{2m}\,\delta_{\sigma\tau}\boldsymbol{\nabla}^2 + v^{\text{ext}}_{\sigma\tau}(\boldsymbol{x}) + \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') \sum_\beta^{\text{OCC}} \varphi^*_\beta(\boldsymbol{x}',\sigma')\,\varphi_\beta(\boldsymbol{x}',\tau')\Bigg\}\varphi_\alpha(\boldsymbol{x},\tau) \\ - \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}')\,\varphi_\alpha(\boldsymbol{x}',\tau') \sum_\beta^{\text{OCC}} \varphi^*_\beta(\boldsymbol{x}',\sigma')\,\varphi_\beta(\boldsymbol{x},\tau) = \sum_\beta^{\text{OCC}} \Lambda_{\alpha\beta}\,\varphi_\beta(\boldsymbol{x},\sigma) \quad .\end{aligned} \tag{8.55}$$

One can show that the matrix Λ must be Hermitian, which means it can be diagonalized by a unitary matrix $U_{a\alpha}$, with $(U\Lambda U^\dagger)_{ab} = \varepsilon_a\,\delta_{ab}$. Defining $\varphi_a(\boldsymbol{x},\sigma) = U_{a\alpha}\,\varphi_\alpha(\boldsymbol{x},\sigma)$, we then obtain the *Hartree-Fock equations*,

$$\begin{aligned}\Bigg\{-\frac{\hbar^2}{2m}\,\delta_{\sigma\tau}\boldsymbol{\nabla}^2 + v^{\text{ext}}_{\sigma\tau}(\boldsymbol{x}) + \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') \sum_b^{\text{OCC}} \varphi^*_b(\boldsymbol{x}',\sigma')\,\varphi_b(\boldsymbol{x}',\tau')\Bigg\}\varphi_a(\boldsymbol{x},\tau) \\ - \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}')\,\varphi_a(\boldsymbol{x}',\tau') \sum_b^{\text{OCC}} \varphi^*_b(\boldsymbol{x}',\sigma')\,\varphi_b(\boldsymbol{x},\tau) = \varepsilon_a\,\varphi_a(\boldsymbol{x},\sigma) \quad ,\end{aligned} \tag{8.56}$$

with no sum on a. The quantities $\{\varepsilon_a\}$ are the *single particle Hartree-Fock energies*. Note that the last term in the curly brackets can be interpreted as a renormalization of the one-body potential, with

$$v^{\text{H}}_{\sigma\tau}(\boldsymbol{x}) = \int d^dx'\; u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') \sum_b^{\text{OCC}} \varphi^*_b(\boldsymbol{x}',\sigma')\,\varphi_b(\boldsymbol{x}',\tau') \quad . \tag{8.57}$$

This is known as the *Hartree potential*. The *Fock term*, arising from exchange, has the interpretation of a *nonlocal* potential, *viz.*

$$v^{\text{F}}_{\sigma\tau}(\boldsymbol{x},\boldsymbol{x}') = -u_{\sigma\tau'\sigma'\tau}(\boldsymbol{x}-\boldsymbol{x}') \sum_b^{\text{OCC}} \varphi^*_b(\boldsymbol{x}',\sigma')\,\varphi_b(\boldsymbol{x},\tau') \quad . \tag{8.58}$$

Thus, the Hartree-Fock (HF) equations may be written

$$\Bigg\{-\frac{\hbar^2}{2m}\,\delta_{\sigma\tau}\boldsymbol{\nabla}^2 + v^{\text{ext}}_{\sigma\tau}(\boldsymbol{x}) + v^{\text{H}}_{\sigma\tau}(\boldsymbol{x})\Bigg\}\varphi_a(\boldsymbol{x},\tau) + \int d^dx'\; v^{\text{F}}_{\sigma\tau}(\boldsymbol{x},\boldsymbol{x}')\,\varphi_a(\boldsymbol{x}',\tau) = \varepsilon_a\,\varphi_a(\boldsymbol{x},\sigma) \quad . \tag{8.59}$$

Note that if we multiply Eqn. 8.56 by $\varphi_a^*(\boldsymbol{x},\sigma)$ and then integrate over $\boldsymbol{x}$ and sum on σ, we obtain the relation

$$\varepsilon_a = \int d^dx\ \varphi_a^*(\boldsymbol{x},\sigma)\left\{-\frac{\hbar^2}{2m}\,\delta_{\sigma\tau}\boldsymbol{\nabla}^2 + v^{\sf ext}_{\sigma\tau}(\boldsymbol{x}) + v^{\sf H}_{\sigma\tau}(\boldsymbol{x})\right\}\varphi_\alpha(\boldsymbol{x},\tau) \\ + \int d^dx \int d^dx'\ v^{\sf F}_{\sigma\tau}(\boldsymbol{x},\boldsymbol{x}')\,\varphi_a^*(\boldsymbol{x},\sigma)\,\varphi_a(\boldsymbol{x}',\tau) \quad . \tag{8.60}$$

If we now sum over all occupied states a, we obtain the result

$$\sum_a^{\sf occ}\varepsilon_a = T + V + 2\,U \quad . \tag{8.61}$$

Thus, the sum over all the single particle HF energies *is not* the total energy $E = \langle\,\Psi\,|\,\hat{H}\,|\,\Psi\,\rangle$. Rather, the interpretation of ε_a is that

$$\frac{\delta E}{\delta N} = E(N) - E(N-1) = \varepsilon_a \tag{8.62}$$

when the electron in state a is removed to form the $(N-1)$-electron system. Put another way, the energy required to transfer an electron from an orbital $|\,a\,\rangle$ to an orbital $|\,b\,\rangle$ is $\varepsilon_b - \varepsilon_a$. These results presume that this transfer does not affect the other wavefunctions $\varphi_c(\boldsymbol{x},\sigma)$ for $c \neq a, b$. This is presumably valid in the thermodynamic limit $N \to \infty$, but need not be so for finite N. The difference in ground state energies is thus given by the smallest value $\mathsf{min}_a\,\varepsilon_a$.

When $v^{\sf ext}_{\sigma\tau}(\boldsymbol{x}) = v(\boldsymbol{x})\,\delta_{\sigma\tau}$ and $u_{\sigma\tau\sigma'\tau'}(\boldsymbol{x}-\boldsymbol{x}') = u(\boldsymbol{x}-\boldsymbol{x}')\,\delta_{\sigma\tau}\,\delta_{\sigma'\tau'}$, the spin degree of freedom is just a spectator, and we may obtain a solution where the states are labeled by an index $j \in \{1,\ldots,N\}$ and a spin polarization $\sigma \in \{\uparrow,\downarrow\}$, with polarization-independent single particle HF energies ε_j. The HF equations then become

$$\left\{-\frac{\hbar^2}{2m}\,\boldsymbol{\nabla}^2 + v_{\sf ext}(\boldsymbol{x}) + v^{\sf H}(\boldsymbol{x})\right\}\varphi_j(\boldsymbol{x}) + \int d^dx'\ v^{\sf F}(\boldsymbol{x}-\boldsymbol{x}')\,\varphi_j(\boldsymbol{x}') = \varepsilon_j\,\varphi_j(\boldsymbol{x}) \quad , \tag{8.63}$$

where

$$v^{\sf H}(\boldsymbol{x}) = 2\int d^dx'\ u(\boldsymbol{x}-\boldsymbol{x}')\sum_l^{\sf occ}|\varphi_l(\boldsymbol{x}')|^2 \\ v^{\sf F}(\boldsymbol{x},\boldsymbol{x}') = -u(\boldsymbol{x}-\boldsymbol{x}')\sum_l^{\sf occ}\varphi_l^*(\boldsymbol{x}')\,\varphi_l(\boldsymbol{x}) \quad . \tag{8.64}$$

Translationally invariant systems

For translationally invariant systems, the plane wave basis $\varphi_{\boldsymbol{k}}(\boldsymbol{x}) = V^{-1/2}\exp(i\boldsymbol{k}\cdot\boldsymbol{x})$ yields a solution. The Hartree and Fock potentials are then

$$\begin{aligned} v^{\mathsf{H}}(\boldsymbol{x}) &= n\,\hat{u}(0) \\ v_{\mathsf{F}}(\boldsymbol{x}-\boldsymbol{x}') &= -u(\boldsymbol{x}-\boldsymbol{x}')\int\frac{d^dk}{(2\pi)^d}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{x}')}\,\Theta(k_{\mathsf{F}}-k) \quad , \end{aligned} \tag{8.65}$$

where

$$n = 2\int\frac{d^dk}{(2\pi)^d}\,\Theta(k_{\mathsf{F}}-k) = \frac{2\Omega_d\,k_{\mathsf{F}}^d}{d\,(2\pi)^d} \tag{8.66}$$

is the number density. Here $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the total solid angle in d space dimensions. One then has the Hartree-Fock energies

$$\varepsilon(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{k}^2}{2m} + n\,\hat{u}(0) - \int\frac{d^dk'}{(2\pi)^d}\,\hat{u}(\boldsymbol{k}-\boldsymbol{k}')\,\Theta(k_{\mathsf{F}}-k') \quad , \tag{8.67}$$

where $\hat{u}(\boldsymbol{k})$ is the Fourier transform of $u(\boldsymbol{x})$.

HF theory for atoms

In atomic physics, we have the one-body ion core potential $v^{\mathsf{ext}}(\boldsymbol{x}) = -Ze^2/|\boldsymbol{x}|$ (neglecting spin-orbit effects) and the two-body electron-electron interaction $u(\boldsymbol{x}-\boldsymbol{x}') = e^2/|\boldsymbol{x}-\boldsymbol{x}'|$. It is then a good approximation to assume that the Hartree-Fock wavefunctions $\varphi_i(\boldsymbol{x})$ are of the form

$$\varphi_\alpha(\boldsymbol{x}) = R_{nl}(r)\,Y_{lm}(\theta,\phi) \quad , \tag{8.68}$$

independent of σ. This follows from the rotational isotropy of the ion core potential. We can then classify the single particle states by the quantum numbers $n \in \{1,2,\ldots\}$, $l \in \{0,1,\ldots,n-1\}$, $m_l \in \{-l,\ldots,+l\}$, and $m_s = \pm\frac{1}{2}$. The essential physics introduced by the Hartree-Fock method is that of *screening*. Close to the origin, a given electron senses a potential $-Ze^2/r$ due to the unscreened nucleus. Farther away, though, the nuclear charge is screened by the core electrons, and the potential decays faster than $1/r$.[2] Whereas states of different l and identical n are degenerate for the noninteracting hydrogenic atom, when the nuclear potential is screened, states of different l are no longer degenerate. Smaller l means lower energy, since these states are localized closer to the nucleus, where the potential is large and negative and relatively unscreened. Based on the HF energy levels, the order in which the electron shells are filled throughout the periodic table is roughly given by that in Fig. 14.1. This is known as the *Aufbau principle* from the German *Aufbau* = "building up" (see Fig. 14.1). The order in which the

[2]Within the Thomas-Fermi approximation, the potential at long distances decays as $-\mathcal{C}e^2a_{\mathrm{B}}^3/r^4$, where $\mathcal{C} \simeq 100$ is a numerical factor, independent of Z.



Figure 8.1: The *Aufbau* principle and the diagonal rule. Image credit: Wikipedia.

orbitals are filled follows the *diagonal rule*, which says that orbitals with lower values of $n + l$ are filled before those with higher values, and that in the case of equal $n + l$ values, the orbital with the lower n is filled first. For a given l and n there are $(2s + 1) \times (2l + 1) = 4l + 2$ states ($s = \frac{1}{2}$), labeled by the angular momentum and spin polarization quantum numbers m_l and m_s ; this group of orbitals is called a *shell*.

HF theory for the electron gas

The so-called *jellium* model of the electron gas consists of N electrons moving in a uniform neutralizing (*i.e.* positively charged) background. The system is translationally invariant, hence the HF wavefunctions are labeled by wavevectors $\boldsymbol{k}$, with $\varphi_{\boldsymbol{k}}(\boldsymbol{x}) = V^{-1/2}\,\exp(i\boldsymbol{k}\cdot\boldsymbol{x})$. In all dimensions, the neutralizing background contributions precisely cancel the Hartree term in the energy. The single particle HF energies are then given by ($d = 3$ dimensions assumed)

$$\varepsilon(\boldsymbol{k}) = \frac{\hbar^2\boldsymbol{k}^2}{2m} - \int \frac{d^3k'}{(2\pi)^3}\,\frac{4\pi e^2}{(\boldsymbol{k}-\boldsymbol{k}')^2}\,\Theta(k_{\rm F}-k') = \varepsilon_0(\boldsymbol{k}) + \Sigma(\boldsymbol{k}) \quad , \tag{8.69}$$

where $n = k_{\rm F}^3/3\pi^2$ is the electron density. The *bare dispersion* for noninteracting electrons is $\varepsilon_0(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$, and the *self-energy* is

$$\Sigma(\boldsymbol{k}) = \frac{e^2 k_{\rm F}}{2\pi}\left(\frac{k^2-k_{\rm F}^2}{kk_{\rm F}}\,\ln\left|\frac{k+k_{\rm F}}{k-k_{\rm F}}\right| - 2\right) \quad . \tag{8.70}$$

If we expand about the Fermi wavevector, writing $\boldsymbol{k} = (k_{\rm F}+q)\,\hat{\boldsymbol{n}}$, where $\hat{\boldsymbol{n}}$ is any unit vector, we obtain

$$\varepsilon(k_{\rm F}+q) = \varepsilon_{\rm F} + \frac{\hbar^2 k_{\rm F}}{m}\,q + \frac{e^2}{\pi}\,q\,\ln\left|\frac{2k_{\rm F}}{q}\right| + \mathcal{O}(q^2) \quad , \tag{8.71}$$

where $\varepsilon_{\rm F} = \frac{\hbar^2 k_{\rm F}^2}{2m} - \frac{e^2 k_{\rm F}}{\pi}$ is the Fermi energy within HF theory. The velocity in the vicinity perpendicular to the Fermi surface is then with

$$v(q) = \frac{1}{\hbar}\frac{\partial\varepsilon(k_{\rm F}+q)}{\partial q} = \frac{\hbar k_{\rm F}}{m} + \frac{e^2}{\pi\hbar}\left(\ln\left|\frac{2k_{\rm F}}{q}\right| - 1\right) \quad , \tag{8.72}$$

which diverges logarithmically in the limit $q \to 0$. This divergence of the Fermi velocity is an artifact of the HF approach, which neglects screening effects, which we shall consider later on.

The total kinetic energy per particle is given by

$$\frac{T}{N} = \frac{1}{N}\times 2\sum_{|\boldsymbol{k}|<k_{\rm F}}\varepsilon_0(k) = \frac{2}{n}\int\frac{d^3k}{(2\pi)^3}\,\frac{\hbar^2k^2}{2m}\Theta(k_{\rm F}-k) = \frac{3\hbar^2k_{\rm F}^2}{10m} \quad . \tag{8.73}$$

The total potential energy comes from the sum of the (i) self-interaction of the neutralizing background, (ii) the energy of interaction between the neutralizing background and the uniform density electron gas, (iii) the Hartree energy of the electron gas $E_{\rm H}$, and (iv) the exchange energy of the electron gas $E_{\rm X}$. The first three of these sum to zero, leaving $E_{\rm X}$, where

$$\begin{aligned}\frac{E_{\rm X}}{N} &= \frac{1}{2N}\times 2\sum_{|\boldsymbol{k}|<k_{\rm F}}\Sigma(k) = \frac{1}{n}\int\frac{d^3k}{(2\pi)^3}\,\Sigma(k)\,\Theta(k_{\rm F}-k) \\ &= -\frac{e^2k_{\rm F}}{4\pi^3}\int_0^1 dx\,x^2\left\{2+\frac{1-x^2}{x}\,\ln\left|\frac{1+x}{1-x}\right|\right\} = -\frac{3e^2k_{\rm F}}{4\pi} \quad .\end{aligned} \tag{8.74}$$

Note the factor of $\frac{1}{2}$ multiplying the above result, which corrects for the factor of 2 in Eqn. 8.61.

It is conventional to define a dimensionless length $r_{\rm s}$ according to $\frac{4}{3}\pi(r_{\rm s}\,a_{\rm B})^3\,n \equiv 1$, where $a_{\rm B} = \hbar^2/me^2$ is the Bohr radius. Thus

$$r_{\rm s} = \left(\frac{3}{4\pi}\right)^{1/3} n^{-1/3}\,a_{\rm B}^{-1} \qquad , \qquad k_{\rm F} = \left(\frac{9\pi}{4}\right)^{1/3} r_{\rm s}^{-1}a_{\rm B}^{-1} \quad . \tag{8.75}$$

The kinetic energy per particle is then

$$\frac{T}{N} = \frac{3\hbar^2}{10m}\left(\frac{9\pi}{4}\right)^{2/3} r_{\rm s}^{-2}\,a_{\rm B}^{-2} = \frac{3}{5}\left(\frac{9\pi}{4}\right)^{2/3}\frac{e^2}{2a_{\rm B}}\cdot\frac{1}{r_{\rm s}^2} \simeq \frac{2.21}{r_{\rm s}^2}\,{\rm Ryd} \quad , \tag{8.76}$$

while the exchange energy per particle is

$$\frac{E_{\rm X}}{N} = -\frac{3}{2\pi}\left(\frac{9\pi}{4}\right)^{1/3}\frac{e^2}{2a_{\rm B}}\cdot\frac{1}{r_{\rm s}} \simeq -\frac{0.916}{r_{\rm s}}\,{\rm Ryd} \quad , \tag{8.77}$$

where $1\,{\rm Ryd} = e^2/2a_{\rm B} = 13.6057\,{\rm eV}$. Thus, interaction effects dominate when $r_{\rm s}$ is large, meaning the density n is small. This is because the kinetic energy term involves two gradients, hence scales as L^{-2}, whereas the Coulomb interaction scales as L^{-1}. For short-ranged interactions, interaction effects dominate at large densities, which perhaps is more intuitive.

8.3 Density Functional Theory

In any interacting electronic system, the kinetic energy of each electron is given by $\boldsymbol{p}^2/2m$ and the interaction between any two electrons is the Coulomb energy $v(\boldsymbol{x}-\boldsymbol{x}') = e^2/|\boldsymbol{x}-\boldsymbol{x}'|$. What differs in the description from one material to the next is the one-body potential $v_{\text{ext}}(\boldsymbol{x})$. This is what distinguishes the Hamiltonian for table salt (NaCl) from that of elemental iron (Fe)[3]. Thus, the ground state wavefunction of any many-electron system $\Psi_{\sigma_1\cdots\sigma_N}(\boldsymbol{x}_1,\ldots\boldsymbol{x}_N)$ is completely determined by $v_{\text{ext}}(\boldsymbol{x})$. From the wavefunction we can also determine the one-body electron density for spin polarization σ, which is

$$n_\sigma(\boldsymbol{x}) = \langle\,\Psi\,|\,\psi_\sigma^\dagger(\boldsymbol{x})\,\psi_\sigma(\boldsymbol{x})\,|\,\Psi\,\rangle = N\sum_{\sigma_2}\cdots\sum_{\sigma_N}\int d^3x_2\cdots\int d^3x_N\,\left|\Psi_{\sigma\,\sigma_2\cdots\sigma_N}(\boldsymbol{x},\boldsymbol{x}_2,\ldots\boldsymbol{x}_N)\right|^2 \quad . \tag{8.78}$$

Summing over both spin polarizations, we obtain the total electron number density,

$$n(\boldsymbol{x}) = \sum_\sigma n_\sigma(\boldsymbol{x}) \quad . \tag{8.79}$$

Though at first consideration counterintuitive, it turns out that the entire ground state wavefunction can be considered to be a *functional* of the electron number density. A number of highly consequential and extremely useful results follow. This is the subject of *density functional theory* (DFT), which revolutionized the study of electronic structure of molecules and solids[4], and which was recognized by the 1998 Nobel Prize in Chemistry to Walter Kohn and John Pople.

8.3.1 Hohenberg-Kohn theorems

The mathematical basis underpinning DFT are two theorems by Hohenberg and Kohn.

THEOREM #1 : The ground state energy of a many-electron system is a functional of the total electron density $n(\boldsymbol{x}) = \sum_\sigma \langle\,\Psi\,|\,\psi_\sigma^\dagger(\boldsymbol{x})\,\psi_\sigma(\boldsymbol{x})\,|\,\Psi\,\rangle$.

To prove this theorem – in our callow physicist's sort of way – let $|\,\Psi\,\rangle$ and $|\,\Psi'\,\rangle$ be the ground states corresponding respectively to the two external potentials $v_{\text{ext}}(\boldsymbol{x})$ and $v'_{\text{ext}}(\boldsymbol{x})$. We assume these ground states are normalized and distinct, meaning that in any finite volume their overlap is less than unity in magnitude. The Hamiltonians are $\hat{H} = \hat{T}+\hat{V}+\hat{U}$ and $\hat{H}' = \hat{T}+\hat{V}'+\hat{U}$, and it must be that

$$E' = \langle\,\Psi'\,|\,H'\,|\,\Psi'\,\rangle < \langle\,\Psi\,|\,\hat{H}'\,|\,\Psi\,\rangle = \langle\,\Psi\,|\,\hat{H}\,|\,\Psi\,\rangle + \langle\,\Psi\,|\,\hat{V}'-\hat{V}\,|\,\Psi\,\rangle \quad , \tag{8.80}$$

[3]For large Z ions, the spin-orbit interaction is also important. This can be included in the one-body potential $v_{\text{ext}}(\boldsymbol{x})$ by extending v_{ext} to a function $v_{\text{ext}}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\sigma})$ of position, momentum, and spin.

[4]Here we follow the discussion in chapter 15 of Girvin and Yang, *Condensed Matter Physics* (Cambridge, 2019). See also G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge, 2005).

where

$$\hat{V} = \int d^3x\, v_{\text{ext}}(\boldsymbol{x})\, \hat{n}(\boldsymbol{x}) \qquad , \qquad \hat{V}' = \int d^3x\, v'_{\text{ext}}(\boldsymbol{x})\, \hat{n}(\boldsymbol{x}) \quad . \tag{8.81}$$

Thus we conclude $E' < E + \langle\, \Psi \,|\, \hat{V}' - \hat{V} \,|\, \Psi \,\rangle$. But then $E < E' + \langle\, \Psi' \,|\, \hat{V} - \hat{V}' \,|\, \Psi' \,\rangle$ as well, simply by exchanging the definitions of the primed and unprimed systems. Adding these two results we obtain

$$0 < \langle\, \Psi \,|\, \hat{V}' - \hat{V} \,|\, \Psi \,\rangle + \langle\, \Psi' \,|\, \hat{V} - \hat{V}' \,|\, \Psi' \,\rangle \quad . \tag{8.82}$$

and thus if $\langle\, \Psi \,|\, \hat{n}(\boldsymbol{x}) \,|\, \Psi \,\rangle = \langle\, \Psi' \,|\, \hat{n}(\boldsymbol{x}) \,|\, \Psi' \,\rangle = n(\boldsymbol{x})$, we arrive at a contradiction: $0 < 0$. We conclude that the ground state wavefunctions for different one-body potentials cannot yield the exact same density $n(\boldsymbol{x})$.

THEOREM #2: The ground state energy may be expressed as a functional of the density, $E[n]$, such that minimizing this functional with respect to $n(\boldsymbol{x})$ yields the true ground state density.

To see that this is the case, note that theorem #1 entails that for each density $n(\boldsymbol{x})$ corresponding to the many-body ground state in some external potential, there is a corresponding ground state wavefunction $|\, \Psi[n] \,\rangle$. Now define the functional

$$E[n] = \overbrace{\langle\, \Psi[n] \,|\, T + U \,|\, \Psi[n] \,\rangle}^{E_{\text{jel}}[n]} + \int d^3x\, v_{\text{ext}}(\boldsymbol{x})\, n(\boldsymbol{x}) \quad . \tag{8.83}$$

Here $E_{\text{jel}}[n]$ is the energy functional for jellium, with $v_{\text{ext}} = 0$. Note that this requires that we take the two-body potential term $\hat{U}$ to be

$$\hat{U} = \frac{1}{2} \int d^3x \int d^3x' \, \big(\hat{n}(\boldsymbol{x}) - n_0\big) \, \frac{e^2}{|\boldsymbol{x} - \boldsymbol{x}'|} \, \big(\hat{n}(\boldsymbol{x}') - n_0\big) \quad , \tag{8.84}$$

where n_0 corresponds to a uniform neutralizing background. Charge neutrality requires that

$$\lim_{N\to\infty} \frac{1}{N} \int d^3x \, \big(n(\boldsymbol{x}) - n_0\big) = 0 \tag{8.85}$$

lest the Coulomb energy diverge. Now, since $E[\tilde{n}] = \langle\, \Psi[\tilde{n}] \,|\, H \,|\, \Psi[\tilde{n}] \,\rangle > E[n]$, which follows by considering $|\, \Psi[\tilde{n}] \,\rangle$ to be a variational ground state for the Hamiltonian whose true ground state density is $n(\boldsymbol{x})$, we conclude that the functional $E[n]$ is indeed minimized when $n(\boldsymbol{x})$ is the true ground state density when the external potential is $v_{\text{ext}}(\boldsymbol{x})$.

How do we know that a given density $n(\boldsymbol{x})$ corresponds to the actual ground state density for Coulomb-interacting electrons in some external potential $v_{\text{ext}}(\boldsymbol{x})$? The short answer is that we don't. Indeed for the single particle system, where there are no Coulomb interactions, any density function $n(\boldsymbol{x})$ which vanishes at any location cannot possibly be the actual ground state density for any nonsingular potential $v_{\text{ext}}(\boldsymbol{x})$ due to the Perron-Frobenius "no nodes theorem". Functions $n(\boldsymbol{x})$ which do correspond to the ground state density of a fermionic system for

some potential $v_{\rm ext}(\boldsymbol{x})$ are called *V-representable*. A weaker condition is that of *N-representability*, which means that for a given density function $n(\boldsymbol{x})$, there exists an N-fermion wavefunction $\Psi_{\sigma_1\cdots\sigma_N}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$ such that

$$\begin{aligned} n(\boldsymbol{x}) &= \sum_\sigma \langle\,\Psi\,|\,\psi_\sigma^\dagger(\boldsymbol{x})\,\psi_\sigma(\boldsymbol{x})\,|\,\Psi\,\rangle \\ &= N\sum_\sigma\sum_{\sigma_2}\cdots\sum_{\sigma_N}\int d^3x_2\cdots\int d^3x_N\,\left|\Psi_{\sigma\sigma_2\cdots\sigma_N}(\boldsymbol{x},\boldsymbol{x}_2,\ldots\boldsymbol{x}_N)\right|^2 \quad . \end{aligned} \tag{8.86}$$

Levy[5] and Lieb [6] showed that one could extend the domain of density functionals thusly, so that the energy minimization is to be carried out over all $|\,\Psi\,\rangle$ such that $\langle\,\Psi\,|\,\hat{n}(\boldsymbol{x})\,|\,\Psi\,\rangle = n(\boldsymbol{x})$, where $\hat{n}(\boldsymbol{x}) = \sum_\sigma \psi_\sigma^\dagger(\boldsymbol{x})\,\psi_\sigma(\boldsymbol{x})$. Note that while obtaining $n(\boldsymbol{x})$ from $|\,\Psi\,\rangle$ is formally defined by Eqn. 8.86, the inverse process, by which one extracts an N-body wavefunction $\Psi_{\sigma_1\cdots\sigma_N}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)$ from a given N-representable density function $n(\boldsymbol{x})$ is impractically complex for any $N > 1$. The virtue of the Kohn-Sham procedure, discussed in the next section, is that it provides us with a constructive way to implement the extremization procedure within a class of many-body wavefunctions.

8.3.2 Kohn-Sham equations

For Coulomb-interacting electrons, the functional $E_{\rm jel}[n]$ is *universal* and is given by

$$E_{\rm jel}[n] = \min_{|\Psi\rangle\to n(\boldsymbol{x})}\,\langle\,\Psi[n]\,|\,\hat{T}+\hat{U}\,|\,\Psi[n]\,\rangle \quad , \tag{8.87}$$

where the minimization is with respect to all totally antisymmetric N-body wavefunctions yielding a one-body density $n(\boldsymbol{x})$. For *noninteracting* systems, the ground state is a Slater determinant $|\,\Psi_{\rm S}[n]\,\rangle$, and we define the functional

$$T_{\rm S}[n] \equiv \min_{|\Psi_{\rm S}\rangle\to n(\boldsymbol{x})}\,\langle\,\Psi_{\rm S}[n]\,|\,\hat{T}\,|\,\Psi_{\rm S}[n]\,\rangle \quad , \tag{8.88}$$

where $|\,\Psi_{\rm S}\,\rangle$ is an N-particle Slater determinant. We may write

$$n(\boldsymbol{x}) = \sum_\alpha\sum_\sigma n_\alpha\left|\varphi_\alpha(\boldsymbol{x},\sigma)\right|^2 \quad , \tag{8.89}$$

where $n_\alpha \in \{0,1\}$ is the occupation of the single particle state $\varphi_\alpha(\boldsymbol{x},\sigma)$, with $N = \sum_\alpha n_\alpha$. We may write

$$T_{\rm S}[n] = -\frac{\hbar^2}{2m}\sum_\alpha n_\alpha\,\langle\,\varphi_\alpha\,|\,\nabla^2\,|\,\varphi_\alpha\,\rangle \quad . \tag{8.90}$$

[5] M. Levy, *Proc. Nat. Acad. Sci* **76**, 6062 (1979).
[6] E. H. Lieb, *Int. J. Quantum Chem.* **24**, 243 (1983).

Aside – I want to comment yet again on the extremely complex and unusual nature of the functionals $E_{\mathsf{jel}}[n]$ and $T_{\mathsf{S}}[n]$. Functions, such as the iconic $f(x)$, eat numbers ($x \in \mathbb{R}$) and excrete numbers ($f(x) \in \mathbb{R}$). *Functionals*, such as $F[f(x)]$, by contrast eat entire *functions* ($f(x) \in C^\infty(\mathbb{R})$) and excrete numbers[7] ($F[f(x)] \in \mathbb{R}$). Usually it is the case that the functionals we deal with are specified explicitly. Such it is with the action functional in classical mechanics, $S[q(t)]$, where

$$S[q(t)] = \int_{t_1}^{t_2} dt\; L(q, \dot{q}, t) \tag{8.91}$$

with $L = \frac{1}{2}m\dot{q}^2 - V(q)$. In this case, if you give me the function $q(t)$, which typically must satisfy certain boundary conditions such as being fixed at the endpoints, I can in principle perform the above integral and hand you back $S[q(t)]$. Things are not so straightforward with regard to $T_{\mathsf{S}}[n]$ and $E_{\mathsf{jel}}[n]$. Rather, the prescriptions are as follows:

- For $T_{\mathsf{S}}[n]$, you give me some function $n(\boldsymbol{x}) \in C^\infty(\mathbb{R}^d)$ and I rummage through my file cabinet of N-particle Slater determinant wavefunctions $\Psi_{\mathsf{S}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \mathsf{det}\{\varphi_{\alpha_i}(\boldsymbol{x}_j)\}$ constructed from inequivalent orthonormal bases, and I set aside only those functions for which the density $\langle\, \Psi_{\mathsf{S}} \,|\, \hat{n}(\boldsymbol{x}) \,|\, \Psi_{\mathsf{S}} \,\rangle = n(\boldsymbol{x})$ agrees with your specified function. Then, one-by-one, I go through this collection, computing $\langle\, \Psi_{\mathsf{S}} \,|\, \hat{T} \,|\, \Psi_{\mathsf{S}} \,\rangle$ for each, and I find which Ψ_{S} yields the lowest expectation value, which is then the value of $T_{\mathsf{S}}[n]$.
- For $E_{\mathsf{jel}}[n]$, you give me a function $n(\boldsymbol{x})$, and I along with a bunch of volunteers search my giant warehouse of totally antisymmetric N-particle functions $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for those Ψ yielding $\langle\, \Psi \,|\, \hat{n}(\boldsymbol{x}) \,|\, \Psi \,\rangle = n(\boldsymbol{x})$.[8] For each of these surviving wavefunctions, we evaluate $\langle\, \Psi \,|\, \hat{T} + \hat{U} \,|\, \Psi \,\rangle$ and find which Ψ yields the lowest expectation value. This is then the value of $E_{\mathsf{jel}}[n]$.

As you can see, $T_{\mathsf{S}}[n]$ and $E_{\mathsf{jel}}[n]$ are indeed functionals of $n(\boldsymbol{x})$, because there is an explicit, if impractical, prescription for how they may be evaluated. Turning the evaluation of $E_{\mathsf{jel}}[n]$ into an implementable variational scheme was the genius of Kohn and Sham, to whose program we now return.

Having defined the functionals $E_{\mathsf{jel}}[n]$ and $T_{\mathsf{S}}[n]$, we next define the *exchange-correlation functional* $E_{\mathsf{XC}}[n]$ according to the relation

$$E_{\mathsf{jel}}[n] = T_{\mathsf{S}}[n] + E_{\mathsf{H}}[n] + E_{\mathsf{XC}}[n] \quad , \tag{8.92}$$

where $E_{\mathsf{H}}[n]$ is the Hartree functional,

$$E_{\mathsf{H}}[n] = \frac{1}{2}\int d^3x \int d^3x'\, \big(n(\boldsymbol{x}) - n_0\big)\, \frac{e^2}{|\boldsymbol{x} - \boldsymbol{x}'|}\, \big(n(\boldsymbol{x}') - n_0\big) \quad . \tag{8.93}$$

[7] We can of course generalize to complex functions and complex functionals and functions of several variables $f(\boldsymbol{x}) \in C^\infty(\mathbb{R}^n)$.

[8] This is of course vastly bigger than my file cabinet of Slater determinants, a copy of which is stored somewhere in the warehouse, since $\big\{\Psi_{\mathsf{S}} \,|\, \langle\, \Psi_{\mathsf{S}} \,|\, \hat{n} \,|\, \Psi_{\mathsf{S}} \,\rangle = n\big\} \in \big\{\Psi \,|\, \langle\, \Psi \,|\, \hat{n} \,|\, \Psi \,\rangle = n\big\}$.

Note that Eqn. 8.92 is a definition of the functional $E_{\rm XC}[n]$ in terms of the universal functionals $E_{\rm jel}[n]$ (which exists but is unknown), $E_{\rm H}[n]$ (which is explicitly given in Eqn. 8.93), and $T_{\rm S}[n]$ (which is given in Eqn. 8.90). Note that $T_{\rm S}[n]$ is the kinetic energy of a fictional noninteracting fermion system which has the same density $n(\boldsymbol{x})$ as the interacting system under consideration.

At this point, rather than vary with respect to $n(\boldsymbol{x})$, we instead vary with respect to each of the single particle wavefunctions $\varphi^*_\alpha(\boldsymbol{x},\sigma)$, subject to the conditions of orthonormality.

This results in the Kohn-Sham equations,

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + v_{\rm ext}(\boldsymbol{x}) + v_{\rm H}(\boldsymbol{x}) + v_{\rm XC}(\boldsymbol{x})\right\}\varphi_\alpha(\boldsymbol{x},\sigma) = \varepsilon_\alpha\,\varphi_\alpha(\boldsymbol{x},\sigma) \quad , \tag{8.94}$$

where $n_\alpha = 1$,

$$v_{\rm H}(\boldsymbol{x}) = \frac{\delta E_{\rm H}[n]}{\delta n(\boldsymbol{x})} = \int d^3x'\,\frac{e^2}{|\boldsymbol{x}-\boldsymbol{x}'|}\left(n(\boldsymbol{x}') - n_0\right) \tag{8.95}$$

and where

$$v_{\rm XC}(\boldsymbol{x}) = \frac{\delta E_{\rm XC}[n]}{\delta n(\boldsymbol{x})} \tag{8.96}$$

are functional derivatives. Note that we have used the functional chain rule,

$$\frac{\delta F[n]}{\delta\varphi^*_\alpha(\boldsymbol{x})} = \frac{\delta F[n]}{\delta n(\boldsymbol{x})}\cdot\frac{\delta n(\boldsymbol{x})}{\delta\varphi^*_\alpha(\boldsymbol{x})} = \frac{\delta F[n]}{\delta n(\boldsymbol{x})}\,\varphi_\alpha(\boldsymbol{x}) \tag{8.97}$$

for any functional $F[n]$. Note that $v_{\rm XC}(\boldsymbol{x})$ is local, unlike in Hartree-Fock theory where the Fock potential $v_{\rm F}(\boldsymbol{x},\boldsymbol{x}')$ is nonlocal.

It is worth emphasizing that while the Kohn-Sham orbitals $\varphi_\alpha(\boldsymbol{x},\sigma)$ have no obvious physical significance, they are often interpreted as Bloch energy bands for the interacting system (whatever that means!). The KS eigenvalues ε_α do not in general correspond to physical excitation energies of the system, and the Slater determinant formed from the N lowest-lying KS orbitals is in general *not* a good approximation to the actual ground state wavefunction. Indeed the HF wavefunction is often a better approximation in that regard. However, in the $N\to\infty$ limit, it can be proven[9] that in gapless systems the eigenvalue ε_N corresponding to the highest occupied KS energy state is indeed the actual Fermi energy of the system.

At this point, the problem has been reduced to finding the best approximation to the unknown functional $E_{\rm XC}[n]$.

[9] G. Giuliani and G. Vignali, *op cit.*

LDA: the local density approximation

The most commonly used such approximation is called the *local density approximation*, or LDA. One writes

$$E_{\rm XC}^{\rm LDA} = \int d^3x\, n(\boldsymbol{x})\, \varepsilon_{\rm XC}\big(n(\boldsymbol{x})\big) \quad . \tag{8.98}$$

The quantity $\varepsilon_{\rm XC}\big(n(\boldsymbol{x})\big)$ is the exchange-correlation energy per particle for uniform density jellium. Taking the functional derivative,

$$\mu_{\rm XC}\big(n(\boldsymbol{x})\big) = \frac{\delta E_{\rm XC}^{\rm LDA}}{\delta n(\boldsymbol{x})} = \varepsilon_{\rm XC}\big(n(\boldsymbol{x})\big) + n(\boldsymbol{x})\, \frac{\partial \varepsilon_{\rm XC}(n)}{\partial n}\bigg|_{n(\boldsymbol{x})} \quad . \tag{8.99}$$

Recall how in the HF approximation in $d = 3$, $E_{\rm X}/V = -(3e^2/4\pi)\, nk_{\rm F} \propto n^{4/3}$, hence

$$\mu_{\rm XC}(n) = \tfrac{4}{3}\, \varepsilon_{\rm XC}(n) = -\frac{e^2 k_{\rm F}(n)}{\pi} \quad , \tag{8.100}$$

where $k_{\rm F}(n) = (3\pi^2 n)^{1/3}$.

Gradient expansions

If we expand about the jellium density n_0, we may write

$$E_{\rm XC}^{\rm GEA}[n_0 + \delta n] = \frac{1}{2}\int d^3x \left\{ A(n_0)\, \big[\delta n(\boldsymbol{x})\big]^2 + B(n_0)\, \big[\boldsymbol{\nabla}\delta n(\boldsymbol{x})\big]^2 + \ldots \right\} \quad . \tag{8.101}$$

This procedure goes under the name *gradient expansion approximation*, or GEA[10]. One can also define spin-resolved expansions *viz.*

$$E_{\rm XC}^{\rm SGEA}[n_{0\uparrow} + \delta n_\uparrow, n_{0\downarrow} + \delta n_\downarrow] = \frac{1}{2}\int d^3x \left\{ A_{\sigma\sigma'}(n_{0\uparrow}, n_{0\downarrow})\, \delta n_\sigma(\boldsymbol{x})\, \delta n_{\sigma'}(\boldsymbol{x}) \right. \tag{8.102}$$
$$\left. + B_{\sigma\sigma'}(n_{0\uparrow}, n_{0\downarrow})\, \boldsymbol{\nabla}\delta n_\sigma(\boldsymbol{x}) \cdot \boldsymbol{\nabla}\delta n_{\sigma'}(\boldsymbol{x}) + \ldots \right\} \quad .$$

Alas, in applications to real materials, the GEA is often less accurate than the LDA.

Major applications of DFT

Girvin and Yang provide a brief list of popular applications of DFT. For each such application there are many thousands of papers in the literature:

[10] Density functional theory is replete with acronyms.

- *Structural determination* : Given a set of atoms, what sort of crystal structure will they form? The external potential is given by

$$v_{\text{ext}}(\boldsymbol{x}) = -\sum_l \frac{Z_l\, e^2}{|\boldsymbol{R}_l - \boldsymbol{x}|} \quad . \tag{8.103}$$

  The total energy is then

$$E_{\text{tot}}\big[\{\boldsymbol{R}_l\}, n(\boldsymbol{x})\big] = E_{\text{jel}}[n] + \int d^3x\, n(\boldsymbol{x})\, v_{\text{ext}}(\boldsymbol{x}) + \sum_{l<l'} \frac{Z_l\, Z_{l'}\, e^2}{|\boldsymbol{R}_l - \boldsymbol{R}_{l'}|} \quad . \tag{8.104}$$

  For a given set of nuclear positions $\{\boldsymbol{R}_l\}$, the energy functional is minimized with respect to the density $n(\boldsymbol{x})$. The resulting energy is then a function of the nuclear positions, and is then minimized with respect to these variables, yielding the crystal structure. Typically one works with periodic boundary conditions and with as large a crystallite cell as one can computationally afford in order to approximate the thermodynamic limit.

- *Cohesive energy* : The difference between the minimum energy per unit cell in the crystalline state and the total atomic energy of each atom in the unit cell is called the *cohesive energy* $E_{\text{coh}} = E_{\text{crystal}} - E_{\text{atomic}}$. If $E_{\text{coh}} < 0$, then crystal formation is advantageous, and the difference is the crystalline *binding energy* per unit cell.

- *Elastic constants* : After the optimal crystalline structure is determined, by varying with respect to the nuclear positions one can obtain the elastic constants.

- *Phase diagram under pressure* : At $T = 0$ the Gibbs free energy $G = E - TS + pV$ is the enthalpy $H = E + pV$. Including the pV term in the energy, one can evaluate the $T = 0$ Gibbs free energy at any finite pressure.

8.4 Response Functions

8.4.1 Linear response theory

What can we do with $E_{\text{jel}}[n]$? For starters, we can compute response functions for the jellium system. If the uniform density for pure jellium is n_0, then upon introducing a potential $v_{\text{ext}}(\boldsymbol{x})$ we may write $n(\boldsymbol{x}) = n_0 + \delta n(\boldsymbol{x})$. Expanding the functional $E_{\text{jel}}[n]$ about $n = n_0$, we have that the total energy functional $E[n] = E_{\text{jel}}[n] + V[n]$, to second order in δn, is given by

$$E[n_0 + \delta n] = E_{\text{jel}}[n_0] + \frac{1}{2}\int d^3x \int d^3x'\, \frac{\delta E_{\text{jel}}[n]}{\delta n(\boldsymbol{x})\, \delta n(\boldsymbol{x}')}\bigg|_{n_0} \delta n(\boldsymbol{x})\, \delta n(\boldsymbol{x}') + \int d^3x\, \big(n_0 + \delta n(\boldsymbol{x})\big) v_{\text{ext}}(\boldsymbol{x}) \quad . \tag{8.105}$$

Note that the first functional variation $\delta E_{\rm jel}[n_0] = 0$ vanishes for $n = n_0$ by definition. Thus,

$$\left.\frac{\delta E[n]}{\delta n(\boldsymbol{x})}\right|_{n_0} = v_{\rm ext}(\boldsymbol{x}) + \int d^3x'\, \chi^{-1}(\boldsymbol{x},\boldsymbol{x}')\,\delta n(\boldsymbol{x}') \quad , \tag{8.106}$$

where[11]

$$\chi^{-1}(\boldsymbol{x},\boldsymbol{x}') \equiv \left.\frac{\delta^2 E_{\rm jel}[n]}{\delta n(\boldsymbol{x})\,\delta n(\boldsymbol{x}')}\right|_{n_0} \quad . \tag{8.107}$$

The function $\chi^{-1}(\boldsymbol{x},\boldsymbol{x}')$ is the inverse density susceptibility. The relation between χ^{-1} and χ is

$$\int d^3x'\, \chi^{-1}(\boldsymbol{x},\boldsymbol{x}')\,\chi(\boldsymbol{x}',\boldsymbol{x}'') = \delta(\boldsymbol{x}-\boldsymbol{x}'') \quad , \tag{8.108}$$

thus Eqn. 8.106 is equivalent to

$$\delta n(\boldsymbol{x}) = -\int d^3x'\, \chi(\boldsymbol{x},\boldsymbol{x}')\, v_{\rm ext}(\boldsymbol{x}') \quad . \tag{8.109}$$

The above formula is an example of *linear response*. Had we expanded $E[n_0+\delta n]$ to higher order in δn, we'd have obtained higher order terms on the RHS, arranged as a functional Taylor series in $v_{\rm ext}(\boldsymbol{x})$.

Since the jellium system is translationally invariant, we must have $\chi(\boldsymbol{x},\boldsymbol{x}') = \chi(\boldsymbol{x}-\boldsymbol{x}')$. We now define the Fourier transform $\hat\chi(\boldsymbol{q})$ as

$$\hat\chi(\boldsymbol{q}) \equiv \int d^dx\, \chi(\boldsymbol{x})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \quad , \tag{8.110}$$

where $d = 3$ in the present discussion. The FT of $\chi^{-1}(\boldsymbol{x}-\boldsymbol{x}')$ is $\hat\chi^{-1}(\boldsymbol{q}) = 1/\hat\chi(\boldsymbol{q})$. Thus, within linear response,

$$\delta\hat n(\boldsymbol{q}) = -\hat\chi(\boldsymbol{q})\,\hat v_{\rm ext}(\boldsymbol{q}) \quad . \tag{8.111}$$

We write, as above,

$$E[n] = T_{\rm S}[n] + V[n] + E_{\rm H}[n] + E_{\rm XC}[n] \quad . \tag{8.112}$$

We define the noninteracting susceptibility according to

$$\chi_0^{-1}(\boldsymbol{x},\boldsymbol{x}') = \left.\frac{\delta^2 T_{\rm S}[n]}{\delta n(\boldsymbol{x})\,\delta n(\boldsymbol{x}')}\right|_{n_0} = \frac{1}{V}\sum_{\boldsymbol{q}} \hat\chi_0^{-1}(\boldsymbol{q})\, e^{i\boldsymbol{q}\cdot(\boldsymbol{x}-\boldsymbol{x}')} \quad , \tag{8.113}$$

where, as we shall derive later,

$$\hat\chi_0(\boldsymbol{q}) = 2\int \frac{d^dk}{(2\pi)^d}\,\frac{f^0(\boldsymbol{k}+\boldsymbol{q}) - f^0(\boldsymbol{k})}{\varepsilon_0(\boldsymbol{k}) - \varepsilon_0(\boldsymbol{k}+\boldsymbol{q})} \quad , \tag{8.114}$$

[11] Note that this definition differs by a minus sign by that in ch. 15 of Girvin and Yang, *Modern Condensed Matter Physics*.

where $\varepsilon_0(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$ is the noninteracting dispersion, and

$$f^0(\boldsymbol{k}) = \frac{1}{\exp\left(\frac{\varepsilon_0(\boldsymbol{k})-\mu}{k_{\rm B}T}\right)+1} \tag{8.115}$$

is the Fermi distribution. At $T = 0$, we have $f^0(\boldsymbol{k}) = \Theta(k_{\rm F} - k)$. Performing the integral in $d = 3$ dimensions, we obtain

$$\hat{\chi}_0(\boldsymbol{q}, T=0) = g(\varepsilon_{\rm F})\, L(q/2k_{\rm F}) \quad , \tag{8.116}$$

where $g(\varepsilon_{\rm F}) = mk_{\rm F}/\pi^2\hbar^2$ is the DOS at the Fermi level, with $k_{\rm F}(n) = (3\pi^2 n)^{1/3}$, and where

$$L(x) = \frac{1}{2} + \frac{1-x^2}{4x}\,\ln\left|\frac{1+x}{1-x}\right| \tag{8.117}$$

is the *Lindhard function*. We may now write

$$\chi^{-1}(\boldsymbol{x},\boldsymbol{x}') = \chi_0^{-1}(\boldsymbol{x},\boldsymbol{x}') + \frac{e^2}{|\boldsymbol{x}-\boldsymbol{x}'|} + \chi_{\rm XC}^{-1}(\boldsymbol{x},\boldsymbol{x}') \tag{8.118}$$

where

$$\chi_{\rm XC}^{-1}(\boldsymbol{x},\boldsymbol{x}') = \left.\frac{\delta^2 E_{\rm XC}[n]}{\delta n(\boldsymbol{x})\,\delta n(\boldsymbol{x}')}\right|_{n_0} \quad . \tag{8.119}$$

Note that

$$\chi_{\rm H}^{-1}(\boldsymbol{x},\boldsymbol{x}') = \left.\frac{\delta^2 E_{\rm H}[n]}{\delta n(\boldsymbol{x})\,\delta n(\boldsymbol{x}')}\right|_{n_0} = \frac{e^2}{|\boldsymbol{x}-\boldsymbol{x}'|} \quad . \tag{8.120}$$

Assuming translational invariance, $\chi_a(\boldsymbol{x},\boldsymbol{x}') = \chi_a(\boldsymbol{x}-\boldsymbol{x}')$ for all labels a (*e.g.*, χ, χ_0, $\chi_{\rm H}$, $\chi_{\rm XC}$), and $\chi_a^{-1}(\boldsymbol{x},\boldsymbol{x}') = \chi_a^{-1}(\boldsymbol{x}-\boldsymbol{x}')$ as well. Taking the Fourier transforms, then, we obtain

$$\hat{\chi}^{-1}(\boldsymbol{q}) = \hat{\chi}_0^{-1}(\boldsymbol{q}) + \frac{4\pi e^2}{\boldsymbol{q}^2} + \hat{\chi}_{\rm XC}^{-1}(\boldsymbol{q}) \equiv \widehat{\Pi}^{-1}(\boldsymbol{q}) + \frac{4\pi e^2}{\boldsymbol{q}^2} \quad , \tag{8.121}$$

where the inverse of the *polarization function* $\widehat{\Pi}(\boldsymbol{q})$ is defined according to

$$\widehat{\Pi}^{-1}(\boldsymbol{q}) = \hat{\chi}_0^{-1}(\boldsymbol{q}) + \hat{\chi}_{\rm XC}^{-1}(\boldsymbol{q}) \quad . \tag{8.122}$$

8.4.2 Static screening

We conclude that in the presence of an external potential $v_{\rm ext}(\boldsymbol{x})$, there is to first order a density response $\delta\hat{n}(\boldsymbol{q}) = -\hat{\chi}(\boldsymbol{q})\,\hat{v}_{\rm ext}(\boldsymbol{q})$. The corresponding charge profile is then $\delta\hat{\varrho}(\boldsymbol{q}) = -e\,\delta\hat{n}(\boldsymbol{q})$.

Hence the potential is *screened*. Within linear response, this results in an effective screened potential

$$\begin{aligned} v_{\rm scr}(\boldsymbol{x}) &= v_{\rm ext}(\boldsymbol{x}) + \int d^3x' \, \frac{e^2}{|\boldsymbol{x}-\boldsymbol{x}'|} \, \delta n(\boldsymbol{x}') \\ \hat{v}_{\rm scr}(\boldsymbol{q}) &= \hat{v}_{\rm ext}(\boldsymbol{q}) - \frac{4\pi e^2}{\boldsymbol{q}^2} \, \hat{\chi}(\boldsymbol{q}) \, \hat{v}_{\rm ext}(\boldsymbol{q}) \equiv \frac{\hat{v}_{\rm ext}(\boldsymbol{q})}{\hat{\epsilon}(\boldsymbol{q})} \quad , \end{aligned} \tag{8.123}$$

where $\hat{\epsilon}(\boldsymbol{q})$ is the static (*i.e.* zero frequency) *dielectric constant*, given. by

$$\begin{aligned} \hat{\epsilon}^{-1}(\boldsymbol{q}) &= 1 - \frac{4\pi e^2}{\boldsymbol{q}^2} \, \hat{\chi}(\boldsymbol{q}) \\ &= 1 - \frac{4\pi e^2/\boldsymbol{q}^2}{\widehat{\Pi}^{-1}(\boldsymbol{q}) + 4\pi e^2/\boldsymbol{q}^2} = \frac{1}{1 + \frac{4\pi e^2}{\boldsymbol{q}^2}\,\widehat{\Pi}(\boldsymbol{q})} \quad . \end{aligned} \tag{8.124}$$

Thus,

$$\hat{\epsilon}(\boldsymbol{q}) = 1 + \frac{4\pi e^2}{\boldsymbol{q}^2} \, \widehat{\Pi}(\boldsymbol{q}) \quad . \tag{8.125}$$

If we ignore the $\boldsymbol{q}$-dependence and approximate $\widehat{\Pi}(\boldsymbol{q}) \approx \widehat{\Pi}(0) \equiv Q^2/4\pi e^2$, which defines a quantity Q with dimensions of inverse length, then

$$\hat{\epsilon}(\boldsymbol{q}) \approx 1 + \frac{Q^2}{\boldsymbol{q}^2} \tag{8.126}$$

and for $v_{\rm ext}(\boldsymbol{r}) = -Ze^2/r$, we have $\hat{v}_{\rm ext}(\boldsymbol{q}) = -4\pi Ze^2/\boldsymbol{q}^2$ and the FT of the screened potential is

$$\hat{v}_{\rm scr}(\boldsymbol{q}) = -\frac{4\pi Ze^2}{\boldsymbol{q}^2 + Q^2} \quad , \tag{8.127}$$

which in real space ($d = 3$) corresponds to a Yukawa potential,

$$v_{\rm scr}(\boldsymbol{r}) = -\frac{Ze^2 \exp(-Qr)}{r} \quad . \tag{8.128}$$

Thus, the screened potential is much weaker at long distances (exponentially so) than the bare $1/r$ Coulomb potential.

Note that the total number of electrons accumulated within linear response theory is

$$\begin{aligned} \delta N &= \int d^3x \, \delta n(\boldsymbol{x}) = -\lim_{\boldsymbol{q}\to 0} \, \hat{\chi}(\boldsymbol{q}) \, \hat{v}_{\rm ext}(\boldsymbol{q}) \\ &= \lim_{\boldsymbol{q}\to 0} \frac{Z}{1 + (\boldsymbol{q}^2/4\pi e^2)\,\widehat{\Pi}^{-1}(\boldsymbol{q})} \quad . \end{aligned} \tag{8.129}$$

Thus, provided $\boldsymbol{q}^2/\widehat{\Pi}(\boldsymbol{q})$ vanishes in the limit $\boldsymbol{q} \to 0$, we obtain *perfect screening* by an induced charge $Q = -e\,\delta N = -Ze$ of the charge $+Ze$ impurity.

We emphasize that throughout this section we are discussing only the linear response of the jellium system. To compute the linear response of a material like elemental Pb, say, we'd need to solve the KS equations and evaluate the various functional derivatives at a number density $n(\boldsymbol{x})$ which is the ground state electron density for Pb.

8.4.3 Approximate forms for the polarization function

The static dielectric function $\hat{\epsilon}(\boldsymbol{q})$ is given to us, in Eqn. 8.125, in terms of the unknown polarization function $\widehat{\Pi}(\boldsymbol{q})$. There are two common approximations we shall discuss here.

The first is the *Lindhard approximation*, in which we ignore $\hat{\chi}_{\rm XC}(\boldsymbol{q})$ and write $\widehat{\Pi}(\boldsymbol{q}) \approx \widehat{\Pi}_{\rm L}(\boldsymbol{q})$ where $\widehat{\Pi}_{\rm L}(\boldsymbol{q}) = \hat{\chi}_0(\boldsymbol{q})$, which is given in Eqns. 8.116 and 8.117 above. Thus

$$\hat{\chi}(\boldsymbol{q}) \approx \hat{\chi}_{\rm L}(\boldsymbol{q}) \equiv \frac{\hat{\chi}_0(\boldsymbol{q})}{1 + \frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat{\chi}_0(\boldsymbol{q})} \quad , \qquad \hat{\epsilon}(\boldsymbol{q}) \approx \hat{\epsilon}_{\rm L}(\boldsymbol{q}) \equiv 1 + \frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat{\chi}_0(\boldsymbol{q}) \quad . \tag{8.130}$$

In the $\boldsymbol{q} \to 0$ limit, $\widehat{\Pi}_{\rm L}(\boldsymbol{q}) = g(\varepsilon_{\rm F}) + \mathcal{O}(\boldsymbol{q}^2)$, which entails perfect screening of a Coulomb impurity, *i.e.* $\delta N = Z$. However, rather than obtaining a Yukawa form for the screened potential, one instead finds

$$v_{\rm scr}(\boldsymbol{r}) \propto \frac{\cos(2k_{\rm F} r)}{r^3} \tag{8.131}$$

in the long distance limit. This arises from the logarithmic singularity in the Lindhard function $L(q/2k_{\rm F})$ (Eqn. 8.117) at $q = 2k_{\rm F}$, which is a feature of the sharp Fermi surface.

A further simplification is the *Thomas-Fermi* (new acronym: TF) approximation, in which we also ignore the $\boldsymbol{q}$-dependence and write $\widehat{\Pi}(\boldsymbol{q}) \approx \widehat{\Pi}_{\rm L}(0) = g(\varepsilon_{\rm F}) \equiv Q_{\rm TF}^2/4\pi e^2$, where $Q_{\rm TF} = \sqrt{4\pi e^2 g(\varepsilon_{\rm F})}$ is the TF wavevector. Thus $\hat{\epsilon}_{\rm TF}(\boldsymbol{q}) = 1 + \frac{Q_{\rm TF}^2}{\boldsymbol{q}^2}$ and the screened potential is of the Yukawa form. For a quick and dirty way to derive TF theory, assume that the electric potential $\phi(\boldsymbol{x})$ varies slowly in space, and imagine locally shifting the Fermi energy by the local electrostatic energy, *i.e.* from $\varepsilon_{\rm F}$ to $\varepsilon_{\rm F} + e\,\phi(\boldsymbol{x})$. This results in a local density accumulation $\delta n(\boldsymbol{x}) = e\,\phi(\boldsymbol{x})\,g(\varepsilon_{\rm F})$, and invoking Poisson's equation,

$$\nabla^2\phi = -4\pi\varrho = 4\pi e\,\delta n = 4\pi e^2 g(\varepsilon_{\rm F})\,\phi = Q_{\rm TF}^2\,\phi \quad , \tag{8.132}$$

whence the Yukawa potential. By ignoring the $\boldsymbol{q}$-dependence, we have missed the Fermi surface singularity which is present in the (more realistic) Lindhard approximation. Note that the TF wavelength, $\lambda_{\rm TF} = Q_{\rm TF}^{-1}$, is given by

$$\lambda_{\rm TF} = \left(\frac{\pi}{12}\right)^{1/6} \sqrt{r_{\rm s}}\,a_{\rm B} \approx 0.800\sqrt{r_{\rm s}}\,a_{\rm B} \quad . \tag{8.133}$$

Recall that $\frac{4}{3}\pi(r_{\rm s}\,a_{\rm B})^3 n \equiv 1$ defines $r_{\rm s}$, hence $r_{\rm s} \propto n^{-1/3}$. TF theory is statistical and can only be justified if there are a large number of electrons within a sphere of radius $\lambda_{\rm TF}$, which says $r_{\rm s} \lesssim (\pi/12)^{1/3} \simeq 0.640$.

There is another kind of screening in solids which is relevant when the temperature is much larger than the Fermi energy. This is called Debye-Hückel screening and the argument goes as follows. Let the background charge density be $\varrho_0 = e\, n_0$. Classical statistical physics then yields a local electron density $n(\boldsymbol{x}) = n_0 \exp\bigl[e\phi(\boldsymbol{x})/k_{\rm B}T\bigr]$, and invoking Poisson results in the equation $\nabla^2\phi = Q_{\sf DH}^2\,\phi$, where $Q_{\sf DH} = \sqrt{k_{\rm B}T/4\pi n_0 e^2}$.

Chapter 9

Landau Fermi Liquid Theory

9.1 Normal ^{3}He Liquid

^{3}He is a neutral atom consisting of two protons, one neutron, and two electrons. A composite of five fermions, it behaves as a hard-sphere (radius $a \approx 1.35$Å) fermion of (nuclear) spin $I = \frac{1}{2}$ at energies below the scale of electronic transitions[1]. It exhibits a fairly rich phase diagram, depicted in the left hand panel of Fig. 9.1. ^{3}He A and ^{3}He B are superfluid phases which differ in the symmetry of their respective order parameters. ^{3}He N is a normal fluid which behaves much like a free Fermi gas, but in which interaction effects play an essential role in its physical properties. It is known as a *Fermi liquid*[2] In a Fermi liquid, as in the noninteracting Fermi gas, the low-temperature specific heat $c_V(T)$ is linear in T and the magnetic susceptibility $\chi(T)$ is Pauli-like ($\chi \propto T^0$), as shown in Fig. 9.2. An important distinction between ^{3}He N and most metals is that the mass of the ^{3}He atom is about 6,000 times greater than that of the electron. Thus at a typical density $n = 1.64 \times 10^{22}\,\mathrm{cm}^{-3}$ and $m_3 = 5.01 \times 10^{-24}\,\mathrm{g}$ one obtains a Fermi temperature

$$T_{\mathsf{F}} = \frac{\hbar^2}{2mk_{\mathrm{B}}}\,(3\pi^2 n)^{2/3} = 4.97\,\mathrm{K} \quad , \tag{9.1}$$

which is much much smaller than $T_{\mathsf{F}}(\mathrm{Cu}) \approx 81,000\,\mathrm{K}$ and $T_{\mathsf{F}}(\mathrm{Al}) \approx 135,000\,\mathrm{K}$. This explains why one begins to see Curie-like behavior in the magnetic susceptibility, *i.e.* $\chi(T) \simeq n\mu_0^2/k_{\mathrm{B}}T$, at temperatures $T \gtrsim 1\,\mathrm{K}$. Here $\mu_0 = -10.746 \times 10^{-27}\,\mathrm{J/T} = -1.1574\,\mu_{\mathrm{B}}$ is the ^{3}He nuclear magnetic moment, and $\mu_{\mathrm{B}} = e\hbar/2m_{\mathrm{e}}c$ is the Bohr magneton, with m_{e} the electron mass. Recall these basic

[1] $E_1 - E_0 \approx 20\,\mathrm{eV}$, and the first ionization energy is $24.6\,\mathrm{eV}$.
[2] The general theory of Fermi liquids was developed principally by the Russian physicist Lev Landau in the 1950s.

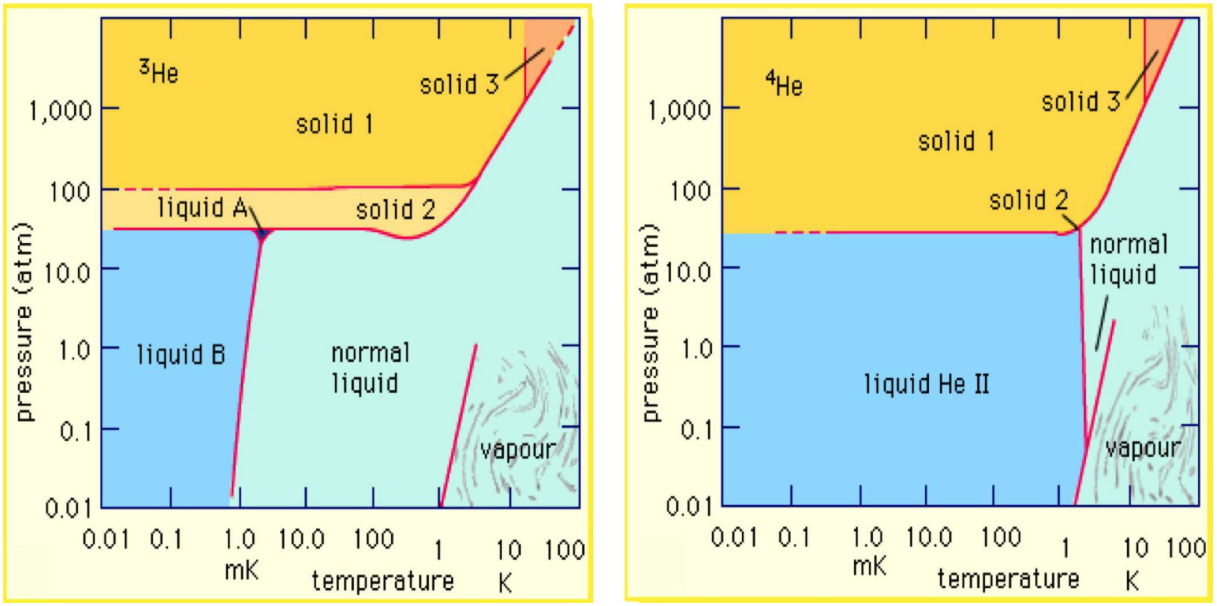


Figure 9.1: Phase diagrams of ^{3}He (left) and ^{4}He (right).

results for the free spin-$\frac{1}{2}$ Fermi gas with ballistic dispersion $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$:

$$\begin{aligned}
\text{Fermi wavevector} \;&:\; k_{\text{F}} = (3\pi^2 n)^{1/3} \\
\text{density of states} \;&:\; g(\varepsilon_{\text{F}}) = \frac{m k_{\text{F}}}{\pi^2\hbar^2} \\
\text{occupancy} \;&:\; f(\varepsilon) = \left[\exp\left(\frac{\varepsilon-\mu}{k_{\text{B}}T}\right) + 1\right]^{-1} \\
\text{specific heat} \;&:\; c_V = \frac{1}{V}\left(\frac{\partial E}{\partial T}\right)_{N,V} = \frac{\pi^2}{3}\, g(\varepsilon_{\text{F}})\, k_{\text{B}}^2 T + \mathcal{O}(T^3) \\
\text{magnetic susceptibility} \;&:\; \chi = \left(\frac{\partial M}{\partial H}\right)_{N,V} = \mu_0^2\, g(\varepsilon_{\text{F}}) + \mathcal{O}(T^2) \\
\text{compressibility} \;&:\; \kappa = n^{-2}\left(\frac{\partial n}{\partial \mu}\right)_T = n^{-2} g(\varepsilon_{\text{F}}) + \mathcal{O}(T^2) \quad .
\end{aligned} \tag{9.2}$$

Experimental data for $c_V(T)$ and $\chi(T)$ in ^{3}He are shown in Fig. 9.2. Note that $c_V(T)/T$ and $\chi(T)$ are each pressure-dependent constants as $T \to 0$. The same is true for the compressibility $\kappa(T)$, which is obtained from measurements of the velocity of thermodynamic sound, $s = (m_3 n\kappa)^{-1/2}$. In a noninteracting Fermi gas, all these quantities are proportional to the density of states $g(\varepsilon_{\text{F}})$, up to constant factors. We can define $c_V^0(T,n)$, $\chi^0(T,n)$, and $\kappa^0(T,n)$ to be the corresponding free Fermi gas expressions for a system of spin-$\frac{1}{2}$ fermions of mass m_3 and density n. One finds that the ratios c_V/c_V^0, χ/χ^0, and κ/κ^0 all tend to different constants as $T \to 0$. Thus, it is impossible to reconcile the data by positing a phenomenological effective

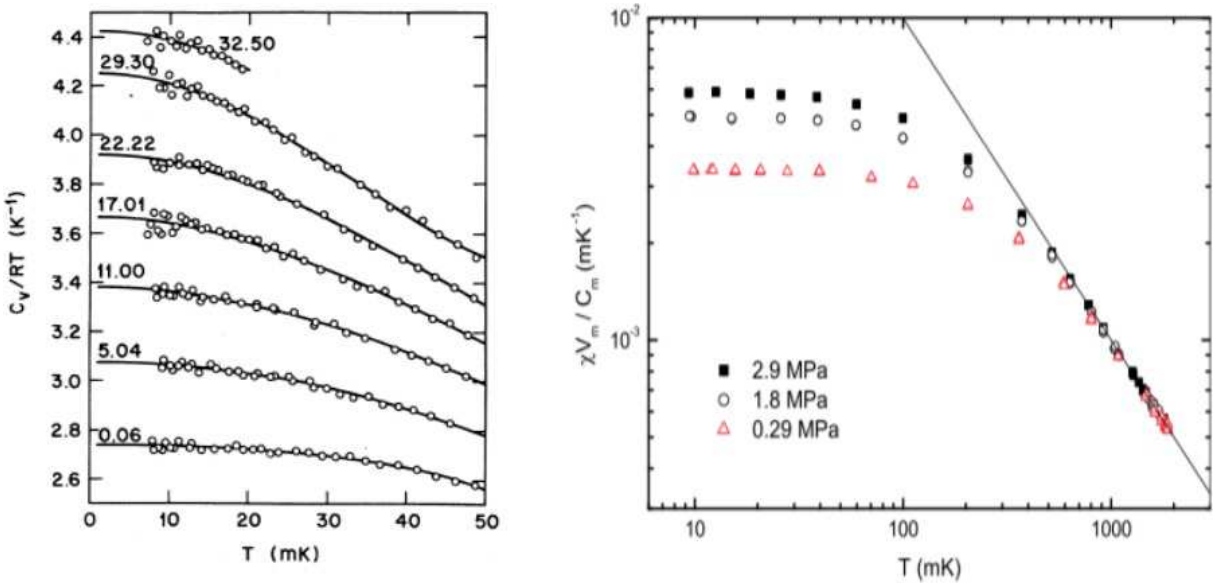


Figure 9.2: Left: $c_V(T)/RT$ for normal ^{3}He. From D. S. Greywall, *Phys. Rev. B* **27**, 2747 (1983). Numbers give the sample pressures in bars at $T = 0.1\,\mathrm{K}$. Right: Normalized magnetic susceptibility $\chi(T)\,v_0/C_\mathrm{m}$ of normal ^{3}He, where v_0 is the molar volume and $C_\mathrm{m} \equiv \lim_{T\to\infty} T\chi(T)$ is the Curie constant. From V. Goudon *et al.*, *J. Phys.: Conf. Ser.* **150**, 032024 (2009).

mass m^*, since that would require that these ratios all tend to the same value. Furthermore, the $T \to 0$ limits of these ratios are all pressure-dependent. Another issue is that the first correction to the low temperature linear specific heat in a Fermi gas go as T^3, whereas experiments yield a co.rrection on the order of $T^3 \ln T$ We shall see below how Landau's theory is capable of reproducing the observed temperature dependences, and moreover introduces additional interaction parameters which allow us to describe all these behaviors in a consistent way. We shall largely follow here the treatments by Nozieres and Pines, and by Baym and Pethick[3].

9.2 Fermi Liquid Theory : Statics and Thermodynamics

9.2.1 Adiabatic continuity

The idea behind Fermi liquid theory is that the many-body eigenstates of the free Fermi gas with Hamiltonian $\hat{H}_0$, which are Slater determinants, each evolve adiabatically into eigenstates of the interacting Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}_1$, where $\hat{H}_1$ is the interaction part. Typically we will

[3]P. Nozieres and D. Pines, *Theory of Quantum Liquids* (Avalon, 1999); G. Baym and C. Pethick, *Landau Fermi-Liquid Theory : Concepts and Applications* (Wiley, 1991).

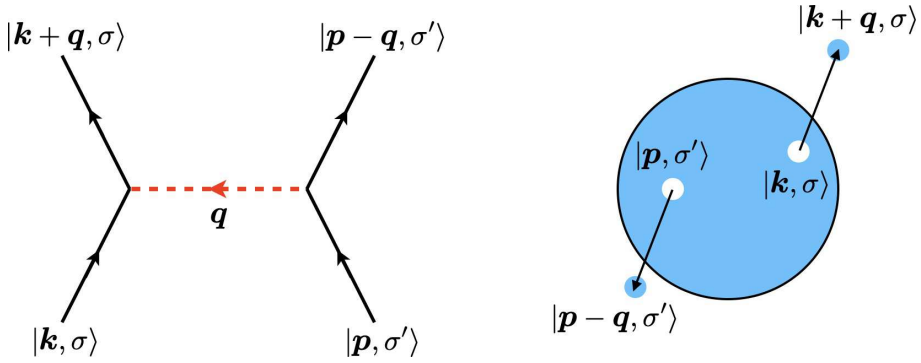


Figure 9.3: Two particle, two hole excitation of the state $|\mathrm{F}\rangle$ obtained via first order perturbation theory in the interaction Hamiltonian $\hat{H}_1$.

consider

$$\hat{H}_0 = \sum_{\boldsymbol{k},\sigma} \varepsilon^0_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k},\sigma} c_{\boldsymbol{k},\sigma} \quad , \tag{9.3}$$

with $\varepsilon^0_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m$. The general form of interactions in a translationally invariant system is

$$\hat{H}_1 = \frac{1}{2}\sum_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}} \sum_{\alpha,\beta} \sum_{\alpha',\beta'} \hat{u}_{\alpha\beta\alpha'\beta'}(\boldsymbol{q})\, c^\dagger_{\boldsymbol{k}+\boldsymbol{q},\alpha}\, c^\dagger_{\boldsymbol{p}-\boldsymbol{q},\alpha'}\, c_{\boldsymbol{p},\beta'}\, c_{\boldsymbol{k},\beta} \quad . \tag{9.4}$$

In systems with spin isotropy, we can write

$$\hat{u}_{\alpha\beta\alpha'\beta'}(\boldsymbol{q}) = \hat{u}^{\mathsf{S}}(\boldsymbol{q})\,\delta_{\alpha\beta}\,\delta_{\alpha'\beta'} + \hat{u}^{\mathsf{H}}(\boldsymbol{q})\,\boldsymbol{\tau}_{\alpha\beta}\cdot\boldsymbol{\tau}_{\alpha'\beta'} \quad , \tag{9.5}$$

where $\hat{u}^{\mathsf{S,H}}(\boldsymbol{q})$ are the scalar and Heisenberg exchange parts of the interaction, respectively.

We will focus here on the case where $\hat{u}^{\mathsf{H}} = 0$, in which case we may write

$$\hat{H}_1 = \frac{1}{2}\sum_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}} \sum_{\sigma,\sigma'} \hat{u}(\boldsymbol{q})\, c^\dagger_{\boldsymbol{k}+\boldsymbol{q},\sigma}\, c^\dagger_{\boldsymbol{p}-\boldsymbol{q},\sigma'}\, c_{\boldsymbol{p},\sigma'}\, c_{\boldsymbol{k},\sigma} \quad . \tag{9.6}$$

When $\hat{H}_1 = 0$, the N-particle ground state is the filled Fermi sphere, $|\,\mathrm{F}\,\rangle = \prod'_{\boldsymbol{k},\sigma} c^\dagger_{\boldsymbol{k},\sigma}\,|\,0\,\rangle$, where the prime denotes the restriction $|\boldsymbol{k}| \leqslant k_{\mathsf{F}}$. Treating the interaction in first order perturbation theory, we have the perturbed ground state $|\,\mathrm{F}'\,\rangle$ is given by

$$|\,\mathrm{F}'\,\rangle = |\,\mathrm{F}\,\rangle + \sum_\alpha \frac{|\,\alpha\,\rangle\langle\,\alpha\,|\,\hat{H}_1\,|\,\mathrm{F}\,\rangle}{E^0_{\mathrm{F}} - E^0_\alpha} + \mathcal{O}(\hat{H}_1^2) \quad . \tag{9.7}$$

This results in contributions such as that depicted in Fig. 9.3. Proceeding to still higher orders of perturbation theory, the perturbed ground state appears as a seething, bubbling 'soup' of particle-hole pairs.

We can associate interacting and noninteracting eigenstates, however, through the process of adiabatic evolution. Define

$$\hat{H}(\lambda) = \hat{H}_0 + \lambda \hat{H}_1 \quad , \tag{9.8}$$

so $\hat{H}(0) = \hat{H}_0$ and $\hat{H}(1) = \hat{H}_0 + \hat{H}_1 = \hat{H}$. Suppose $\lambda(t)$ is a monotonically increasing function of t for $t < 0$, with $\lambda(-\infty) = 0$ and $\lambda(0) = 1$. The unitary evolution operator is then

$$\begin{aligned}\hat{U}(0,-\infty) &= \mathcal{T} \exp\left\{ -\frac{i}{\hbar} \int_{-\infty}^{0} dt\, \hat{H}(t) \right\} \\ &= \mathcal{T} \exp\left\{ -\frac{i}{\hbar} \int_0^1 \frac{d\lambda}{\dot{\lambda}}\, \hat{H}(\lambda) \right\} = \mathcal{T} \exp\left\{ -\frac{i}{\hbar\epsilon} \int_0^1 \frac{d\lambda}{\lambda}\, \hat{H}(\lambda) \right\} \equiv \hat{U}_\epsilon \quad ,\end{aligned} \tag{9.9}$$

where in the final expression we take $\lambda(t) = \exp(-\epsilon|t|)$. Thus, we can consider the adiabatic map,

$$\hat{U}_\epsilon \,:\, |\,\mathrm{F}\,\rangle \to |\,\mathrm{F}'_\epsilon\,\rangle = \hat{U}_\epsilon\,|\,\mathrm{F}\,\rangle \tag{9.10}$$

where $\hat{H}\,|\,\mathrm{F}'\,\rangle = E'\,|\,\mathrm{F}'\,\rangle$. We then consider the limit as $\epsilon \to 0$. One wrinkle here is that the phase of $|\,\mathrm{F}'_\epsilon\,\rangle$ in the limit $\epsilon \to 0$ is generally divergent, and to cancel it out we could instead define the state

$$|\,\widetilde{\mathrm{F}}'\,\rangle \equiv \lim_{\epsilon\to 0} \left\{ \left(\frac{\langle\,\mathrm{F}\,|\,U_\epsilon^\dagger\,|\,\mathrm{F}\,\rangle}{\langle\,\mathrm{F}\,|\,U_\epsilon\,|\,\mathrm{F}\,\rangle} \right)^{1/2} \hat{U}_\epsilon\,|\,\mathrm{F}\,\rangle \right\} \quad , \tag{9.11}$$

in which the phase cancels.

Suppose that rather starting with the N-particle state $|\,\mathrm{F}\,\rangle$, we start with the state $c^\dagger_{\boldsymbol{k},\sigma}\,|\,\mathrm{F}\,\rangle$, where $|\boldsymbol{k}| > k_{\mathrm{F}}$. We then adiabatically evolve with $\hat{U}_\epsilon$ as described above (including our nifty phase divergence cancellation protocol). We then obtain a state $|\,\Psi_{\boldsymbol{k},\sigma}\,\rangle$, about which we know three things: (i) its total particle number is $N+1$, (ii) its total momentum is $\hbar\boldsymbol{k}$, and (iii) its total spin polarization is σ. We may write

$$|\,\Psi'_{\boldsymbol{k},\sigma}\,\rangle = q^\dagger_{\boldsymbol{k},\sigma}\,|\,\mathrm{F}'\,\rangle \quad , \tag{9.12}$$

where

$$\begin{aligned} q^\dagger_{\boldsymbol{k},\sigma} &= \lim_{\epsilon\to 0} \left\{ U_\epsilon\, c^\dagger_{\boldsymbol{k},\sigma}\, U_\epsilon^\dagger \right\} \\ &= Z_{\boldsymbol{k},\sigma}\, c^\dagger_{\boldsymbol{k},\sigma} + \sum_{\boldsymbol{k}_1,\boldsymbol{k}_2} \sum_{\sigma_1,\sigma_2} A^{\sigma_1,\sigma_2}_{\boldsymbol{k}_1,\boldsymbol{k}_2}\, c^\dagger_{\boldsymbol{k}_1,\sigma_1} c^\dagger_{\boldsymbol{k}_2,\sigma_2} c_{\boldsymbol{k}_1+\boldsymbol{k}_2-\boldsymbol{k},\sigma_1+\sigma_2-\sigma} + \dots \quad .\end{aligned} \tag{9.13}$$

Thus, the operator which *when acting on the interacting ground state* $|\,\mathrm{F}'\,\rangle$ creates the excited state $|\,\Psi_{\boldsymbol{k},\sigma}\,\rangle$ is a complicated linear combination of products of creation and annihilation operators where each term has fixed total particle number, momentum, and spin polarization. We say that $q^\dagger_{\boldsymbol{k},\sigma}$ creates a *quasiparticle* of momentum $\hbar\boldsymbol{k}$ and spin polarization σ. The factor $Z_{\boldsymbol{k},\sigma}$ is called the *quasiparticle weight* (typically independent of σ in unmagnetized systems) and tells

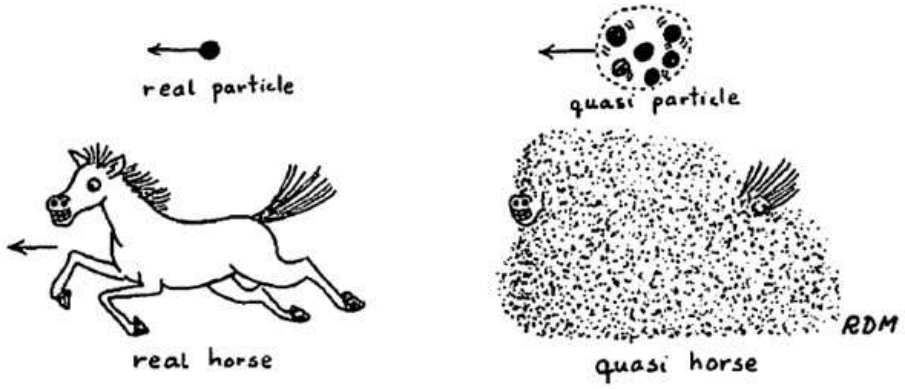


Figure 9.4: A quasi-particle is to a real particle as a quasi-horse is to a real horse. From R. D. Mattuck, *A Guide to Feynman Diagrams in the Many-Body Problem* (Dover, 1992).

us what fraction of the quasiparticle content is the single bare fermion $c^\dagger_{\boldsymbol{k},\sigma}$. The rest is what we in the many-body biz call *dressing*. The bare particle, or what's left of it, is surrounded by a cloud of particle-hole pairs in various combinations. See Fig. 9.4 for a vivid analogy.

Now imagine starting with a general Fock basis state,

$$\big|\,\Psi_0\big[\{N_{\boldsymbol{k},\sigma}\}\big]\,\big\rangle = \prod_{\boldsymbol{k},\sigma} \big(c^\dagger_{\boldsymbol{k},\sigma}\big)^{N_{\boldsymbol{k},\sigma}} \big|\,0\,\big\rangle \quad , \tag{9.14}$$

which is an eigenstate of $\hat{H}_0$ with eigenvalue $E^0\big[\{N_{\boldsymbol{k},\sigma}\}\big] = \sum_{\boldsymbol{k},\sigma} N_{\boldsymbol{k},\sigma}\,\varepsilon^0_{\boldsymbol{k},\sigma}$. We then perform our adiabatic evolution, which generates the interacting eigenstate $\big|\Psi\big[\{N_{\boldsymbol{k},\sigma}\}\big]\big\rangle$, which must be an eigenstate of $\hat{H} = \hat{H}_0 + \hat{H}_1$. Its associated eigenvalue E must then be a function, however complicated, of the set $\{N_{\boldsymbol{k},\sigma}\}$, *i.e.* $E = E\big[\{N_{\boldsymbol{k},\sigma}\}\big]$. Since we can adiabatically evolve any many-body eigenstate of $\hat{H}_0$, we can also evolve a *density matrix* of the form

$$\varrho_0\big[\{n_{\boldsymbol{k},\sigma}\}\big] = \bigotimes_{\boldsymbol{k},\sigma} \Big[\big(1-n_{\boldsymbol{k},\sigma}\big)\,|\,0\,\rangle\langle\,0\,| + n_{\boldsymbol{k},\sigma}\,c^\dagger_{\boldsymbol{k},\sigma}\,|\,0\,\rangle\langle\,0\,|\,c_{\boldsymbol{k},\sigma}\Big] \tag{9.15}$$

Here we may take the distribution $\{n_{\boldsymbol{k},\sigma}\}$ to be smooth as a function of $\boldsymbol{k}$ for each σ, and regard the energy to be a function (or functional[4]) of the distributions $\{n_{\boldsymbol{k},\sigma}\}$.

It is important to note that the principle of adiabatic continuity can easily fail, for example when a phase boundary is crossed as λ evolves over the interval $\lambda \in [0,1]$. This is indeed the case for phases of matter such as charge and spin density waves, exciton condensates, superconductors, *etc.*

[4]If we regard $\boldsymbol{k}$ as a continuous variable, then $E[\{n_{\boldsymbol{k},\sigma}\}]$ is a functional of the functions $n_{\boldsymbol{k},\uparrow}$ and $n_{\boldsymbol{k},\downarrow}$.

9.2.2 First law of thermodynamics for Fermi liquids

We begin with the formula for the entropy of a distribution of fermions,

$$\begin{aligned} S[\{n_{\boldsymbol{k},\sigma}\}] &= -k_{\rm B}\,\mathsf{Tr}\,(\varrho_0 \ln \varrho_0) \\ &= -k_{\rm B} \sum_{\boldsymbol{k},\sigma} \Big\{ n_{\boldsymbol{k},\sigma} \ln n_{\boldsymbol{k},\sigma} + (1-n_{\boldsymbol{k},\sigma}) \ln(1-n_{\boldsymbol{k},\sigma}) \Big\} \quad . \end{aligned} \tag{9.16}$$

Note that the entropy does not change under adiabatic evolution of the density matrix. The first variation of the entropy is then

$$\delta S = -k_{\rm B} \sum_{\boldsymbol{k},\sigma} \ln\left(\frac{n_{\boldsymbol{k},\sigma}}{1-n_{\boldsymbol{k},\sigma}}\right) \delta n_{\boldsymbol{k},\sigma} \quad . \tag{9.17}$$

The total particle number operator is $\hat{N} = \sum_{\boldsymbol{k},\sigma} \hat{n}_{\boldsymbol{k},\sigma}$, hence

$$N = \mathsf{Tr}\left(\varrho_0\,\hat{N}\right) = \sum_{\boldsymbol{k},\sigma} n_{\boldsymbol{k},\sigma} \qquad , \qquad \delta N = \sum_{\boldsymbol{k},\sigma} \delta n_{\boldsymbol{k},\sigma} \quad . \tag{9.18}$$

Note that the particle number, like the entropy, is preserved by adiabatic evolution.

Finally, the energy E, as discussed in the previous section, is a functional of the distribution, which means that we may write

$$\delta E = \sum_{\boldsymbol{k},\sigma} \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k},\sigma} \qquad , \qquad \widetilde{\varepsilon}_{\boldsymbol{k},\sigma} = \frac{\delta E}{\delta n_{\boldsymbol{k},\sigma}} \tag{9.19}$$

is the first functional variation of E. The energy is *not* an adiabatic invariant. It is crucial to note that $\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}$ is simultaneously a function of $\boldsymbol{k}$ and σ *and* a functional of the distribution. Indeed, we shall write

$$\frac{\delta^2 E}{\delta n_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k}',\sigma'}} = \frac{\delta \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}}{\delta n_{\boldsymbol{k}',\sigma'}} \equiv \frac{1}{V}\,\widetilde{f}_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} \quad , \tag{9.20}$$

where $\widetilde{f}_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}$ has dimensions of energy $\times$ volume and is itself, in principle, a functional of the distribution.

Writing the First Law as

$$T\,\delta S = \delta E - \mu\,\delta N \quad , \tag{9.21}$$

and using the fact that the $\delta n_{\boldsymbol{k},\sigma}$ are all independent variations, we have

$$-k_{\rm B}T\,\ln\left(\frac{n_{\boldsymbol{k},\sigma}}{1-n_{\boldsymbol{k},\sigma}}\right) = \widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu \quad , \tag{9.22}$$

for each $(\boldsymbol{k},\sigma)$, which is equivalent to

$$n_{\boldsymbol{k},\sigma} = \frac{1}{\exp\left(\frac{\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}-\mu}{k_{\rm B}T}\right)+1} \quad . \tag{9.23}$$

This has the innocent appearance of the Fermi distribution familiar from elementary quantum statistical physics, but it must be emphasized again that $\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}$ is a functional of the distribution, hence Eqn. 9.23 is in fact a complicated implicit, nonlinear equation for the individual occupations $n_{\boldsymbol{k},\sigma}$.

At $T=0$, however, we have

$$n_{\boldsymbol{k},\sigma}(T=0) = \Theta(\mu - \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}) \equiv n^0_{\boldsymbol{k},\sigma} \quad . \tag{9.24}$$

It is now convenient to define the deviation

$$\delta n_{\boldsymbol{k},\sigma} \equiv n_{\boldsymbol{k},\sigma} - n^0_{\boldsymbol{k},\sigma} \quad , \tag{9.25}$$

where $n^0_{\boldsymbol{k},\sigma}$ is the ground state distribution at $T=0$. In an isotropic system with no external magnetic field, we have $n^0_{\boldsymbol{k},\sigma} = \Theta(k_{\rm F}-k)$. We may now write the energy E as a functional of the $\delta n_{\boldsymbol{k},\sigma}$, *viz.*

$$E = E_0 + \sum_{\boldsymbol{k},\sigma} \varepsilon_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k},\sigma} + \frac{1}{2V}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta n_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k}',\sigma'} + \ldots \quad . \tag{9.26}$$

Though it may not be obvious at this stage, it turns out that this is as far as we need to go in the expansion of the energy as a functional Taylor series in the $\delta n_{\boldsymbol{k},\sigma}$. Note that

$$\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} = \frac{\delta E}{\delta n_{\boldsymbol{k},\sigma}} = \varepsilon_{\boldsymbol{k},\sigma} + \frac{1}{V}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta n_{\boldsymbol{k}',\sigma'} + \ldots \tag{9.27}$$

and thus

$$\varepsilon_{\boldsymbol{k},\sigma} = \left.\frac{\delta E}{\delta n_{\boldsymbol{k},\sigma}}\right|_{\delta n=0} \quad . \tag{9.28}$$

Similarly,

$$\left.\frac{\delta^2 E}{\delta n_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k}',\sigma'}}\right|_{\delta n=0} = \left.\frac{\delta\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}}{\delta n_{\boldsymbol{k}',\sigma'}}\right|_{\delta n=0} \equiv \frac{1}{V}\,f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} \quad . \tag{9.29}$$

Compare with Eqn. 9.20. In isotropic systems, the Fermi velocity is given by

$$\frac{1}{\hbar}\left.\frac{\partial\varepsilon_{\boldsymbol{k},\sigma}}{\partial\boldsymbol{k}}\right|_{k=k_{\rm F}} = v_{\rm F}\,\hat{\boldsymbol{k}} \quad , \tag{9.30}$$

and we define the *effective mass* m^* by the relation $v_{\rm F} = \hbar k_{\rm F}/m^*$. The Fermi energy is then given by $\varepsilon_{\rm F} = \varepsilon_{\boldsymbol{k},\sigma}\big|_{k=k_{\rm F}}$, and the density of states at the Fermi energy is

$$g(\varepsilon_{\rm F}) = \sum_\sigma \int \frac{d^3k}{(2\pi)^3}\,\delta(\varepsilon_{\rm F} - \varepsilon_{\boldsymbol{k},\sigma}) = \frac{m^* k_{\rm F}}{\pi^2\hbar^2} \quad , \tag{9.31}$$

where, recall, $k_{\rm F} = (3\pi^2 n)^{1/3}$.

In systems with spin isotropy, we may define the functions $f^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'}$ and $f^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'}$ as follows:

$$\begin{aligned} f_{\boldsymbol{k}\uparrow,\boldsymbol{k}'\uparrow} &= f_{\boldsymbol{k}\downarrow,\boldsymbol{k}'\downarrow} \equiv f^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'} + f^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'} \\ f_{\boldsymbol{k}\uparrow,\boldsymbol{k}'\downarrow} &= f_{\boldsymbol{k}\downarrow,\boldsymbol{k}'\uparrow} \equiv f^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'} - f^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'} \quad . \end{aligned} \tag{9.32}$$

Equivalently,

$$f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} = f^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'} + \sigma\sigma' f^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'} \quad . \tag{9.33}$$

Recall that $f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}$ has dimensions of energy $\times$ volume. Thus we may define the dimensionless function $F_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}$ by multiplying $f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}$ by the density of states $g(\varepsilon_{\rm F})$:

$$F_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} \equiv g(\varepsilon_{\rm F})\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} \qquad , \qquad F^{\rm s,a}_{\boldsymbol{k},\boldsymbol{k}'} \equiv g(\varepsilon_{\rm F})\, f^{\rm s,a}_{\boldsymbol{k},\boldsymbol{k}'} \quad , \tag{9.34}$$

with $F_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} = F^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'} + \sigma\sigma' F^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'}$. When $\boldsymbol{k}$ and $\boldsymbol{k}'$ both lie on the Fermi surface, we may write

$$F^{\rm s,a}_{k_{\rm F}\hat{\boldsymbol{k}},k_{\rm F}\hat{\boldsymbol{k}}'} \equiv F^{\rm s,a}(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'}) \quad , \tag{9.35}$$

where $\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}' = \cos\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'}$. Furthermore, we may expand $F^{\rm s,a}(\vartheta)$ in terms of the Legendre polynomials, *viz.*

$$F^{\rm s,a}(\vartheta) = \sum_{n=0}^{\infty} F^{\rm s,a}_n\, P_n(\cos\vartheta) \quad . \tag{9.36}$$

Recall the generating function for the Legendre polynomials,

$$(1 - 2xt + t^2)^{-1/2} = \sum_{n=0}^{\infty} t^n\, P_n(x) \quad , \tag{9.37}$$

as well as the recurrence relation

$$P_{n+1}(x) = \frac{2n+1}{n+1}\,x\,P_n(x) - \frac{n}{n+1}\,P_{n-1}(x) \quad , \tag{9.38}$$

and the orthogonality relation

$$\int_{-1}^{1} dx\, P_m(x)\, P_n(x) = \frac{2}{2n+1}\,\delta_{mn} \quad . \tag{9.39}$$

Therefore if $F(\vartheta) = \sum_\ell F_\ell\, P_\ell(\vartheta)$ then

$$\int \frac{d\Omega}{4\pi}\, F(\vartheta)\, P_n(\cos\vartheta) = \frac{F_n}{2n+1} \quad , \tag{9.40}$$

where $d\Omega$ is the differential solid angle.

parameter	$p = 0$ bar	$p = 27$ bar
m^*/m	2.80	5.17
$F_0^{\rm s}$	9.28	68.17
$F_1^{\rm s}$	5.39	12.79
$F_0^{\rm a}$	-0.696	-0.760
$(F_1^{\rm a})^*$	-0.54	-1.00
$(F_1^{\rm a})^*$	-0.46	-0.27
$v_{\rm F}$ (cm/sec)	5.90×10^3	3.57×10^3
c_1 (cm/sec)	1.829×10^4	3.893×10^4

Table 9.1: Fermi liquid parameters for ^{3}He N (from Baym and Pethick, p. 117). Two estimates for the parameter $F_1^{\rm a}$ are given, based on two different methods.

9.2.3 Low temperature equilibrium properties

Entropy and specific heat

From the first law, we have

$$\begin{aligned} T\,\delta S &= \sum_{\boldsymbol{k},\sigma} (\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu)\,\delta n_{\boldsymbol{k},\sigma} \\ &= \sum_{\boldsymbol{k},\sigma} (\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu) \left\{ \frac{\partial n_{\boldsymbol{k},\sigma}}{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}}\,\delta\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} + \frac{\partial n_{\boldsymbol{k},\sigma}}{\partial \mu}\,\delta\mu + \frac{\partial n_{\boldsymbol{k},\sigma}}{\partial T}\,\delta T \right\} \\ &= \sum_{\boldsymbol{k},\sigma} (\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu) \left(\frac{\partial n_{\boldsymbol{k},\sigma}}{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}} \right) \left\{ (\delta\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \delta\mu) - \left(\frac{\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu}{T} \right) \delta T \right\} \quad . \end{aligned} \tag{9.41}$$

It turns out that the contribution of the $(\delta\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \delta\mu)$ term inside the curly brackets results in a contribution of order $T^3 \ln T$, which we shall accept on faith for the time being[5]. Thus, we are left with

$$\begin{aligned} \delta S &= -\sum_{\boldsymbol{k},\sigma} \left(\frac{\partial n_{\boldsymbol{k},\sigma}}{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}} \right) (\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} - \mu)^2\,\frac{\delta T}{T^2} = -V g(\varepsilon_{\rm F})\,\frac{\delta T}{T^2} \int_0^\infty d\varepsilon\,\frac{\partial n}{\partial \varepsilon}\,(\varepsilon - \mu)^2 \\ &= -V g(\varepsilon_{\rm F})\,k_{\rm B}^2\,\delta T \int_{-\infty}^{\infty} dx\,\frac{\partial}{\partial x}\left(\frac{1}{\exp(x)+1} \right) x^2 = \frac{\pi^2}{3}\,V\,g(\varepsilon_{\rm F})\,k_{\rm B}^2\,\delta T \quad . \end{aligned} \tag{9.42}$$

We conclude

$$S(T, V, N) = V\,\frac{\pi^2}{3}\,g(\varepsilon_{\rm F})\,k_{\rm B}^2\,T \tag{9.43}$$

[5]For a justification, see §1.4 of Baym and Pethick.

and

$$c_V(T,n) = \frac{T}{V}\left(\frac{\partial S}{\partial T}\right)_{V,N} = \frac{\pi^2}{3}\, g(\varepsilon_{\rm F})\, k_{\rm B}^2\, T \quad . \tag{9.44}$$

The difference between this result and that of the free fermi gas is the appearance of the effective mass m^* in the density of states $g(\varepsilon_{\rm F})$. If $c_V^0(T)$ is defined to be the low-temperature specific heat in a free Fermi gas of particles of mass m at the same density n, then

$$\frac{c_V(T)}{c_V^0(T)} = \frac{m^*}{m} \quad . \tag{9.45}$$

From $\delta F\big|_{V,N} = -S\,\delta T$, we integrate and obtain the temperature dependence of the ltz free energy,

$$F(T,V,N) = E_0(V,N) + V\,\frac{\pi^2}{6}\, g(\varepsilon_{\rm F})\,(k_{\rm B}T)^2 \quad . \tag{9.46}$$

Thus the chemical potential is

$$\begin{aligned}\mu(n,T) &= -\frac{\partial F}{\partial N}\bigg|_{T,V} = -\left(\frac{\partial (F/V)}{\partial (N/V)}\right)_T \\ &= \mu(n,T=0) + \frac{\pi^2}{6}\,(k_{\rm B}T)^2\,\frac{\partial g(\varepsilon_{\rm F})}{\partial n} \\ &= \mu(n,0) - \frac{\pi^2}{4}\,k_{\rm B}\left(\frac{1}{3} + \frac{\partial \ln m^*}{\partial \ln n}\right)\frac{T^2}{T_{\rm F}} \quad ,\end{aligned} \tag{9.47}$$

where $k_{\rm B}T_{\rm F} \equiv \hbar^2 k_{\rm F}^2/2m^*$.

Compressibility and sound velocity

Consider a swollen Fermi surface of radius $k_{\rm F} + dk_{\rm F}$, as depicted in Fig. 9.5. The change in the chemical potential is then given by

$$d\mu = \widetilde{\varepsilon}_{k_{\rm F}+dk_{\rm F}} - \widetilde{\varepsilon}_{k_{\rm F}} = d\widetilde{\varepsilon}_{k_{\rm F}} \quad , \tag{9.48}$$

where we assume no spin dependence in the dispersion. Thus,

$$\begin{aligned} d\mu &= d\varepsilon_{k_{\rm F}} + \frac{1}{V}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}_{\rm F}\sigma,\boldsymbol{k}'\sigma'}\,\delta n_{\boldsymbol{k}',\sigma'} = \hbar v_{\rm F}\,dk_{\rm F}\left\{1 + \int\frac{d^3k'}{(2\pi)^3}\sum_{\sigma'} f_{\boldsymbol{k}_{\rm F}\sigma,\boldsymbol{k}'\sigma'}\,\delta(\varepsilon_{\boldsymbol{k}'}-\mu)\right\} \\ &= \hbar v_{\rm F}\,dk_{\rm F}\left\{1 + 2\int\frac{d\Omega}{4\pi}\,f^{\rm s}(\vartheta)\int\frac{d^3k'}{(2\pi)^3}\,\delta(\varepsilon_{\boldsymbol{k}'}-\mu)\right\} = \hbar v_{\rm F}\,dk_{\rm F}\left\{1 + F_0^{\rm s}\right\} \quad .\end{aligned} \tag{9.49}$$

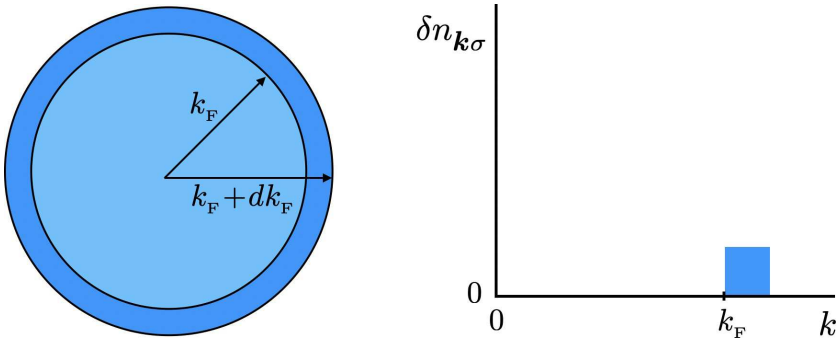


Figure 9.5: $\delta n_{\boldsymbol{k}\sigma}$ for a swollen Fermi surface.

We can now write

$$\begin{aligned}\kappa &= n^{-2}\,\frac{\partial n}{\partial \mu} = n^{-2}\,\frac{\partial n}{\partial k_{\rm F}}\,\frac{\partial k_{\rm F}}{\partial \mu} \\ &= n^{-2}\,\frac{k_{\rm F}^2}{\pi^2}\,\frac{1}{\hbar v_{\rm F}\,(1+F_0^{\rm s})} = \frac{n^{-2}\,g(\varepsilon_{\rm F})}{1+F_0^{\rm s}} = \frac{9\pi^2 m^*}{\hbar^2 k_{\rm F}^5\,(1+F_0^{\rm s})} \quad .\end{aligned} \tag{9.50}$$

Thus, if $\kappa^0 = n^{-2}\,g_0(\varepsilon_{\rm F})$ is the compressibility of the free Fermi gas with mass m at the same density n, we have

$$\frac{\kappa}{\kappa^0} = \frac{m^*/m}{1+F_0^{\rm s}} \quad . \tag{9.51}$$

To derive the connection with sound propagation, we examine the inviscid, weak flow limit of the Navier-Stokes equations, yielding $\partial_t(\varrho\boldsymbol{u}) = -\boldsymbol{\nabla} p$, where $\varrho = mn$ is the density, with m the *bare* mass and n the number density, and p the pressure. Local thermodynamics then gives $\boldsymbol{\nabla} p = (\partial p/\partial\varrho)\,\boldsymbol{\nabla}\varrho = (1/\varrho\kappa)\boldsymbol{\nabla}\varrho$. Taking the divergence,

$$-\frac{1}{\kappa}\boldsymbol{\nabla}\cdot\left(\frac{1}{\varrho}\boldsymbol{\nabla}\varrho\right) = \frac{\partial}{\partial t}\,\boldsymbol{\nabla}\cdot(\varrho\boldsymbol{u}) = -\frac{\partial^2\varrho}{\partial t^2} \quad , \tag{9.52}$$

where in the last equality we have invoked the continuity equation $\partial_t\varrho + \boldsymbol{\nabla}\cdot(\varrho\boldsymbol{u}) = 0$. Since $\boldsymbol{\nabla}\varrho$ is presumed to be small, we arrive at the Helmholtz equation,

$$\frac{1}{\bar\varrho\kappa}\,\nabla^2\varrho = \frac{\partial^2\varrho}{\partial t^2} \quad , \tag{9.53}$$

with wave propagation speed $s = 1/\sqrt{\bar\varrho\kappa}$, where $\bar\varrho$ is the average density.

Uniform magnetic susceptibility

In the presence of an external magnetic field B, there is an additional Zeeman contribution to the Hamiltonian, $\hat H_{\rm Z} = -\mu_0 B\sum_{\boldsymbol{k},\sigma}\sigma n_{\boldsymbol{k},\sigma}$. This causes the $\uparrow$ Fermi surface to expand and the $\downarrow$

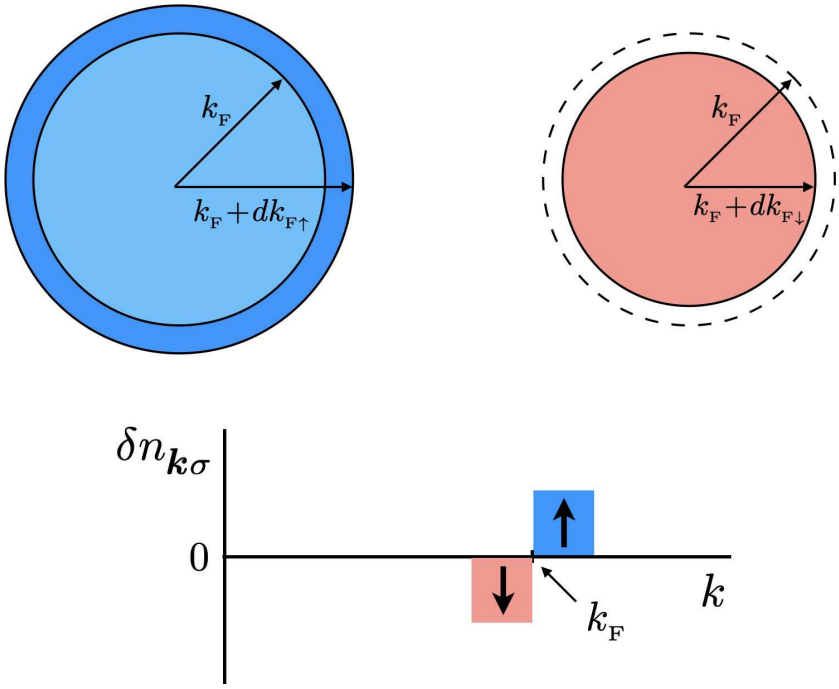


Figure 9.6: $\delta n_{\boldsymbol{k}\sigma}$ in the presence of a magnetic field.

Fermi surface to contract. Thus $dk_{F\uparrow} = -dk_{F\downarrow} \equiv dk_F$ and $\delta n_{\boldsymbol{k},\sigma} = \sigma\,\delta(k_F - k)\,dk_F$. The situation is depicted in Fig. 9.6. If particle number is conserved, then the chemical potential, which is the same for each spin species, is unchanged to lowest order in B. Thus,

$$\begin{aligned} 0 = d\widetilde{\varepsilon}_{k_F,\sigma} &= -\sigma\mu_0\,dB + d\varepsilon_{k_F,\sigma} + \frac{1}{V}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}_F\sigma,\boldsymbol{k}'\sigma'}\,\delta n_{\boldsymbol{k}',\sigma'} \\ &= -\sigma\mu_0\,dB + \hbar v_F\,dk_F\left\{\sigma + \int\frac{d^3k'}{(2\pi)^3}\,f_{\boldsymbol{k}_F\sigma,\boldsymbol{k}'\sigma'}\,\sigma'\,\delta(\varepsilon_{\boldsymbol{k}'}-\mu)\right\} \\ &= -\sigma\mu_0\,dB + \sigma\hbar v_F\,dk_F\left\{1 + g(\varepsilon_F)\int\frac{d\Omega}{4\pi}\,f^{\rm a}(\vartheta)\right\} \\ &= -\sigma\mu_0\,dB + \sigma\hbar v_F\,(1+F_0^{\rm a})\,dk_F \quad . \end{aligned} \tag{9.54}$$

Note that we have invoked the fact that $\sum_{\sigma'}\sigma' f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} = 2\sigma f^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'}$. We conclude that

$$\frac{\partial k_F}{\partial B} = \frac{\mu_0}{\hbar v_F\,(1+F_0^{\rm a})} \quad . \tag{9.55}$$

The magnetic susceptibility is then

$$\chi = \frac{1}{V}\left(\frac{\partial M}{\partial B}\right)_{N,V,B=0} = \mu_0\left(\frac{\partial n_\uparrow}{\partial B} - \frac{\partial n_\downarrow}{\partial B}\right) = \mu_0\left(\frac{\partial n_\uparrow}{\partial k_{F\uparrow}} + \frac{\partial n_\downarrow}{\partial k_{F\downarrow}}\right)\left(\frac{\partial k_F}{\partial B}\right)_{B=0} = \frac{\mu_0^2\,g(\varepsilon_F)}{1+F_0^{\rm a}} \quad , \tag{9.56}$$

and therefore

$$\frac{\chi}{\chi^0} = \frac{m^*/m}{1+F_0^{\rm a}} \quad , \tag{9.57}$$

where $\chi^0 = \mu_0^2\, g_0(\varepsilon_{\rm F})$

Galilean invariance

Consider now a Galilean transformation to an inertial primed frame of reference moving at constant velocity $\boldsymbol{u}$ with respect to our unprimed inertial laboratory frame. The Hamiltonian in the primed frame is

$$\begin{aligned}\hat{H}' &= \sum_{i=1}^N \frac{(\boldsymbol{p}_i - m\boldsymbol{u})^2}{2m} + \hat{H}_1 \\ &= \hat{H} - \boldsymbol{u}\cdot\boldsymbol{P} + \tfrac{1}{2}M\boldsymbol{u}^2 \quad ,\end{aligned} \tag{9.58}$$

where $\boldsymbol{P} = \sum_i \boldsymbol{p}_i$ is the total momentum and $M = Nm$ is the total mass. Let's now add a particle of momentum $\boldsymbol{p} = \hbar\boldsymbol{k}$ and spin polarization σ in the lab frame at $T = 0$, where its energy is then $\varepsilon_{\boldsymbol{k},\sigma}$. In the primed frame, however, the added particle has momentum $\hbar\boldsymbol{k} - m\boldsymbol{u}$ and energy $\widetilde{\varepsilon}_{\boldsymbol{k},\sigma} = \varepsilon_{\boldsymbol{k},\sigma} - \hbar\boldsymbol{k}\cdot\boldsymbol{u} + \frac{1}{2}m\boldsymbol{u}^2$. Thus, $\tilde{\varepsilon}'_{\boldsymbol{k}-\hbar^{-1}m\boldsymbol{u},\sigma} = \varepsilon_{\boldsymbol{k},\sigma} - \hbar\boldsymbol{k}\cdot\boldsymbol{u} + \frac{1}{2}m\boldsymbol{u}^2$, or, equivalently,

$$\tilde{\varepsilon}'_{\boldsymbol{k},\sigma} = \varepsilon_{\boldsymbol{k}+\hbar^{-1}m\boldsymbol{u},\sigma} - \hbar\boldsymbol{k}\cdot\boldsymbol{u} - \tfrac{1}{2}m\boldsymbol{u}^2 \quad . \tag{9.59}$$

Note though that $\tilde{\varepsilon}'_{\boldsymbol{k},\sigma} = \tilde{\varepsilon}'_{\boldsymbol{k},\sigma}[\{n'_{\boldsymbol{k},\sigma}\}]$, with

$$\begin{aligned}n'_{\boldsymbol{k},\sigma} = n^0_{\boldsymbol{k}+\hbar^{-1}m\boldsymbol{u},\sigma} &= n^0_{\boldsymbol{k},\sigma} + \frac{m\boldsymbol{u}}{\hbar}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\, n^0_{\boldsymbol{k},\sigma} \\ &= n^0_{\boldsymbol{k}} - m v_{\rm F}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}}\,\delta(\varepsilon_{\boldsymbol{k},\sigma} - \mu) \quad .\end{aligned} \tag{9.60}$$

This relation is illustrated in Fig. 9.7. Thus, we have

$$\begin{aligned}\tilde{\varepsilon}'_{\boldsymbol{k},\sigma} &= \varepsilon_{\boldsymbol{k},\sigma} + \frac{1}{V}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta n'_{\boldsymbol{k}'\sigma'} \\ &= \varepsilon_{\boldsymbol{k},\sigma} - m v_{\rm F}\sum_{\sigma'}\int\frac{d^3k'}{(2\pi)^3}\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}}'\,\delta(\varepsilon_{\boldsymbol{k}',\sigma'} - \mu) \\ &= \varepsilon_{\boldsymbol{k},\sigma} - m v_{\rm F}\, g(\varepsilon_{\rm F})\,\boldsymbol{u}\cdot\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,\hat{\boldsymbol{k}}'\, f^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'_{\rm F}}\end{aligned} \tag{9.61}$$

We are only interested in the case where $|\boldsymbol{k}| \approx k_{\rm F}$, and thus we may write

$$\begin{aligned}\tilde{\varepsilon}'_{\boldsymbol{k}_{\rm F},\sigma} &= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} - m v_{\rm F}\,\boldsymbol{u}\cdot\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,\hat{\boldsymbol{k}}'\, F^{\rm s}_{\boldsymbol{k}_{\rm F},\boldsymbol{k}'_{\rm F}} \\ &= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} - m v_{\rm F}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\, F^{\rm s}_{\boldsymbol{k}_{\rm F},\boldsymbol{k}'_{\rm F}} \\ &= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} - \tfrac{1}{3}F^{\rm s}_1\, m v_{\rm F}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}} \quad .\end{aligned} \tag{9.62}$$

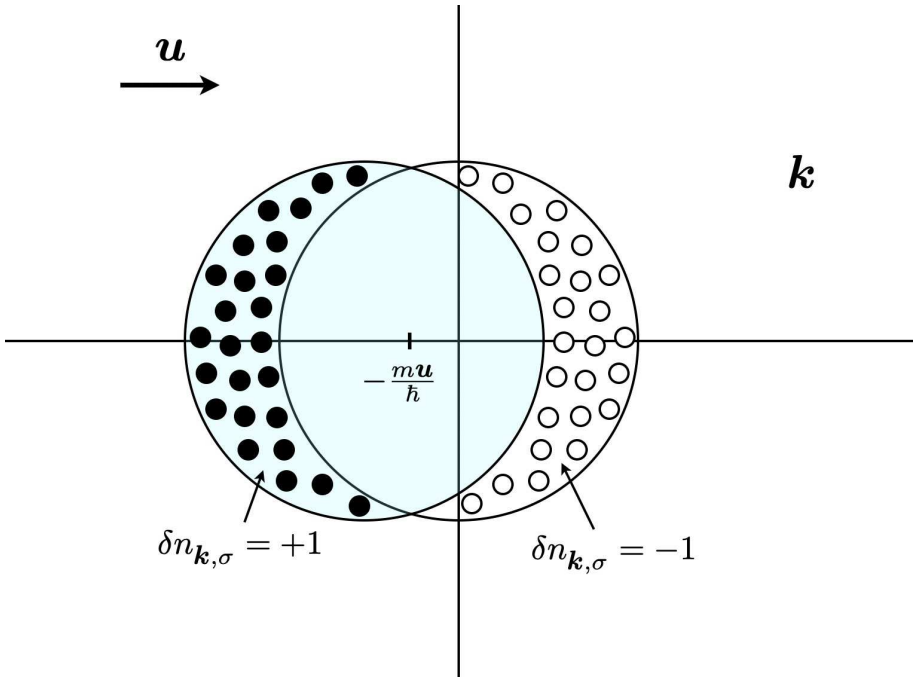


Figure 9.7: Distribution of quasiparticle occupancies in a frame moving with velocity $\boldsymbol{u}$.

Note that we have used above the fact that the integral

$$\int \frac{d\hat{\boldsymbol{k}}'}{4\pi}\, \hat{\boldsymbol{k}}'\, F^{\rm s}_{\boldsymbol{k}_{\rm F},\boldsymbol{k}'_{\rm F}} = C\hat{\boldsymbol{k}} \tag{9.63}$$

must by rotational isotropy lie along $\hat{\boldsymbol{k}}$. Taking the dot product with $\hat{\boldsymbol{k}}$ then gives

$$C = \int \frac{d\hat{\boldsymbol{k}}'}{4\pi}\, \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{k}}'\, F^{\rm s}(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'}) = \tfrac{1}{3}F^{\rm s}_1 \quad . \tag{9.64}$$

Putting this all together, we have

$$\begin{aligned}
\tilde{\varepsilon}'_{\boldsymbol{k}_{\rm F},\sigma} &= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} - \tfrac{1}{3}\, F^{\rm s}_1\, m v_{\rm F}\, \boldsymbol{u}\cdot\hat{\boldsymbol{k}} \\
&= \varepsilon_{\boldsymbol{k}_{\rm F}+\hbar^{-1}m\boldsymbol{u},\sigma} - \hbar\boldsymbol{k}\cdot\boldsymbol{u} - \tfrac{1}{2}m\boldsymbol{u}^2 \\
&= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} + \frac{m\boldsymbol{u}}{\hbar}\cdot\boldsymbol{\nabla}_{\boldsymbol{k}}\,\varepsilon_{\boldsymbol{k},\sigma}\big|_{\boldsymbol{k}=\boldsymbol{k}_{\rm F}} - \hbar k_{\rm F}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}} - \tfrac{1}{2}m\boldsymbol{u}^2 \\
&= \varepsilon_{\boldsymbol{k}_{\rm F},\sigma} + (m-m^*)v_{\rm F}\,\boldsymbol{u}\cdot\hat{\boldsymbol{k}} - \tfrac{1}{2}m\boldsymbol{u}^2 \quad ,
\end{aligned} \tag{9.65}$$

Thus, to lowest order in $\boldsymbol{u}$, we have

$$(m-m^*) = -\tfrac{1}{3}F^{\rm s}_1\, m \quad \Rightarrow \quad \frac{m^*}{m} = 1 + \tfrac{1}{3}F^{\rm s}_1 \quad . \tag{9.66}$$

This result is connected with the following point. The total particle current is given by

$$\boldsymbol{J} = \sum_{\boldsymbol{k},\sigma} \frac{1}{\hbar} \frac{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}}{\partial \boldsymbol{k}} \, n_{\boldsymbol{k},\sigma} \quad , \tag{9.67}$$

where it is $\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}$ and not $\varepsilon_{\boldsymbol{k},\sigma}$ which appears.

We again stress that this relationship between m^*/m and $F_1^{\rm s}$ is valid only in Galilean invariant systems, such as liquid ^{3}He N. The imposition of a crystalline lattice potential breaks the Galilean symmetry and invalidates the above result.

9.2.4 Thermodynamic stability at $T = 0$

Consider a $T = 0$ distortion of the Fermi surface. The Landau free energy $\Omega = E - TS + \mu N$ must be a minimum with respect to all possible such distortions. We adopt the parameterization

$$\begin{aligned} n_{\boldsymbol{k},\sigma} &= \Theta(k_{\rm F}(\hat{\boldsymbol{k}},\sigma) - k) = \Theta(k_{\rm F} + \delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma) - k) \\ &= \Theta(k_{\rm F} - k) + \delta(k_{\rm F} - k)\,\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma) + \tfrac{1}{2}\delta'(k_{\rm F} - k)\left[\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\right]^2 + \ldots \quad , \end{aligned} \tag{9.68}$$

where $\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)$ is the local FS distortion in the direction $\hat{\boldsymbol{k}}$ for spin polarization σ. We now evaluate $\Omega(T=0) = E - \mu N$ to second order in $\delta k_{\rm F}$:

$$\begin{aligned} \Omega &= \Omega_0 + \sum_{\boldsymbol{k},\sigma} (\varepsilon_{\boldsymbol{k},\sigma} - \mu)\,\delta n_{\boldsymbol{k},\sigma} + \frac{1}{2V}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta n_{\boldsymbol{k},\sigma}\,\delta n_{\boldsymbol{k}',\sigma'} \\ &= \Omega_0 + \sum_{\boldsymbol{k},\sigma} (\varepsilon_{\boldsymbol{k},\sigma} - \mu)\left\{\delta(k_{\rm F} - k)\,\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma) + \tfrac{1}{2}\delta'(k_{\rm F} - k)\left[\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\right]^2\right\} \\ &\qquad + \frac{1}{2V}\sum_{\boldsymbol{k},\sigma}\sum_{\boldsymbol{k}',\sigma'} f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta(k_{\rm F} - k)\,\delta(k_{\rm F} - k')\,\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\,\delta k_{\rm F}(\hat{\boldsymbol{k}}',\sigma') \quad , \end{aligned} \tag{9.69}$$

which entails

$$\begin{aligned} \frac{\Omega - \Omega_0}{V} &= \sum_\sigma \int \frac{d^3k}{(2\pi)^3}\left\{-\frac{\partial}{\partial k}\,\delta(k_{\rm F} - k)\right\}\left[\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\right]^2 \\ &\qquad + \frac{k_{\rm F}^4}{8\pi^4}\sum_{\sigma,\sigma'}\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\, f_{\sigma,\sigma'}(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\,\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\,\delta k_{\rm F}(\hat{\boldsymbol{k}}',\sigma') \\ &= \frac{\hbar^2 k_{\rm F}^3}{4\pi^2 m^*}\left\{\sum_\sigma\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\left[\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\right]^2\right. \\ &\qquad \left. + \frac{1}{2}\sum_{\sigma,\sigma'}\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\, F_{\sigma,\sigma'}(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\,\delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma)\,\delta k_{\rm F}(\hat{\boldsymbol{k}}',\sigma')\right\} \quad . \end{aligned} \tag{9.70}$$

Recall now that $F_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'} = F^{\rm s}_{\boldsymbol{k},\boldsymbol{k}'} + \sigma\sigma' F^{\rm a}_{\boldsymbol{k},\boldsymbol{k}'}$, so if we define the symmetric and antisymmetric components of the FS distortion

$$\delta k^{\rm s}_{\rm F}(\hat{\boldsymbol{k}}) \equiv \sum_\sigma \delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma) \quad , \quad \delta k^{\rm a}_{\rm F}(\hat{\boldsymbol{k}}) \equiv \sum_\sigma \sigma\, \delta k_{\rm F}(\hat{\boldsymbol{k}},\sigma) \quad , \tag{9.71}$$

then

$$\frac{\Omega-\Omega_0}{V} = \frac{\hbar^2 k_{\rm F}^3}{8\pi^2 m^*} \sum_{\nu={\rm s,a}} \left\{ \int \frac{d\hat{\boldsymbol{k}}}{4\pi} \left[\delta k^\nu_{\rm F}(\hat{\boldsymbol{k}})\right]^2 + \int \frac{d\hat{\boldsymbol{k}}}{4\pi} \int \frac{d\hat{\boldsymbol{k}}'}{4\pi}\, F^\nu(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\, \delta k^\nu_{\rm F}(\hat{\boldsymbol{k}})\, \delta k^\nu_{\rm F}(\hat{\boldsymbol{k}}') \right\} \quad . \tag{9.72}$$

Having resolved the free energy into contributions from the spin symmetric and antisymmetric distortions of the FS, we now further resolve it into angular momentum channels, writing

$$\delta k^\nu_{\rm F}(\hat{\boldsymbol{k}}) = \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell A^\nu_{\ell,m}\, Y_{\ell,m}(\hat{\boldsymbol{k}}) \quad , \tag{9.73}$$

where $A^\nu_{\ell,-m} = A^{\nu *}_{\ell,m}$ since $\delta k^\nu_{\rm F}(\hat{\boldsymbol{k}})$ is real. We also have

$$F^\nu(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'}) = \sum_{\ell=0}^\infty F^\nu_\ell\, P_\ell(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'}) = \sum_{\ell=0}^\infty \sum_{m=-\ell}^\ell \frac{4\pi}{2\ell+1}\, F^\nu_\ell\, Y^*_{\ell,m}(\hat{\boldsymbol{k}})\, Y_{\ell,m}(\hat{\boldsymbol{k}}') \quad , \tag{9.74}$$

and invoking the orthonormality of the spherical harmonics,

$$\int d\hat{\boldsymbol{k}}\, Y^*_{\ell,m}(\hat{\boldsymbol{k}})\, Y_{\ell' m'}(\hat{\boldsymbol{k}}) = \delta_{\ell\ell'}\, \delta_{mm'} \quad , \tag{9.75}$$

we obtain the pleasingly compact expression

$$\frac{\Omega-\Omega_0}{V} = \frac{\hbar^2 k_{\rm F}^3}{32\,\pi^3 m^*} \sum_{\nu={\rm s,a}} \left(1 + \frac{F^\nu_\ell}{2\ell+1}\right) |A^\nu_{\ell,m}|^2 \quad . \tag{9.76}$$

The stability criterion in each angular momentum channel is then

$$F^\nu_\ell > -(2\ell+1) \quad , \tag{9.77}$$

where $\nu \in \{{\rm s,a}\}$.

What happens when these stability criteria are violated? According to Eqn. 9.76, the free energy can be made arbitrarily negative by increasing the amplitude(s) $A^\nu_{\ell,m}$ of any FS distortion for which $F^\nu_\ell < -(2\ell+1)$. This is unphysical, and an artifact of going only to order $(\delta k^\nu_{\rm F})^2$ in the expansion of the Landau free energy. Suppose though we add a fourth order correction to Ω of the form

$$\frac{\Delta\Omega}{V} = \frac{\hbar^2 k_{\rm F}^3}{4\pi m^*} \sum_{\nu={\rm s,a}} \lambda_\nu \left(\int \frac{d\hat{\boldsymbol{k}}}{4\pi} \left[\delta k^\nu_{\rm F}(\hat{\boldsymbol{k}})\right]^2 \right)^2 = \frac{\hbar^2 k_{\rm F}^3}{64\pi^3 m^*} \sum_{\nu={\rm s,a}} \lambda_\nu \left(\sum_{\ell,m} |A^\nu_{\ell,m}|^2 \right)^2 \tag{9.78}$$

so that

$$\frac{\Omega + \Delta\Omega - \Omega_0}{V} = \frac{\hbar^2 k_{\rm F}^3}{32\,\pi^3 m^*} \sum_{\nu={\rm s,a}} \left\{ \sum_{\ell,m} \left(1 + \frac{F_\ell^\nu}{2\ell+1}\right) |A_{\ell,m}^\nu|^2 + \tfrac{1}{2}\,\lambda_\nu \left(\sum_{\ell,m} |A_{\ell,m}^\nu|^2 \right)^2 \right\} \quad . \tag{9.79}$$

Such a term lies beyond the expansion for the internal energy of a Fermi liquid that we have considered thus far. To minimize the free energy, we set the variation with respect to each $A_{\ell,m}^{\nu *}$ to zero. For stable channels where $F_\ell^\nu > -(2\ell+1)$, we then find $A_{\ell,m}^\nu = 0$. But for unstable channels, we obtain

$$\sum_{m=-\ell}^{\ell} |A_{\ell,m}^\nu|^2 = -\frac{1}{\lambda_\nu}\left(1 + \frac{F_\ell^\nu}{2\ell+1}\right) > 0 \quad . \tag{9.80}$$

Thus, the weight of the distortion in each unstable (ν, ℓ) sector is distributed over all $(2\ell+1)$ of the coefficients $A_{\ell,m}^\nu$ such that the sum of their squares is fixed as specified above. Thus, an $\ell = 1$ instability results in a dipolar distortion of the FS, while an $\ell = 2$ instability results in a quadrupolar distortion of the FS, *etc.*

9.3 Collective Dynamics of the Fermi Surface

9.3.1 Landau-Boltzmann equation

We first review some basic features of the Boltzmann equation, which was discussed earlier in §5.6. Consider the classical dynamical system governing flow on an N-dimensional phase space Γ, where $\boldsymbol{X} = (X^1, \ldots, X^N) \in \Gamma$ is a point in phase space. The dynamical system is

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{V}(\boldsymbol{X}) \tag{9.81}$$

where each $V^\mu = V^\mu(X^1, \ldots, X^N)$[6]. Now consider a distribution function $f(\boldsymbol{X}, t)$. The continuity equation says

$$\frac{\partial f}{\partial t} + \boldsymbol{\nabla}\cdot(\boldsymbol{V} f) = 0 \quad , \tag{9.82}$$

where $\boldsymbol{\nabla} = \left(\frac{\partial}{\partial X^1}, \ldots, \frac{\partial}{\partial X^N}\right)$. Assuming phase flow is *incompressible*, $\boldsymbol{\nabla}\cdot\boldsymbol{V} = 0$ and the continuity equation takes the form

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + \boldsymbol{V}\cdot\boldsymbol{\nabla} f = 0 \quad , \tag{9.83}$$

where $\frac{Df}{Dt} = \frac{d}{dt} f(\boldsymbol{X}(t), t)$, called the *convective derivative*, is the total derivative of the distribution in the frame comoving with the flow.

[6]This autonomous system can be extended to a time-dependent one, *i.e.* $\dot{\boldsymbol{X}} = \boldsymbol{V}(\boldsymbol{X}, t)$, which is a dynamical system in one higher $(N+1)$ dimensions, taking $X^{N+1} = t$ and $V^{N+1} = 1$.

For our application, phase space has dimension $N = 6$, with $\boldsymbol{X} = (\boldsymbol{r}, \boldsymbol{k})$. We also add to the RHS a source/sink term corresponding to *collisions* between particles. Typically these are local in position $\boldsymbol{r}$ but nonlocal in the wavevector $\boldsymbol{k}$. An example is shown in Fig. 9.3, where a collision results in an instantaneous wavevector $\boldsymbol{q}$ transfer between two interacting particles. We also must account for spin, and the most straightforward way to do this is to specify independent distributions for each spin polarization. Writing $f(\boldsymbol{r}, \boldsymbol{k}, \sigma, t) = n_{\boldsymbol{k},\sigma}(\boldsymbol{r}, t)$, our Boltzmann equation takes the form

$$\frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial t} + \langle \dot{\boldsymbol{r}} \rangle_\sigma \cdot \frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{r}} + \langle \dot{\boldsymbol{k}} \rangle_\sigma \cdot \frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{k}} = I[n] \quad , \tag{9.84}$$

where $I[n]$ is the collision term. We now invoke Landau's Fermi liquid theory, but on a local scale, and write the energy density $\mathcal{E}(\boldsymbol{r}, t)$ as a functional of the distribution $\delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r}, t)$, *viz.*

$$\mathcal{E}(\boldsymbol{r},t) = \mathcal{E}_0 + \sum_\sigma \int \frac{d^3k}{(2\pi)^3}\, \varepsilon_{\boldsymbol{k},\sigma}\, \delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t) + \frac{1}{2}\sum_{\sigma,\sigma'} \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3}\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\, \delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)\, \delta n_{\boldsymbol{k}',\sigma'}(\boldsymbol{r},t) \quad , \tag{9.85}$$

where $\delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r}, t)$ is dimensionless and indicates the local number density of fermions of wavevector $\boldsymbol{k}$ and spin polarization σ in units of the bulk number density n. Note that the above expression is local in position space. We then have the Landau-Boltzmann equation[7],

$$\frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial t} + \overbrace{\frac{1}{\hbar}\frac{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{k}}}^{\langle \dot{\boldsymbol{r}} \rangle_\sigma} \cdot \frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{r}} - \overbrace{\frac{1}{\hbar}\frac{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{r}}}^{-\langle \dot{\boldsymbol{k}} \rangle_\sigma} \cdot \frac{\partial n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)}{\partial \boldsymbol{k}} = I[n] \quad , \tag{9.86}$$

where

$$\widetilde{\varepsilon}_{\boldsymbol{k},\sigma}(\boldsymbol{r},t) = V_\sigma(\boldsymbol{r},t) + \varepsilon_{\boldsymbol{k},\sigma}(\boldsymbol{r},t) + \sum_{\sigma'} \int \frac{d^3k'}{(2\pi)^3}\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\, \delta n_{\boldsymbol{k}',\sigma'}(\boldsymbol{r},t) \quad . \tag{9.87}$$

Here we have included $V_\sigma(\boldsymbol{r}, t)$, the external local potential for particles at position $\boldsymbol{r}$ at time t. Note that

$$\frac{\partial}{\partial \boldsymbol{r}}\, \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}(\boldsymbol{r},t) = \frac{\partial}{\partial \boldsymbol{r}}\, V_\sigma(\boldsymbol{r},t) + \sum_{\sigma'} \int \frac{d^3k'}{(2\pi)^3}\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\, \frac{\partial}{\partial \boldsymbol{r}}\, \delta n_{\boldsymbol{k}',\sigma'}(\boldsymbol{r},t) \tag{9.88}$$

Now we write linearize, writing $n = n^0 + \delta n$, obtaining

$$\frac{\partial\, \delta n_{\boldsymbol{k},\sigma}}{\partial t} + \frac{1}{\hbar}\frac{\partial \varepsilon_{\boldsymbol{k},\sigma}}{\partial \boldsymbol{k}} \cdot \frac{\partial\, \delta n_{\boldsymbol{k},\sigma}}{\partial \boldsymbol{r}} - \frac{1}{\hbar}\frac{\partial n^0_{\boldsymbol{k},\sigma}}{\partial \boldsymbol{k}} \cdot \frac{\partial \widetilde{\varepsilon}_{\boldsymbol{k},\sigma}}{\partial \boldsymbol{r}} = I[n^0 + \delta n] \quad . \tag{9.89}$$

If $V_\sigma(\boldsymbol{r}, t) = \delta \hat{V}_\sigma\, e^{i(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t)}$, then the solution for the distribution in the linearized theory will be $\delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r}, t) = \delta \hat{n}_{\boldsymbol{k},\sigma}\, e^{i(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t)}$, with

$$\omega\, \delta \hat{n}_{\boldsymbol{k},\sigma} - \boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k},\sigma}\, \delta \hat{n}_{\boldsymbol{k},\sigma} + \left(\frac{\partial n^0_{\boldsymbol{k},\sigma}}{\partial \varepsilon_{\boldsymbol{k},\sigma}}\right) \boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k},\sigma} \left[\delta \hat{V}_\sigma + \sum_{\sigma'} \int \frac{d^3k'}{(2\pi)^3}\, f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\, \delta \hat{n}_{\boldsymbol{k}',\sigma'}\right] = -\big[\mathcal{L}\, \delta \hat{n}\big]_{\boldsymbol{k},\sigma} \quad , \tag{9.90}$$

[7] We assume no curvature $\boldsymbol{\Omega}(\boldsymbol{k})$ contributing to the velocity $\dot{\boldsymbol{r}}$.

where $\mathcal{L}$ is the *linearized collision operator*. Note that this is a linear integral (or integrodifferential, depending on the form of $\mathcal{L}$) equation for $\delta\hat{n}_{\boldsymbol{k},\sigma}$ in terms of $\delta\hat{V}_\sigma$.

9.3.2 Zero sound : free FS oscillations in the collisionless limit

We now consider the case of free oscillations of the Fermi surface, *i.e.* the case $V_\sigma(\boldsymbol{r},t)=0$, in the collisionless limit ($\mathcal{L}=0$). We are left with

$$(\omega-\boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k},\sigma})\,\delta\hat{n}_{\boldsymbol{k},\sigma}+\boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k},\sigma}\left(\frac{\partial n^0_{\boldsymbol{k},\sigma}}{\partial\varepsilon_{\boldsymbol{k},\sigma}}\right)\sum_{\sigma'}\int\frac{d^3k'}{(2\pi)^3}\,f_{\boldsymbol{k}\sigma,\boldsymbol{k}'\sigma'}\,\delta\hat{n}_{\boldsymbol{k}',\sigma'}=0\quad . \tag{9.91}$$

This is an *eigenvalue equation* for $\omega(\boldsymbol{q})$, where the eigenvector is the distribution $\delta\hat{n}_{\boldsymbol{k},\sigma}$. If we write

$$\delta n_{\boldsymbol{k},\sigma}(\boldsymbol{r},t)=\hbar v_{\text{F}}\,\delta(\varepsilon_{\text{F}}-\varepsilon_{\boldsymbol{k},\sigma})\,\delta k_{\text{F}}(\hat{\boldsymbol{k}},\sigma)\,e^{i(\boldsymbol{q}\cdot\boldsymbol{r}-\omega t)}\quad , \tag{9.92}$$

then we arrive at

$$(\omega-\boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k}_{\text{F}},\sigma})\,\delta k_{\text{F}}(\hat{\boldsymbol{k}},\sigma)-\boldsymbol{q}\cdot\boldsymbol{v}_{\boldsymbol{k}_{\text{F}},\sigma}\sum_{\sigma'}\int\frac{d^3k'}{(2\pi)^3}\,\delta(\varepsilon_{\text{F}}-\varepsilon_{\boldsymbol{k}',\sigma'})\,f_{\boldsymbol{k}_{\text{F}}\sigma,\boldsymbol{k}'_{\text{F}}\sigma'}\,\delta k_{\text{F}}(\hat{\boldsymbol{k}}',\sigma')=0\quad . \tag{9.93}$$

We now take $\boldsymbol{v}_{\boldsymbol{k},\sigma}=v_{\text{F}}\hat{\boldsymbol{k}}$, independent of σ. Thus,

$$(\lambda-\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}})\,\delta k_{\text{F}}(\hat{\boldsymbol{k}},\sigma)-\tfrac{1}{2}\,\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,F_{\sigma,\sigma'}(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\,\delta k_{\text{F}}(\hat{\boldsymbol{k}}',\sigma')=0\quad , \tag{9.94}$$

where $\lambda\equiv\omega/v_{\text{F}}q$. This is immediately resolved into symmetric and antisymmetric channels $\nu\in\{\text{s},\text{a}\}$, *viz.*

$$(\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}-\lambda)\,\delta k^\nu_{\text{F}}(\hat{\boldsymbol{k}})+\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,F^\nu(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\,\delta k^\nu_{\text{F}}(\hat{\boldsymbol{k}}')=0 \tag{9.95}$$

Thus,

$$\delta k^\nu_{\text{F}}(\hat{\boldsymbol{k}})=\frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda-\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,F^\nu(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\,\delta k^\nu_{\text{F}}(\hat{\boldsymbol{k}}')\quad , \tag{9.96}$$

and resolving into angular momentum channels as before, writing

$$F^\nu(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})=\sum_{\ell,m}\frac{4\pi\,F^\nu_\ell}{2\ell+1}\,Y_{\ell,m}(\hat{\boldsymbol{k}})\,Y^*_{\ell,m}(\hat{\boldsymbol{k}}')\qquad,\qquad\delta k^\nu_{\text{F}}(\hat{\boldsymbol{k}})=\sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell}A^\nu_{\ell,m}\,Y_{\ell,m}(\hat{\boldsymbol{k}})\quad , \tag{9.97}$$

multiplying the above equation by $Y^*_{\ell,m}(\hat{\boldsymbol{k}})$ and then integrating over the unit $\hat{\boldsymbol{k}}$ sphere, we obtain

$$A^\nu_{\ell,m}=\sum_{\ell',m'}\frac{F^\nu_{\ell'}}{2\ell'+1}\left[\int d\hat{\boldsymbol{k}}\,\frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda-\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}\,Y^*_{\ell,m}(\hat{\boldsymbol{k}})\,Y_{\ell',m'}(\hat{\boldsymbol{k}})\right]A^\nu_{\ell',m'} \tag{9.98}$$

The oscillations of the FS are called *zero sound*.

Simple model for zero sound

Eqn. 9.98 defines an eigenvalue equation for the infinite length vector $\boldsymbol{A} = \{A_{0,0}, A_{1,-1}, A_{1,0}, \ldots\}$. So simplify matters, consider the case where $F_\ell^\nu = F_0^\nu \, \delta_{\ell,0}$. We drop the ν superscript for clarity. Eqn. 9.98 then reduces to

$$1 = F_0 \int \frac{d\hat{\boldsymbol{k}}}{4\pi} \, \frac{\hat{\boldsymbol{q}} \cdot \hat{\boldsymbol{k}}}{\lambda - \hat{\boldsymbol{q}} \cdot \hat{\boldsymbol{k}}} = F_0 \left[\frac{\lambda}{2} \, \ln\!\left(\frac{\lambda + 1}{\lambda - 1} \right) - 1 \right] \quad , \tag{9.99}$$

which is equivalent to

$$\left(1 + \frac{1}{F_0} \right) \lambda^{-1} = \tanh^{-1}(\lambda^{-1}) \quad . \tag{9.100}$$

This is a transcendental equation for $\lambda(F_0)$. It may be solved graphically by plotting the LHS and RHS *versus* the quantity $u \equiv \lambda^{-1}$. One finds that a nontrivial solution with real λ exists provided $F_0 > 0$. For $F_0 \in [-1, 0]$, a complex solution exists, corresponding to a damped oscillation. We may also solve explicitly in two limits:

$$\begin{aligned} F_0 \to 0 \quad &\Rightarrow \quad \lambda \to 1 \quad \Rightarrow \quad \frac{\lambda}{2} \ln\left(\frac{\lambda + 1}{\lambda - 1} \right) = \frac{1}{2} \ln\left(\frac{2}{\lambda - 1} \right) + \ldots \quad \Rightarrow \quad \lambda \simeq 1 + 2\, e^{-2/F_0} \\ F_0 \to \infty \quad &\Rightarrow \quad \lambda \to \infty \quad \Rightarrow \quad \frac{\lambda}{2} \ln\left(\frac{\lambda + 1}{\lambda - 1} \right) = 1 + \frac{1}{3\lambda^2} + \ldots \quad \Rightarrow \quad \lambda \simeq \sqrt{\frac{F_0}{3}} \end{aligned} \tag{9.101}$$

The ratio of zero sound to first sound velocities is thus

$$\frac{c_0}{c_1} = \frac{\sqrt{3}\, \lambda(F_0^{\rm s})}{\sqrt{(1 + F_0^{\rm s})(1 + \frac{1}{3} F_1^{\rm s})}} \quad . \tag{9.102}$$

Another zero sound mode

Consider next the truncated Landau interaction function

$$\begin{aligned} F(\vartheta_{\hat{\boldsymbol{k}}, \hat{\boldsymbol{k}}'}) &= F_0 + F_1 \, \hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}' \\ &= F_0 + F_1 \, \cos\theta \cos\theta' + \tfrac{1}{2} F_1 \, \sin\theta \sin\theta' \Big(e^{i\phi} \, e^{-i\phi'} + e^{-i\phi} e^{i\phi'} \Big) \quad . \end{aligned} \tag{9.103}$$

We posit a Fermi surface distortion of the form $\delta k_{\rm F}(\hat{\boldsymbol{k}}) = u(\theta)\, e^{i\phi}$, resulting in the eigenvalue equation

$$u(\theta) = \frac{F_1}{4} \, \frac{\sin\theta \cos\theta}{\lambda - \cos\theta} \int_0^\pi d\theta' \sin^2\!\theta' \, u(\theta') \quad . \tag{9.104}$$

Multiply by $\sin\theta$ and integrate to obtain

$$\frac{4}{F_1} = \int\limits_{-1}^{1} dx\, \frac{x - x^3}{\lambda - x} = -\lambda(\lambda^2 - 1)\ln\left(\frac{\lambda+1}{\lambda-1}\right) + 2\lambda^2 - \tfrac{4}{3} \quad , \tag{9.105}$$

where $x = \cos\theta$. Note that at the limiting value $\lambda = 0$ the integral returns a value of $\frac{2}{3}$, corresponding to $F_1 = 6$. In the opposite limit $\lambda \to \infty$, the RHS takes the value $2/3\lambda^2$. Thus, there should be a solution for $F_1 \in [6, \infty]$. According to Tab. 9.1, in ^{3}He N at high pressure one indeed has $F_1^{\rm s} > 6$, yet so far as I am aware this mode has yet to be observed.

Separable kernel

Finally, consider the case of the *separable kernel*,

$$F(\hat{\boldsymbol{k}}, \hat{\boldsymbol{k}}') = L\, w(\hat{\boldsymbol{k}})\, w(\hat{\boldsymbol{k}}') \quad , \tag{9.106}$$

resulting in the eigenvalue equation

$$\delta k_{\rm F}(\hat{\boldsymbol{k}}) = \frac{L\, \hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}\, w(\hat{\boldsymbol{k}})}{\lambda - \hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}} \int \frac{d\hat{\boldsymbol{k}}'}{4\pi}\, w(\hat{\boldsymbol{k}}')\, \delta k_{\rm F}(\hat{\boldsymbol{k}}') \quad . \tag{9.107}$$

Multiplying by $w(\hat{\boldsymbol{k}})$ and integrating, we obtain

$$\int \frac{d\hat{\boldsymbol{k}}}{4\pi} \left(\frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda - \hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}\right) w^2(\hat{\boldsymbol{k}}) = L^{-1} \quad . \tag{9.108}$$

Note that $\lambda = \lambda(\hat{\boldsymbol{q}})$ will in general be a function of direction if the function $w(\hat{\boldsymbol{k}})$ is not isotropic.

9.4 Dynamic Response of the Fermi Liquid

We now restore the driving term $V(\boldsymbol{r}, t) = \delta\hat{V}(\boldsymbol{q}, \omega)\, e^{i(\boldsymbol{q}\cdot\boldsymbol{r} - \omega t)}$, taken to be spin-independent, and solve the inhomogeneous linear equation Eqn. 9.89 at $T = 0$ for $\delta\hat{n}_{\boldsymbol{k},\sigma}(\boldsymbol{q}, \omega)$ in the collisionless limit. The Fourier components of the bulk density are given by

$$\delta\hat{n}(\boldsymbol{q}, \omega) = \int \frac{d^3k}{(2\pi)^3}\, \delta\hat{n}_{\boldsymbol{k},\sigma}(\boldsymbol{q}, \omega) \equiv -\chi(\boldsymbol{q}, \omega)\, \delta\hat{V}(\boldsymbol{q}, \omega) \quad , \tag{9.109}$$

where $\chi(\boldsymbol{q}, \omega)$ is the dynamical density response function, which we first met in chapter 9. We work in the symmetric channel and suppress the symmetry index $\nu = {\rm s}$. The linearized collisionless Landau-Boltzmann equation then takes the form

$$\delta k_{\rm F}(\hat{\boldsymbol{k}}) = \frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda - \hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}} \left\{ \int \frac{d\hat{\boldsymbol{k}}'}{4\pi}\, F(\vartheta_{\hat{\boldsymbol{k}},\hat{\boldsymbol{k}}'})\, \delta k_{\rm F}(\hat{\boldsymbol{k}}') + \frac{\delta\hat{V}(\hat{\boldsymbol{q}}, \omega)}{\hbar v_{\rm F}} \right\} \quad , \tag{9.110}$$

with $\lambda = \omega/qv_{\rm F}$ as before. The density response is related to the Fermi surface distortion according to

$$\delta\hat{n}(\boldsymbol{q},\omega) = \frac{k_{\rm F}^2}{\pi^2}\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\,\delta k_{\rm F}(\hat{\boldsymbol{k}}) \quad . \tag{9.111}$$

Note that $\delta k_{\rm F}(\hat{\boldsymbol{k}})$ is implicity a function of $\boldsymbol{q}$ and ω.

The difficulty in solving the above equation is that the different angular momentum channels don't decouple. However, in the simplified model where the interaction function $F(\vartheta) = F_0$ is isotropic, we can make progress. We then have

$$\delta\hat{n}(\boldsymbol{q},\omega) = \overbrace{\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\left(\frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda-\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}\right)}^{\equiv -G(\lambda)}\left\{F_0\,\delta\hat{n}(\hat{\boldsymbol{q}},\omega) + \frac{k_{\rm F}^2}{\pi^2}\,\frac{\delta\hat{V}(\hat{\boldsymbol{q}},\omega)}{\hbar v_{\rm F}}\right\} \tag{9.112}$$

where

$$G(\lambda) = -\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\left(\frac{\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}{\lambda-\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{k}}}\right) = 1 - \frac{\lambda}{2}\,\ln\!\left(\frac{\lambda+1}{\lambda-1}\right) \quad . \tag{9.113}$$

Thus we find

$$\chi(\boldsymbol{q},\omega) = \frac{g(\varepsilon_{\rm F})\,G(\omega/v_{\rm F}|\boldsymbol{q}|)}{1+F_0\,G(\omega/v_{\rm F}|\boldsymbol{q}|)} \quad . \tag{9.114}$$

Note that the pole of the response function lies at the natural frequency of the FL oscillations, *i.e.* when $1 + F_0\,G(\omega/qv_{\rm F}) = 0$.

Chapter 10

Linear Response of Quantum Systems

10.1 Response and Resonance

10.1.1 Forced damped oscillator

Consider a damped harmonic oscillator subjected to a time-dependent forcing:

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x = f(t) \quad , \tag{10.1}$$

where γ is the damping rate ($\gamma > 0$) and ω_0 is the natural frequency in the absence of damping[1]. We adopt the following convention for the Fourier transform of a function $H(t)$:

$$H(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \hat{H}(\omega)\, e^{-i\omega t} \qquad , \qquad \hat{H}(\omega) = \int_{-\infty}^{\infty} dt\, H(t)\, e^{+i\omega t} \quad . \tag{10.2}$$

Note that if $H(t)$ is a real function, then $\hat{H}(-\omega) = \hat{H}^*(\omega)$. In Fourier space, then, eqn. (10.1) becomes

$$(\omega_0^2 - 2i\gamma\omega - \omega^2)\, \hat{x}(\omega) = \hat{f}(\omega) \quad , \tag{10.3}$$

with the solution

$$\hat{x}(\omega) = \frac{\hat{f}(\omega)}{\omega_0^2 - 2i\gamma\omega - \omega^2} \equiv \hat{\chi}(\omega)\, \hat{f}(\omega) \tag{10.4}$$

where $\hat{\chi}(\omega)$ is the *susceptibility* or *response function*:

$$\hat{\chi}(\omega) = \frac{1}{\omega_0^2 - 2i\gamma\omega - \omega^2} = \frac{-1}{(\omega - \omega_+)(\omega - \omega_-)} \quad , \tag{10.5}$$

[1]Note that $f(t)$ has dimensions of acceleration.

with

$$\omega_\pm = -i\gamma \pm \sqrt{\omega_0^2 - \gamma^2} \quad . \tag{10.6}$$

The complete solution to (10.1) is then

$$x(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \, \frac{\hat{f}(\omega)\, e^{-i\omega t}}{\omega_0^2 - 2i\gamma\omega - \omega^2} + x_{\rm h}(t) \tag{10.7}$$

where $x_{\rm h}(t)$ is the homogeneous solution,

$$x_{\rm h}(t) = A_+ e^{-i\omega_+ t} + A_- e^{-i\omega_- t} \quad . \tag{10.8}$$

Since $\mathrm{Im}\,(\omega_\pm) < 0$, $x_{\rm h}(t)$ is a *transient* which decays in time. The coefficients $A_\pm$ may be chosen to satisfy initial conditions on $x(0)$ and $\dot{x}(0)$, but the system 'loses its memory' of these initial conditions after a finite time, and in steady state all that is left is the inhomogeneous piece, which is completely determined by the forcing.

In the time domain, we can write

$$x(t) = \int_{-\infty}^{\infty} dt' \; \chi(t-t')\, f(t') \qquad , \qquad \chi(s) \equiv \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \; \hat{\chi}(\omega)\, e^{-i\omega s} \quad , \tag{10.9}$$

which brings us to a very important and sensible result:

Claim: The response is *causal*, *i.e.* $\chi(t-t') = 0$ when $t < t'$, provided that $\hat{\chi}(\omega)$ is analytic in the upper half plane of the variable ω.

Proof: Consider eqn. (10.9). Since $\hat{\chi}(\omega)$ is analytic in the upper half plane, closing in the UHP we obtain $\chi(s<0) = 0$.

For our example (10.5), we close in the LHP for $s > 0$ and obtain

$$\begin{aligned} \chi(s>0) &= (-2\pi i) \sum_{\omega \,\in\, \mathrm{LHP}} \mathrm{Res} \left\{ \frac{1}{2\pi} \, \hat{\chi}(\omega)\, e^{-i\omega s} \right\} \\ &= \frac{i e^{-i\omega_+ s}}{\omega_+ - \omega_-} + \frac{i e^{-i\omega_- s}}{\omega_- - \omega_+} \quad , \end{aligned} \tag{10.10}$$

i.e.

$$\chi(s) = \begin{cases} \frac{e^{-\gamma s}}{\sqrt{\omega_0^2 - \gamma^2}} \, \sin\left(\sqrt{\omega_0^2 - \gamma^2}\right) \Theta(s) & \text{if } \omega_0^2 > \gamma^2 \\ \frac{e^{-\gamma s}}{\sqrt{\gamma^2 - \omega_0^2}} \, \sinh\left(\sqrt{\gamma^2 - \omega_0^2}\right) \Theta(s) & \text{if } \omega_0^2 < \gamma^2 \end{cases} \quad , \tag{10.11}$$

where $\Theta(s)$ is the step function: $\Theta(s \geqslant 0) = 1$, $\Theta(s<0) = 0$. Causality simply means that events occuring after the time t cannot influence the state of the system at t. Note that, in general, $\chi(t)$ describes the time-dependent response to a δ-function impulse at $t = 0$.

10.1.2 Energy dissipation

How much work is done by the force $f(t)$? Since the power applied is $P(t) = f(t)\,\dot{x}(t)$, we have

$$\begin{aligned} P(t) &= f(t)\,\frac{d}{dt}\int\limits_{-\infty}^{\infty} dt' \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\ \hat{\chi}(\omega)\,e^{-i\omega(t-t')} f(t') \\ &= f(t)\int\limits_{-\infty}^{\infty} dt' \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\ (-i\omega)\,\hat{\chi}(\omega)\,e^{-i\omega(t-t')} f(t') \\ \Delta E &= \int\limits_{-\infty}^{\infty} dt\,P(t) = \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\,(-i\omega)\,\hat{\chi}(\omega)\left|\hat{f}(\omega)\right|^2 \quad . \end{aligned} \tag{10.12}$$

Separating $\hat{\chi}(\omega)$ into real and imaginary parts,

$$\hat{\chi}(\omega) = \hat{\chi}'(\omega) + i\hat{\chi}''(\omega) \quad , \tag{10.13}$$

we find for our example

$$\begin{aligned} \hat{\chi}'(\omega) &= \frac{\omega_0^2 - \omega^2}{(\omega_0^2-\omega^2)^2 + 4\gamma^2\omega^2} = +\hat{\chi}'(-\omega) \\ \hat{\chi}''(\omega) &= \frac{2\gamma\omega}{(\omega_0^2-\omega^2)^2 + 4\gamma^2\omega^2} = -\hat{\chi}''(-\omega) \quad . \end{aligned} \tag{10.14}$$

The energy dissipated may now be written

$$\Delta E = \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\,\omega\,\hat{\chi}''(\omega)\left|\hat{f}(\omega)\right|^2 \quad . \tag{10.15}$$

The even function $\hat{\chi}'(\omega)$ is called the *reactive* part of the susceptibility; the odd function $\hat{\chi}''(\omega)$ is the *dissipative* part. When experimentalists measure a *lineshape*, they usually are referring to features in $\omega\,\hat{\chi}''(\omega)$, which describes the absorption rate as a function of driving frequency.

10.1.3 Kramers-Kronig relations

Let $\hat{\chi}(z)$ be a complex function of the complex variable z which is analytic in the upper half plane. Then the following integral must vanish,

$$\oint\limits_{\mathcal{C}} \frac{dz}{2\pi i}\,\frac{\hat{\chi}(z)}{z-\zeta} = 0 \quad , \tag{10.16}$$

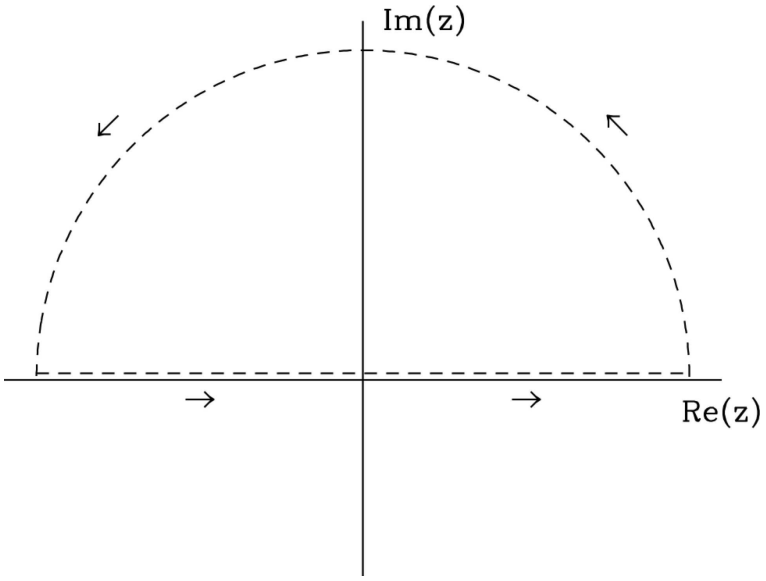


Figure 10.1: The complex integration contour $\mathcal{C}$.

whenever $\mathsf{Im}\,(\zeta) \leqslant 0$, where $\mathcal{C}$ is the contour depicted in fig. 10.1.

Now let $\omega \in \mathbb{R}$ be real, and define the complex function $\hat{\chi}(\omega)$ of the real variable ω by

$$\hat{\chi}(\omega) \equiv \lim_{\epsilon \to 0^+} \hat{\chi}(\omega + i\epsilon) \quad . \tag{10.17}$$

Assuming $\hat{\chi}(z)$ vanishes sufficiently rapidly that Jordan's lemma may be invoked (*i.e.* that the integral of $\hat{\chi}(z)$ along the arc of $\mathcal{C}$ vanishes), we have

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} \frac{d\nu}{2\pi i} \, \frac{\hat{\chi}(\nu)}{\nu - \omega + i\epsilon} \\ &= \int_{-\infty}^{\infty} \frac{d\nu}{2\pi i} \left[\hat{\chi}'(\nu) + i\hat{\chi}''(\nu)\right] \left[\frac{\mathcal{P}}{\nu - \omega} - i\pi\delta(\nu - \omega)\right] \end{aligned} \tag{10.18}$$

where $\mathcal{P}$ stands for 'principal part'. Taking the real and imaginary parts of this equation reveals the *Kramers-Kronig relations*:

$$\hat{\chi}'(\omega) = \mathcal{P}\int_{-\infty}^{\infty} \frac{d\nu}{\pi} \, \frac{\hat{\chi}''(\nu)}{\nu - \omega} \qquad , \qquad \hat{\chi}''(\omega) = -\mathcal{P}\int_{-\infty}^{\infty} \frac{d\nu}{\pi} \, \frac{\hat{\chi}'(\nu)}{\nu - \omega} \quad . \tag{10.19}$$

The Kramers-Kronig relations are valid for any function $\hat{\chi}(z)$ which is analytic in the upper half plane. If $\hat{\chi}(z)$ is analytic everywhere off the $\mathsf{Im}\,(z) = 0$ axis, we may write

$$\hat{\chi}(z) = \int_{-\infty}^{\infty} \frac{d\nu}{\pi} \, \frac{\hat{\chi}''(\nu)}{\nu - z} = -i\,\mathsf{sgn}(\mathsf{Im}\,z) \int_{-\infty}^{\infty} \frac{d\nu}{\pi} \, \frac{\hat{\chi}'(\nu)}{\nu - z} \quad . \tag{10.20}$$

This immediately yields the result

$$\lim_{\epsilon\to 0^+}\left[\hat\chi(\omega+i\epsilon)-\hat\chi(\omega-i\epsilon)\right] = 2i\,\hat\chi''(\omega) \quad . \tag{10.21}$$

As an example, consider the function

$$\hat\chi''(\omega) = \frac{\omega}{\omega^2+\gamma^2} \quad . \tag{10.22}$$

Then, choosing $\gamma > 0$,

$$\hat\chi(z) = \int_{-\infty}^{\infty}\frac{d\omega}{\pi}\,\frac{1}{\omega-z}\cdot\frac{\omega}{\omega^2+\gamma^2} = \begin{cases} +i/(z+i\gamma) & \text{if } \mathsf{Im}\,(z)>0 \\ -i/(z-i\gamma) & \text{if } \mathsf{Im}\,(z)<0 \end{cases} \quad . \tag{10.23}$$

Note that $\hat\chi(z)$ is separately analytic in the UHP and the LHP, but that there is a branch cut along the $\mathsf{Re}\,z$ axis, where $\hat\chi(\omega\pm i\epsilon) = \pm i/(\omega\pm i\gamma)$.

EXERCISE: Show that eqn. (10.21) is satisfied for $\hat\chi(\omega) = \omega/(\omega^2+\gamma^2)$.

If we *analytically continue* $\hat\chi(z)$ from the UHP into the LHP, we find a pole and no branch cut:

$$\tilde{\hat\chi}(z) = \frac{i}{z+i\gamma} \quad . \tag{10.24}$$

The pole lies in the LHP at $z=-i\gamma$.

10.2 Quantum Mechanical Response Functions

10.2.1 First order perturbation theory

Consider a time-dependent quantum system with Hamiltonian $\hat H = \hat H_0 + \hat H_1(t)$, where

$$\hat H_1(t) = -\sum_i \hat Q_i\,\phi_i(t) \quad , \tag{10.25}$$

where each $\hat Q_i$ is an operator labeled by an index i. In continuous systems, the operators may carry spatial labels as well, in which case we would write

$$\hat H(t) = \hat H_0 - \sum_i \int d^dx\,\hat Q_i(\boldsymbol{x})\,\phi_i(\boldsymbol{x},t) \quad . \tag{10.26}$$

The quantities $\phi_i(t)$ or $\phi_i(\boldsymbol{x},t)$ are spatiotemporally varying fields, such as a local scalar potential or local magnetic field. Some examples:

$$\hat{H}_1(t) = \begin{cases} -\hat{\boldsymbol{M}} \cdot \boldsymbol{B}(t) & \text{magnetic moment – magnetic field} \\ +\int d^3x\, \hat{\varrho}(\boldsymbol{x})\, \phi(\boldsymbol{x},t) & \text{charge density – scalar potential} \\ -\frac{1}{c}\int d^3x\, \hat{\boldsymbol{\jmath}}(\boldsymbol{x}) \cdot \boldsymbol{A}(\boldsymbol{x},t) & \text{electromagnetic current – vector potential} \end{cases} \tag{10.27}$$

Let's suppress for now the spatial label $\boldsymbol{x}$, which may be subsumed by the label i if we so desire. The time-dependent expectation value of the operator $\hat{Q}_i$ is given by

$$Q_i(t) \equiv \langle\, \Psi(t)\,|\,\hat{Q}_i\,|\,\Psi(t)\,\rangle \tag{10.28}$$

where $i\hbar\partial_t\,|\Psi(t)\rangle = \hat{H}(t)\,|\Psi(t)\rangle$. Without loss of generality, we assume that the operators $\{\hat{Q}_i\}$ are each defined in such a way that $\langle\Psi_0|\,\hat{Q}_i\,|\Psi_0\rangle = 0$ for all i, where $|\Psi_0\rangle$ is the ground state of H_0, which itself is in general an interacting many-body Hamiltonian. We therefore expect that, to lowest nontrivial order in the fields $\{\phi_i\}$, that the observed response should be linear, *i.e.*

$$Q_i(t) = \int\limits_{-\infty}^{\infty} dt'\,\chi_{ij}(t-t')\,\phi_j(t') + \mathcal{O}(\phi^2) \quad . \tag{10.29}$$

The function $\chi_{ij}(t-t')$ is called a *response function*. It describes how the operator $\hat{Q}_i$ responds at time t to the imposition of a field ϕ_j at time t'. We presume that the responses are all causal, *i.e.* $\chi_{ij}(t-t') = 0$ for $t < t'$. To compute $\chi_{ij}(t-t')$, we will use first order perturbation theory to obtain $\big\langle \hat{Q}_i(t) \big\rangle$ and then functionally differentiate with respect to $\phi_j(t')$:

$$\chi_{ij}(t-t') = \left.\frac{\delta\big\langle \hat{Q}_i(t)\big\rangle}{\delta\phi_j(t')}\right|_{\boldsymbol{\phi}=0} \quad . \tag{10.30}$$

The first step is to establish the result,

$$|\,\Psi(t)\,\rangle = \hat{\mathcal{T}}\,\exp\left\{-\frac{i}{\hbar}\int\limits_{t_0}^{t} dt'\,\Big[\hat{H}_0 + \hat{H}_1(t')\Big]\right\}|\,\Psi(t_0)\,\rangle \quad , \tag{10.31}$$

where $\hat{\mathcal{T}}$ is the *time ordering operator*, which places earlier times to the right. This is easily derived starting with the Schrödinger equation,

$$i\hbar\frac{d}{dt}\,|\,\Psi(t)\,\rangle = \hat{H}(t)|\,\Psi(t)\,\rangle \quad , \tag{10.32}$$

where $\hat{H}(t) = \hat{H}_0 + \hat{H}_1(t)$. Integrating this equation from t to $t+\epsilon$ with infinitesimal ϵ gives

$$|\,\Psi(t+\epsilon)\,\rangle = \left(1 - \frac{i\epsilon}{\hbar}\,\hat{H}(t)\right)|\,\Psi(t)\,\rangle \quad . \tag{10.33}$$

Now let us integrate the Schrödinger equation from $t = t_1$ to $t = t_2$ where $t_1 < t_2$. We have

$$| \Psi(t_2) \rangle = \hat{U}(t_2, t_1) \, | \Psi(t_1) \rangle \tag{10.34}$$

where

$$\begin{aligned} \hat{U}(t_2, t_1) &= \lim_{N \to \infty} \left(1 - \frac{i\epsilon}{\hbar} \, \hat{H}\big(t_1 + (N-1)\epsilon\big) \right) \cdots \left(1 - \frac{i\epsilon}{\hbar} \, \hat{H}(t_1) \right) \\ &\equiv \hat{\mathcal{T}} \exp \left\{ -\frac{i}{\hbar} \int_{t_1}^{t_2} dt \, \hat{H}(t) \right\} \quad , \end{aligned} \tag{10.35}$$

where $\epsilon \equiv (t_2 - t_1)/N$. The operator $\hat{U}(t_2, t_1)$ is unitary operator (*i.e.* $\hat{U}^\dagger = \hat{U}^{-1}$), and is known as the *time evolution operator* between times t_1 and t_2.

EXERCISE: Show that, for $t_1 < t_2 < t_3$ that $\hat{U}(t_3, t_1) = \hat{U}(t_3, t_2) \, \hat{U}(t_2, t_1)$.

If $t_1 < t < t_2$, then differentiating $U(t_2, t_1)$ with respect to $\phi_i(t)$ yields

$$\frac{\delta \hat{U}(t_2, t_1)}{\delta \phi_j(t)} = \frac{i}{\hbar} \, \hat{U}(t_2, t) \, \hat{Q}_j \, \hat{U}(t, t_1) \quad , \tag{10.36}$$

since $\partial \hat{H}(t) / \partial \phi_j(t) = -\hat{Q}_j$. We may therefore write (assuming $t_0 < t, t'$)

$$\begin{aligned} \left. \frac{\delta \, |\Psi(t)\rangle}{\delta \phi_j(t')} \right|_{\{\phi_i = 0\}} &= \frac{i}{\hbar} \, e^{-i\hat{H}_0 (t - t')/\hbar} \, \hat{Q}_j \, e^{-i\hat{H}_0 (t' - t_0)/\hbar} \, | \, \Psi(t_0) \, \rangle \, \Theta(t - t') \\ &= \frac{i}{\hbar} \, e^{-i\hat{H}_0 t/\hbar} \, \hat{Q}_j(t') \, e^{+i\hat{H}_0 \, t_0/\hbar} \, | \, \Psi(t_0) \, \rangle \, \Theta(t - t') \quad , \end{aligned} \tag{10.37}$$

where

$$\hat{Q}_j(t) \equiv e^{i\hat{H}_0 t/\hbar} \, \hat{Q}_j \, e^{-i\hat{H}_0 t/\hbar} \tag{10.38}$$

is the operator Q_j in the time-dependent *interaction representation*. Finally, we have

$$\begin{aligned} \chi_{ij}(t - t') &= \frac{\delta}{\delta \phi_j(t')} \, \langle \, \Psi(t) \, | \, \hat{Q}_i \, | \, \Psi(t) \, \rangle = \frac{\delta \langle \Psi(t)|}{\delta \phi_j(t')} \, \hat{Q}_i \, |\Psi(t)\rangle + \langle \Psi(t)| \, \hat{Q}_i \, \frac{\delta |\Psi(t)\rangle}{\delta \phi_j(t')} \\ &= \Big\{ -\frac{i}{\hbar} \, \langle \, \Psi(t_0) \, | \, e^{-i\hat{H}_0 \, t_0/\hbar} \, \hat{Q}_j(t') \, e^{+i\hat{H}_0 \, t/\hbar} \, \hat{Q}_i \, | \, \Psi(t) \, \rangle \\ &\qquad + \frac{i}{\hbar} \, \langle \, \Psi(t) \, | \, \hat{Q}_i \, e^{-i\hat{H}_0 \, t/\hbar} \, \hat{Q}_j(t') \, e^{+i\hat{H}_0 \, t_0/\hbar} \, | \, \Psi(t_0) \, \rangle \Big\} \, \Theta(t - t') \\ &= \frac{i}{\hbar} \, \big\langle \big[\hat{Q}_i(t), \hat{Q}_j(t') \big] \big\rangle \, \Theta(t - t') \quad , \end{aligned} \tag{10.39}$$

were averages are with respect to the wavefunction $|\,\widetilde{\Psi}_0\,\rangle \equiv \exp(i\hat{H}_0\,t_0/\hbar)\,|\,\Psi(t_0)\,\rangle$, where we take $t_0 \to -\infty$, or, at finite temperature, with respect to a Boltzmann-weighted distribution of such states. To reiterate,

$$\chi_{ij}(t-t') = \frac{i}{\hbar}\,\big\langle\,\big[\hat{Q}_i(t),\hat{Q}_j(t')\big]\,\big\rangle\,\Theta(t-t') \quad . \tag{10.40}$$

This is sometimes known as the *retarded* response function.

10.2.2 Spectral representation

We now derive an expression for the response functions in terms of the spectral properties of the Hamiltonian $\hat{H}_0$, which may describe a fully interacting system. Write $\hat{H}_0\,|\,n\,\rangle = E_n\,|\,n\,\rangle$, in which case[2]

$$\begin{aligned}\hat{\chi}_{ij}(\omega) &= \frac{i}{\hbar}\int\limits_0^\infty dt\;e^{i\omega t}\,\big\langle\,\big[\hat{Q}_i(t),\hat{Q}_j(0)\big]\,\big\rangle \\ &= \frac{i}{\hbar}\int\limits_0^\infty dt\;e^{i\omega t}\sum_{m,n}P_m\,\Big\{\langle\,m\,|\,\hat{Q}_i\,|\,n\,\rangle\,\langle\,n\,|\,\hat{Q}_j\,|\,m\,\rangle\,e^{+i(\omega_m-\omega_n)t} \\ &\qquad\qquad - \langle\,m\,|\,\hat{Q}_j\,|\,n\,\rangle\,\langle\,n\,|\,\hat{Q}_i\,|\,m\,\rangle\,e^{+i(\omega_n-\omega_m)t}\Big\} \quad ,\end{aligned} \tag{10.41}$$

where $\beta = 1/k_{\rm B}T$, $P_m = Z^{-1}\exp(-\beta E_m)$ the Boltzmann weight, with $Z = \mathsf{Tr}\,\exp(-\beta\hat{H}_0)$ the partition function, and the excitation frequencies are defined as $\omega_m \equiv (E_m - E_0)/\hbar$ where E_0 is the ground state energy of $\hat{H}_0$, which, recall, is the Hamiltonian of a fully interacting system. Regularizing the integrals at $t \to \infty$ with $\exp(-\epsilon t)$ with $\epsilon = 0^+$, we use

$$\int\limits_0^\infty dt\;e^{i(\omega-\Omega+i\epsilon)t} = \frac{i}{\omega-\Omega+i\epsilon} \tag{10.42}$$

to obtain the *spectral representation* of the (retarded) response function[3],

$$\hat{\chi}_{ij}(\omega+i\epsilon) = \frac{1}{\hbar}\sum_{m,n}P_m\,\left\{\frac{\langle\,m\,|\,\hat{Q}_j\,|\,n\,\rangle\,\langle\,n\,|\,\hat{Q}_i\,|\,m\,\rangle}{\omega-\omega_m+\omega_n+i\epsilon} - \frac{\langle\,m\,|\,\hat{Q}_i\,|\,n\,\rangle\,\langle\,n\,|\,\hat{Q}_j\,|\,m\,\rangle}{\omega+\omega_m-\omega_n+i\epsilon}\right\} \quad . \tag{10.43}$$

We will refer to this as $\hat{\chi}_{ij}(\omega)$, although formally $\hat{\chi}_{ij}(\omega)$ has poles or a branch cut (for continuous spectra) along the $\mathsf{Re}\,(\omega)$ axis. Note that $\hat{\chi}^*_{ij}(\omega) = \hat{\chi}_{ij}(-\omega)$, which also follows from the fact that

[2]Note that we are using hats to denote operators such as $\hat{H}$ and $\hat{\mathcal{T}}$ as well as Fourier transforms such as $\hat{\chi}_{ij}(\omega)$. Be alert and understand what all the hatted symbols mean!

[3]The spectral representation is sometimes known as the *Lehmann representation*.

$\chi_{ij}(t-t') \in \mathbb{R}$ is a real function of its argument. Diagrammatic perturbation theory does not give us $\hat{\chi}_{ij}(\omega)$, but rather the *time-ordered* response function,

$$\begin{aligned}\chi^{\rm T}_{ij}(t-t') &\equiv \frac{i}{\hbar}\left\langle \hat{\mathcal{T}}\,\hat{Q}_i(t)\,\hat{Q}_j(t')\right\rangle \\ &= \frac{i}{\hbar}\left\langle \hat{Q}_i(t)\,\hat{Q}_j(t')\right\rangle\Theta(t-t') + \frac{i}{\hbar}\left\langle \hat{Q}_j(t')\,\hat{Q}_i(t)\right\rangle\Theta(t'-t) \quad .\end{aligned}$$

The spectral representation of $\hat{\chi}^{\rm T}_{ij}(\omega)$ is

$$\hat{\chi}^{\rm T}_{ij}(\omega+i\epsilon) = \frac{1}{\hbar}\sum_{m,n} P_m\left\{\frac{\langle\, m\,|\,\hat{Q}_j\,|\,n\,\rangle\langle\, n\,|\,\hat{Q}_i\,|\,m\,\rangle}{\omega-\omega_m+\omega_n-i\epsilon} - \frac{\langle\, m\,|\,\hat{Q}_i\,|\,n\,\rangle\langle\, n\,|\,\hat{Q}_j\,|\,m\,\rangle}{\omega+\omega_m-\omega_n+i\epsilon}\right\} \quad . \tag{10.44}$$

The difference between $\hat{\chi}_{ij}(\omega)$ and $\hat{\chi}^{\rm T}_{ij}(\omega)$ is thus only in the sign of the infinitesimal $\pm i\epsilon$ term in one of the denominators.

Let us now define the real and imaginary parts of the product of expectations values encountered above:

$$\langle\, m\,|\,\hat{Q}_i\,|\,n\,\rangle\langle\, n\,|\,\hat{Q}_j\,|\,m\,\rangle \equiv A_{mn}(ij) + iB_{mn}(ij) \quad . \tag{10.45}$$

That is[4],

$$\begin{aligned}A_{mn}(ij) &= \frac{1}{2}\,\langle\, m\,|\,\hat{Q}_i\,|\,n\,\rangle\,\langle\, n\,|\,\hat{Q}_j\,|\,m\,\rangle + \frac{1}{2}\,\langle\, m\,|\,\hat{Q}_j\,|\,n\,\rangle\,\langle\, n\,|\,\hat{Q}_i\,|\,m\,\rangle \\ B_{mn}(ij) &= \frac{1}{2i}\langle\, m\,|\,\hat{Q}_i\,|\,n\,\rangle\,\langle\, n\,|\,\hat{Q}_j\,|\,m\,\rangle - \frac{1}{2i}\langle\, m\,|\,\hat{Q}_j\,|\,n\,\rangle\,\langle\, n\,|\,\hat{Q}_i\,|\,m\,\rangle.\end{aligned} \tag{10.46}$$

Note that $A_{mn}(ij)$ is separately symmetric under interchange of either m and n, or of i and j, whereas $B_{mn}(ij)$ is separately antisymmetric under these operations:

$$\begin{aligned}A_{mn}(ij) &= +A_{nm}(ij) = A_{nm}(ji) = +A_{mn}(ji) \\ B_{mn}(ij) &= -B_{nm}(ij) = B_{nm}(ji) = -B_{mn}(ji) \quad .\end{aligned} \tag{10.47}$$

We define the *spectral densities*

$$\left\{\begin{matrix}\varrho^A_{ij}(\omega)\\ \varrho^B_{ij}(\omega)\end{matrix}\right\} \equiv \hbar^{-1}\sum_{m,n} P_m\left\{\begin{matrix}A_{mn}(ij)\\ B_{mn}(ij)\end{matrix}\right\}\delta(\omega-\omega_n+\omega_m) \quad , \tag{10.48}$$

which satisfy

$$\begin{aligned}\varrho^A_{ij}(\omega) &= +\varrho^A_{ji}(\omega) \quad , \quad \varrho^A_{ij}(-\omega) = +e^{-\beta\hbar\omega}\,\varrho^A_{ij}(\omega) \\ \varrho^B_{ij}(\omega) &= -\varrho^B_{ji}(\omega) \quad , \quad \varrho^B_{ij}(-\omega) = -e^{-\beta\hbar\omega}\,\varrho^B_{ij}(\omega) \quad .\end{aligned} \tag{10.49}$$

[4]We assume all the $\hat{Q}_i$ are Hermitian, *i.e.* $\hat{Q}_i = \hat{Q}_i^\dagger$.

In terms of these spectral densities,

$$\begin{aligned}\hat{\chi}'_{ij}(\omega) &= \mathcal{P}\int_{-\infty}^{\infty} d\nu\, \frac{2\nu}{\nu^2-\omega^2}\,\varrho^A_{ij}(\nu) - \pi(1-e^{-\beta\hbar\omega})\,\varrho^B_{ij}(\omega) = +\hat{\chi}'_{ij}(-\omega) \\ \hat{\chi}''_{ij}(\omega) &= \mathcal{P}\int_{-\infty}^{\infty} d\nu\, \frac{2\omega}{\nu^2-\omega^2}\,\varrho^B_{ij}(\nu) + \pi(1-e^{-\beta\hbar\omega})\,\varrho^A_{ij}(\omega) = -\hat{\chi}''_{ij}(-\omega).\end{aligned} \tag{10.50}$$

For the time ordered response functions, we find

$$\begin{aligned}\hat{\chi}'^{\,\mathrm{T}}_{ij}(\omega) &= \mathcal{P}\int_{-\infty}^{\infty} d\nu\, \frac{2\nu}{\nu^2-\omega^2}\,\varrho^A_{ij}(\nu) - \pi(1+e^{-\beta\hbar\omega})\,\varrho^B_{ij}(\omega) \\ \hat{\chi}''^{\,\mathrm{T}}_{ij}(\omega) &= \mathcal{P}\int_{-\infty}^{\infty} d\nu\, \frac{2\omega}{\nu^2-\omega^2}\,\varrho^B_{ij}(\nu) + \pi(1+e^{-\beta\hbar\omega})\,\varrho^A_{ij}(\omega) \quad .\end{aligned} \tag{10.51}$$

Hence, knowledge of either the retarded or the time-ordered response functions is sufficient to determine the full behavior of the other:

$$\hat{\chi}'_{ij}(\omega) = \frac{e^{\beta\hbar\omega}\,\hat{\chi}'^{\,\mathrm{T}}_{ij}(\omega)}{e^{\beta\hbar\omega}+1} + \frac{\hat{\chi}'^{\,\mathrm{T}}_{ji}(\omega)}{e^{\beta\hbar\omega}+1} \quad , \quad \hat{\chi}''_{ij}(\omega) = \frac{e^{\beta\hbar\omega}\,\hat{\chi}''^{\,\mathrm{T}}_{ij}(\omega)}{e^{\beta\hbar\omega}+1} - \frac{\hat{\chi}''^{\,\mathrm{T}}_{ji}(\omega)}{e^{\beta\hbar\omega}+1} \quad . \tag{10.52}$$

For the diagonal responses, with $i=j$, we then have

$$\hat{\chi}'_{jj}(\omega) = \hat{\chi}'^{\,\mathrm{T}}_{jj}(\omega) \quad , \quad \hat{\chi}''_{jj}(\omega) = \hat{\chi}''^{\,\mathrm{T}}_{jj}(\omega)\,\tanh(\tfrac{1}{2}\beta\hbar\omega) \quad . \tag{10.53}$$

10.2.3 Energy dissipation

The rate at which work is done by the external fields is the power dissipated, and is given by

$$\begin{aligned}P(t) &= \frac{d}{dt}\,\langle\,\Psi(t)\,|\,\hat{H}_0\,|\,\Psi(t)\,\rangle = -\frac{i}{\hbar}\,\langle\,\Psi(t)\,|\,[\hat{H}_0,\hat{H}_1(t)]\,|\,\Psi(t)\,\rangle \\ &= \frac{i}{\hbar}\sum_i \phi_i(t)\,\langle\,\Psi(t)\,|\,[\hat{H}_0,\hat{Q}_i]\,|\,\Psi(t)\,\rangle\end{aligned} \tag{10.54}$$

Now recall $|\,\Psi(t)\,\rangle = \hat{U}(t,t_0)\,|\,\Psi(t_0)\,\rangle$, where we shall take $t_0\to-\infty$. We will evaluate the power dissipated to quadratic order in the fields, and for this we need the expansion of the evolution operator $\hat{U}(t,t_0)$ to linear order in the fields, which is to say to linear order in the perturbation $\hat{H}_1$. From Eqn. 10.35 we have

$$\hat{U}(t,t_0) = \hat{U}_0(t,t_0) - \frac{i}{\hbar}\int_{t_0}^{t} ds\,\hat{U}_0(t,s)\,\hat{H}_1(s)\,\hat{U}_0(s,t_0) + \ldots \tag{10.55}$$

where $\hat{U}_0(t_2,t_1) = e^{-i\hat{H}_0(t_2-t_1)/\hbar}$ and $\hat{H}_1(s) = -\sum_i \hat{Q}_i\,\phi_i(t)$. Thus, we have

$$\begin{aligned} |\,\Psi(t)\,\rangle &= \hat{U}(t,t_0)\,|\,\Psi_0\,\rangle \\ &= e^{-i\hat{H}_0\,t/\hbar}\left\{1 - \frac{i}{\hbar}\int\limits_{t_0}^{t} dt'\,e^{i\hat{H}_0\,t'/\hbar}\hat{H}_1(s)\,e^{-i\hat{H}_0\,t'/\hbar} + \ldots\right\}|\,\widetilde{\Psi}_0\,\rangle \end{aligned} \tag{10.56}$$

Thus, to lowest nontrivial order in the fields,

$$\begin{aligned} P(t) &= \frac{i}{\hbar}\sum_i \phi_i(t)\,\langle\,\Psi(t)\,|\,[\hat{H}_0,\hat{Q}_i]\,|\,\Psi(t)\,\rangle \\ &= -\frac{1}{\hbar^2}\sum_{i,j}\phi_i(t)\int\limits_{t_0}^{t} dt'\,\phi_j(t')\,\langle\,\widetilde{\Psi}_0\,|\,\big[[\hat{H}_0,\hat{Q}_i(t)],\hat{Q}_j(t')\big]\,|\,\widetilde{\Psi}_0\,\rangle \\ &= \frac{i}{\hbar}\sum_{i,j}\phi_i(t)\int\limits_{t_0}^{t} dt'\,\phi_j(t')\,\frac{d}{dt}\,\langle\,\widetilde{\Psi}_0\,|\,[\hat{Q}_i(t),\hat{Q}_j(t')\big]\,|\,\widetilde{\Psi}_0\,\rangle \\ &= \sum_{i,j}\phi_i(t)\int\limits_{-\infty}^{\infty} dt'\,\frac{d}{dt}\,\chi_{ij}(t-t')\,\phi_j(t') = \sum_i \phi_i(t)\,\frac{d\langle\hat{Q}_j(t)\rangle}{dt} \quad . \end{aligned} \tag{10.57}$$

The total energy dissipated is thus a functional of the external fields $\{\phi_i(t)\}$:

$$\begin{aligned} W &= \int\limits_{-\infty}^{\infty} dt\,P(t) = -\int\limits_{-\infty}^{\infty} dt\int\limits_{-\infty}^{\infty} dt'\,\chi_{ij}(t-t')\,\dot{\phi}_i(t)\,\phi_j(t') \\ &= \int\limits_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,(-i\omega)\,\hat{\phi}_i^*(\omega)\,\hat{\chi}_{ij}(\omega)\,\hat{\phi}_j(\omega) \quad , \end{aligned} \tag{10.58}$$

where we now adopt the convention that we sum on the repeated indices i and j. Since the $\{\hat{Q}_i\}$ are Hermitian observables, the $\{\phi_i(t)\}$ must be real fields, in which case their conjugates are given by $\hat{\phi}_i^*(\omega) = \hat{\phi}_j(-\omega)$, whence

$$\begin{aligned} W &= \int\limits_{-\infty}^{\infty}\frac{d\omega}{4\pi}\,(-i\omega)\big[\hat{\chi}_{ij}(\omega) - \hat{\chi}_{ji}(-\omega)\big]\,\hat{\phi}_i^*(\omega)\,\hat{\phi}_j(\omega) \\ &= \int\limits_{-\infty}^{\infty}\frac{d\omega}{2\pi}\,\mathcal{M}_{ij}(\omega)\,\hat{\phi}_i^*(\omega)\,\hat{\phi}_j(\omega) \end{aligned} \tag{10.59}$$

where

$$\begin{aligned}\mathcal{M}_{ij}(\omega) &\equiv \tfrac{1}{2}(-i\omega)\left\{\hat{\chi}_{ij}(\omega) - \hat{\chi}_{ji}(-\omega)\right\} \\ &= \pi\omega\left(1 - e^{-\beta\hbar\omega}\right)\left(\varrho^A_{ij}(\omega) + i\varrho^B_{ij}(\omega)\right) \quad .\end{aligned} \tag{10.60}$$

Note that as a matrix $\mathcal{M}(\omega) = \mathcal{M}^\dagger(\omega)$, so that $\mathcal{M}(\omega)$ has real eigenvalues.

10.2.4 Correlation functions

We define the *correlation function*

$$S_{ij}(t - t') \equiv \frac{1}{2\pi}\left\langle \hat{Q}_i(t)\,\hat{Q}_j(t')\right\rangle \quad , \tag{10.61}$$

which has the spectral representation

$$\begin{aligned}\hat{S}_{ij}(\omega) &= \hbar\varrho^A_{ij}(\omega) + i\hbar\varrho^B_{ij}(\omega) \\ &= \sum_{m,n} P_m \left\langle\, m \,|\, \hat{Q}_i \,|\, n \,\right\rangle \left\langle\, n \,|\, \hat{Q}_j \,|\, m \,\right\rangle \delta(\omega - \omega_n + \omega_m) \quad .\end{aligned} \tag{10.62}$$

Note that

$$\hat{S}_{ij}(-\omega) = e^{-\beta\hbar\omega}\,\hat{S}^*_{ij}(\omega) \qquad , \qquad \hat{S}_{ji}(\omega) = \hat{S}^*_{ij}(\omega) \quad . \tag{10.63}$$

and that

$$\mathcal{M}_{ij}(\omega) = \frac{\omega}{2i}\left\{\hat{\chi}_{ij}(\omega) - \hat{\chi}_{ji}(-\omega)\right\} = \frac{\pi\omega}{\hbar}\left(1 - e^{-\beta\hbar\omega}\right)\hat{S}_{ij}(\omega) \quad . \tag{10.64}$$

This result is known as the *fluctuation-dissipation theorem*, as it relates the equilibrium fluctuations $S_{ij}(\omega)$ to the dissipation kernel $\mathcal{M}_{ij}(\omega)$.

Time Reversal Symmetry

If the operators $\hat{Q}_i$ have a definite symmetry under time reversal, say

$$\hat{\mathcal{T}}\,\hat{Q}_i\hat{\mathcal{T}}^{-1} = \eta_i\,\hat{Q}_i \quad , \tag{10.65}$$

then the correlation function satisfies

$$\hat{S}_{ij}(\omega) = \eta_i\,\eta_j\,\hat{S}_{ji}(\omega) \quad . \tag{10.66}$$

10.2.5 Continuous systems

The indices i and j could contain spatial information as well. Typically we will separate out spatial degrees of freedom, and write

$$S_{ij}(\boldsymbol{x}-\boldsymbol{x}',t-t') = \frac{1}{2\pi}\left\langle \hat{Q}_i(\boldsymbol{x},t)\,\hat{Q}_j(\boldsymbol{x}',t')\right\rangle \quad , \tag{10.67}$$

where we have assumed space and time translation invariance. The Fourier transform is defined as

$$\begin{aligned}\hat{S}_{ij}(\boldsymbol{k},\omega) &= \int\limits_{-\infty}^{\infty} dt \int d^3x\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\, e^{i\omega t}\, S_{ij}(\boldsymbol{x},t) \\ &= \frac{1}{2\pi V}\int\limits_{-\infty}^{\infty} dt\; e^{+i\omega t}\left\langle \hat{Q}_i(\boldsymbol{k},t)\,\hat{Q}_j(-\boldsymbol{k},0)\right\rangle \quad .\end{aligned} \tag{10.68}$$

10.3 A Spin in a Magnetic Field

Consider a $S=\frac{1}{2}$ object in an external field, described by the Hamiltonian

$$\hat{H}_0 = \gamma B_0\, S^z \tag{10.69}$$

with $B_0 > 0$. (Without loss of generality, we can take the DC external field $\boldsymbol{B}_0$ to lie along $\hat{\boldsymbol{z}}$.) The eigenstates are $|\pm\rangle$, with $\omega_\pm = \pm\frac{1}{2}\gamma B_0$. We apply a perturbation,

$$\hat{H}_1(t) = \gamma \boldsymbol{S}\cdot\boldsymbol{B}_1(t) \quad . \tag{10.70}$$

At $T=0$, the susceptibility tensor is

$$\begin{aligned}\chi_{\alpha\beta}(\omega) &= \frac{\gamma^2}{\hbar}\sum_n \left\{ \frac{\langle -|\,S^\beta\,|\,n\,\rangle\langle\,n\,|\,S^\alpha\,|-\rangle}{\omega-\omega_- +\omega_n+i\epsilon} - \frac{\langle -|\,S^\alpha\,|\,n\,\rangle\langle\,n\,|\,S^\beta\,|-\rangle}{\omega+\omega_- -\omega_n+i\epsilon}\right\} \\ &= \frac{\gamma^2}{\hbar}\left\{ \frac{\langle -|\,S^\beta\,|+\rangle\langle+|\,S^\alpha\,|-\rangle}{\omega+\gamma B_0+i\epsilon} - \frac{\langle -|\,S^\alpha\,|+\rangle\langle+|\,S^\beta\,|-\rangle}{\omega-\gamma B_0+i\epsilon}\right\} \quad ,\end{aligned} \tag{10.71}$$

where we have dropped the hat on $\hat{\chi}_{\alpha\beta}(\omega)$ for notational convenience. The only nonzero matrix elements are

$$\chi_{+-}(\omega) = \frac{\hbar\gamma^2}{\omega+\gamma B_0+i\epsilon} \qquad , \qquad \chi_{-+}(\omega) = \frac{-\hbar\gamma^2}{\omega-\gamma B_0+i\epsilon} \quad , \tag{10.72}$$

or, equivalently,

$$\begin{aligned}\chi_{xx}(\omega) &= \tfrac{1}{4}\hbar\gamma^2\left\{\frac{1}{\omega+\gamma B_0+i\epsilon}-\frac{1}{\omega-\gamma B_0+i\epsilon}\right\} = +\chi_{yy}(\omega)\\ \chi_{xy}(\omega) &= \tfrac{i}{4}\hbar\gamma^2\left\{\frac{1}{\omega+\gamma B_0+i\epsilon}+\frac{1}{\omega-\gamma B_0+i\epsilon}\right\} = -\chi_{yx}(\omega) \quad .\end{aligned} \tag{10.73}$$

10.3.1 Bloch Equations

The torque exerted on a magnetic moment $\boldsymbol{\mu}$ by a magnetic field $\boldsymbol{H}$ is $\boldsymbol{N} = \boldsymbol{\mu} \times \boldsymbol{H}$, which is equal to the rate of change of the total angular momentum: $\dot{\boldsymbol{J}} = \boldsymbol{N}$. Since $\boldsymbol{\mu} = \gamma \boldsymbol{J}$, where γ is the gyromagnetic factor, we have $\dot{\boldsymbol{\mu}} = \gamma \boldsymbol{\mu} \times \boldsymbol{H}$. For noninteracting spins, the total magnetic moment, $\boldsymbol{M} = \sum_i \boldsymbol{\mu}_i$ then satisfies

$$\frac{d\boldsymbol{M}}{dt} = \gamma \boldsymbol{M} \times \boldsymbol{H} \quad . \tag{10.74}$$

Now suppose that $\boldsymbol{H} = H_0\,\hat{\boldsymbol{z}} + \boldsymbol{H}_\perp(t)$, where $\hat{\boldsymbol{z}} \cdot \boldsymbol{H}_\perp = 0$. In equilibrium, we have $\boldsymbol{M} = M_0\,\hat{\boldsymbol{z}}$, with $M_0 = \chi_0 H_0$, where χ_0 is the static susceptibility. Phenomenologically, we assume that the relaxation to this equilibrium state is described by a longitudinal and transverse relaxation time, respectively known as T_1 and T_2:

$$\begin{aligned}\dot{M}_x &= \gamma M_y H_z - \gamma M_z H_y - \frac{M_x}{T_2}\\ \dot{M}_y &= \gamma M_z H_x - \gamma M_x H_z - \frac{M_y}{T_2}\\ \dot{M}_z &= \gamma M_x H_y - \gamma M_y H_x - \frac{M_z - M_0}{T_1} \quad .\end{aligned} \tag{10.75}$$

These are known as the *Bloch equations*. Mathematically, they are a set of coupled linear, first order, time-dependent, inhomogeneous equations. These may be recast in the form

$$\dot{M}^\alpha + R_{\alpha\beta} M^\beta = \psi^\alpha \quad , \tag{10.76}$$

with $R_{\alpha\beta}(t) = T^{-1}_{\alpha\beta} - \gamma\,\epsilon_{\alpha\beta\delta} H^\delta(t)$, $\psi^\alpha = T^{-1}_{\alpha\beta}\,M^\beta_0$, and

$$T_{\alpha\beta} = \begin{pmatrix} T_2 & 0 & 0 \\ 0 & T_2 & 0 \\ 0 & 0 & T_1 \end{pmatrix} \quad . \tag{10.77}$$

The formal solution is written

$$\boldsymbol{M}(t) = \int_0^t dt'\, U(t-t')\,\boldsymbol{\psi}(t') + U(t)\,\boldsymbol{\psi}(0) \quad , \tag{10.78}$$

where the evolution matrix,

$$U(t) = \hat{\mathcal{T}}\,\exp\left\{-\int_0^t dt'\, R(t')\right\} \quad , \tag{10.79}$$

is given in terms of the time-ordered exponential (earlier times to the right).

We can make analytical progress if we write write $\boldsymbol{M} = M_0\,\hat{\boldsymbol{z}} + \boldsymbol{m}$ and suppose $|\boldsymbol{H}_\perp| \ll H_0$ and $|\boldsymbol{m}| \ll M_0$, in which case we have

$$\begin{aligned}
\dot m_x &= \gamma\, H_0\, m_y - \gamma\, H_y\, M_0 - \frac{m_x}{T_2} \\
\dot m_y &= \gamma\, H_x\, M_0 - \gamma\, H_0\, m_x - \frac{m_y}{T_2} \\
\dot m_z &= -\frac{m_z}{T_1} \quad ,
\end{aligned} \tag{10.80}$$

which are equivalent to the following:

$$\begin{aligned}
\ddot m_x + 2T_2^{-1}\dot m_x + \big(\gamma^2 H_0^2 + T_2^{-2}\big) m_x &= \gamma\, M_0\big(\gamma\, H_0 H_x - T_2^{-1} H_y - \dot H_y\big) \\
\ddot m_y + 2T_2^{-1}\dot m_y + \big(\gamma^2 H_0^2 + T_2^{-2}\big) m_y &= \gamma\, M_0\big(\gamma\, H_0 H_y + T_2^{-1} H_x + \dot H_x\big)
\end{aligned} \tag{10.81}$$

and $m_z(t) = m_z(0)\,\exp(-t/T_1)$. Solving the first two by Fourier transform,

$$\begin{aligned}
\big(\gamma^2 H_0^2 + T_2^{-2} - \omega^2 - 2iT_2^{-2}\omega\big)\,\hat m_x(\omega) &= \gamma\, M_0\big(\gamma\, H_0\, H_x(\omega) + (i\omega - T_2^{-1}) H_y(\omega)\big) \\
\big(\gamma^2 H_0^2 + T_2^{-2} - \omega^2 - 2iT_2^{-2}\omega\big)\,\hat m_y(\omega) &= \gamma\, M_0\big(\gamma\, H_0\, H_y(\omega) - (i\omega - T_2^{-1}) H_x(\omega)\big) \quad ,
\end{aligned} \tag{10.82}$$

from which we read off

$$\begin{aligned}
\chi_{xx}(\omega) &= \frac{\gamma^2 H_0\, M_0}{\gamma^2 H_0^2 + T_2^{-2} - \omega^2 - 2iT_2^{-1}\omega} = \chi_{yy}(\omega) \\
\chi_{xy}(\omega) &= \frac{(i\omega - T_2^{-1})\,\gamma\, M_0}{\gamma^2 H_0^2 + T_2^{-2} - \omega^2 - 2iT_2^{-1}\omega} = -\chi_{yx}(\omega) \quad .
\end{aligned} \tag{10.83}$$

Note that Onsager reciprocity is satisfied:

$$\chi_{xy}(\omega, H_0) = \chi^{\mathsf{T}}_{yx}(\omega, H_0) = \chi_{yx}(\omega, -H_0) = -\chi_{yx}(\omega, H_0) \quad . \tag{10.84}$$

The lineshape is given by

$$\begin{aligned} \chi'_{xx}(\omega) &= \frac{(\gamma^2 H_0^2 + T_2^{-2} - \omega^2)\,\gamma^2\, H_0\, M_0}{\left(\gamma^2 H_0^2 + T_2^{-2} - \omega^2\right)^2 + 4\, T_2^{-2}\omega^2} \\ \chi''_{xx}(\omega) &= \frac{2\,\gamma\, H_0\, M_0\, T_2^{-1}\omega}{\left(\gamma^2 H_0^2 + T_2^{-2} - \omega^2\right)^2 + 4\, T_2^{-2}\omega^2} \quad , \end{aligned} \tag{10.85}$$

so a measure of the linewidth is a measure of T_2^{-1}.

10.4 Density Response and Correlations

In many systems, external probes couple to the number density $\hat{n}(\boldsymbol{x}) = \sum_{i=1}^N \delta(\boldsymbol{x} - \boldsymbol{x}_i)$, and we may write the perturbing Hamiltonian as

$$\hat{H}_1(t) = -\int d^3x\; \hat{n}(\boldsymbol{x})\, \phi(\boldsymbol{x}, t) \quad . \tag{10.86}$$

The response $\delta n \equiv n - \langle n \rangle_0$ is given by

$$\begin{aligned} \langle \delta n(\boldsymbol{x}, t) \rangle &= \int dt' \int d^3x'\; \chi(\boldsymbol{x} - \boldsymbol{x}', t - t')\, \phi(\boldsymbol{x}', t') \\ \langle \delta \hat{n}(\boldsymbol{q}, \omega) \rangle &= \hat{\chi}(\boldsymbol{q}, \omega)\, \hat{\phi}(\boldsymbol{q}, \omega) \quad , \end{aligned} \tag{10.87}$$

where

$$\begin{aligned} \hat{\chi}(\boldsymbol{q}, \omega) &= \frac{1}{\hbar V} \sum_{m,n} P_m \left\{ \frac{\left| \langle\, m \,|\, \hat{n}_{\boldsymbol{q}} \,|\, n \,\rangle \right|^2}{\omega - \omega_m + \omega_n + i\epsilon} - \frac{\left| \langle\, m \,|\, \hat{n}_{\boldsymbol{q}} \,|\, n \,\rangle \right|^2}{\omega + \omega_m - \omega_n + i\epsilon} \right\} \\ &= \frac{1}{\hbar} \int\limits_{-\infty}^{\infty} d\nu\; S(\boldsymbol{q}, \nu) \left\{ \frac{1}{\omega + \nu + i\epsilon} - \frac{1}{\omega - \nu + i\epsilon} \right\} \quad . \end{aligned} \tag{10.88}$$

The function

$$S(\boldsymbol{q}, \omega) = \frac{1}{V} \sum_{m,n} P_m \left| \langle\, m \,|\, \hat{n}_{\boldsymbol{q}} \,|\, n \,\rangle \right|^2 \delta(\omega - \omega_n + \omega_m) \tag{10.89}$$

is known as the *dynamic structure factor* (dsf). Note that $\hat{n}_{\boldsymbol{q}} = \sum_{i=1}^N e^{-i\boldsymbol{q}\cdot\boldsymbol{x}_i}$ and that $\hat{n}^\dagger_{\boldsymbol{q}} = \hat{n}_{-\boldsymbol{q}}$. In a scattering experiment, where an incident probe (*e.g.* a neutron) interacts with the system

via a potential $\phi(\boldsymbol{x} - \boldsymbol{R})$, where $\boldsymbol{R}$ is the probe particle position, Fermi's Golden Rule says that the rate at which the incident particle deposits momentum $\hbar\boldsymbol{q}$ and energy $\hbar\omega$ into the system is given by

$$\begin{aligned}\mathcal{I}(\boldsymbol{q},\omega) &= \frac{2\pi}{\hbar}\sum_{m,n} P_m \left|\left\langle\, m\,;\boldsymbol{p}\,\right|\hat{H}_1\left|\,n\,;\boldsymbol{p}-\hbar\boldsymbol{q}\,\right\rangle\right|^2 \delta(\omega-\omega_n+\omega_m) \\ &= \frac{2\pi}{\hbar}\,|\hat{\phi}(\boldsymbol{q})|^2\,S(\boldsymbol{q},\omega) \quad .\end{aligned} \tag{10.90}$$

The quantity $|\hat{\phi}(\boldsymbol{q})|^2$ is called the *form factor*. In neutron scattering, the "on-shell" condition requires that the incident energy ε and momentum $\boldsymbol{p}$ are related via the ballistic dispersion $\varepsilon = \boldsymbol{p}^2/2m_{\rm n}$. Similarly, the final energy and momentum are related, hence

$$\varepsilon - \hbar\omega = \frac{\boldsymbol{p}^2}{2m_{\rm n}} - \hbar\omega = \frac{(\boldsymbol{p}-\hbar\boldsymbol{q})^2}{2m_{\rm n}} \qquad \Longrightarrow \qquad \hbar\omega = \frac{\hbar\boldsymbol{q}\cdot\boldsymbol{p}}{m_{\rm n}} - \frac{\hbar^2\boldsymbol{q}^2}{2m_{\rm n}} \quad . \tag{10.91}$$

Hence, for fixed momentum transfer $\hbar\boldsymbol{q}$, the frequency ω can be varied by changing the incident momentum $\boldsymbol{p}$.

Another case of interest is the response of a system to a foreign object moving with trajectory $\boldsymbol{R}(t) = \boldsymbol{V}t$. In this case, $\phi(\boldsymbol{x},t) = \phi\big(\boldsymbol{x} - \boldsymbol{R}(t)\big)$, and

$$\hat{\phi}(\boldsymbol{q},\omega) = \int\!dt \int\!d^3x\, e^{-i\boldsymbol{q}\cdot\boldsymbol{x}}\, e^{i\omega t}\,\phi(\boldsymbol{x}-\boldsymbol{V}t) = 2\pi\,\delta(\omega-\boldsymbol{q}\cdot\boldsymbol{V})\,\hat{\phi}(\boldsymbol{q}) \tag{10.92}$$

so that

$$\langle \delta n(\boldsymbol{q},\omega)\rangle = 2\pi\,\delta(\omega-\boldsymbol{q}\cdot\boldsymbol{V})\,\hat{\chi}(\boldsymbol{q},\omega)\,\hat{\phi}(\boldsymbol{q}) \quad . \tag{10.93}$$

10.4.1 Sum rules

From eqn. (10.89) we find

$$\begin{aligned}\int_{-\infty}^{\infty}\!d\omega\,\omega\,S(\boldsymbol{q},\omega) &= \frac{1}{V}\sum_{m,n} P_m\left|\left\langle\, m\,\right|\hat{n}_{\boldsymbol{q}}\left|\,n\,\right\rangle\right|^2(\omega_n-\omega_m) \\ &= \frac{1}{\hbar V}\sum_{m,n} P_m\left\langle\, m\,\right|\hat{n}_{\boldsymbol{q}}\left|\,n\,\right\rangle\left\langle\, n\,\right|[\hat{H},\hat{n}^\dagger_{\boldsymbol{q}}]\left|\,m\,\right\rangle \\ &= \frac{1}{\hbar V}\left\langle \hat{n}_{\boldsymbol{q}}\,[\hat{H},\hat{n}^\dagger_{\boldsymbol{q}}]\right\rangle = \frac{1}{2\hbar V}\left\langle\big[\hat{n}_{\boldsymbol{q}},[\hat{H},\hat{n}^\dagger_{\boldsymbol{q}}]\big]\right\rangle \quad ,\end{aligned} \tag{10.94}$$

where the last equality is guaranteed by $\boldsymbol{q} \to -\boldsymbol{q}$ symmetry. Now if the potential is velocity independent, *i.e.* if

$$\hat{H} = -\frac{\hbar^2}{2m}\sum_{i=1}^{N}\boldsymbol{\nabla}_i^2 + V(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) \quad , \tag{10.95}$$

then with $\hat{n}^\dagger_{\boldsymbol{q}} = \sum_{i=1}^N e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i}$ we obtain

$$\begin{aligned}
[\hat{H}, \hat{n}^\dagger_{\boldsymbol{q}}] &= -\frac{\hbar^2}{2m}\sum_{i=1}^N \left[\boldsymbol{\nabla}_i^2\,,\, e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i}\right] = \frac{\hbar^2}{2im}\,\boldsymbol{q}\cdot\sum_{i=1}^N\Big(\boldsymbol{\nabla}_i\, e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i} + e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i}\,\boldsymbol{\nabla}_i\Big) \\
\big[\hat{n}_{\boldsymbol{q}}, [\hat{H}, \hat{n}^\dagger_{\boldsymbol{q}}]\big] &= \frac{\hbar^2}{2im}\,\boldsymbol{q}\cdot\sum_{i=1}^N\sum_{j=1}^N\Big[e^{-i\boldsymbol{q}\cdot\boldsymbol{x}_j}, \boldsymbol{\nabla}_i\, e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i} + e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i}\,\boldsymbol{\nabla}_i\Big] = \frac{N\hbar^2\boldsymbol{q}^2}{m} \quad .
\end{aligned} \tag{10.96}$$

We have derived the *f-sum rule*:

$$\int_{-\infty}^{\infty} d\omega\,\omega\,S(\boldsymbol{q},\omega) = \frac{n\hbar\boldsymbol{q}^2}{2m} \quad , \tag{10.97}$$

where $n = N/V$ is the overall number density. Note that this integral, which is the first moment of the structure factor, is *independent of the potential!*. The $n^{\rm th}$ moment of the dsf distribution is given by

$$\int_{-\infty}^{\infty} d\omega\,\omega^n\,S(\boldsymbol{q},\omega) = \frac{1}{\hbar V}\left\langle \hat{n}_{\boldsymbol{q}}\Big[\overbrace{\hat{H}, \big[\hat{H}, \cdots [\hat{H}}^{n\ \text{times}}, \hat{n}^\dagger_{\boldsymbol{q}}]\cdots\big]\Big]\right\rangle \quad . \tag{10.98}$$

Moments with $n \neq 1$ in general depend on the potential, unlike the $n = 1$ moment from the f-sum rule. The $n = 0$ moment gives

$$\begin{aligned}
S(\boldsymbol{q}) \equiv \int_{-\infty}^{\infty} d\omega\, S(\boldsymbol{q},\omega) &= \frac{1}{\hbar V}\left\langle \hat{n}^\dagger_{\boldsymbol{q}}\,\hat{n}_{\boldsymbol{q}}\right\rangle \\
&= \frac{1}{\hbar}\int d^3x\,\left\langle n(\boldsymbol{x})\,n(0)\right\rangle e^{-i\boldsymbol{q}\cdot\boldsymbol{x}} \quad ,
\end{aligned} \tag{10.99}$$

which is the Fourier transform of the density-density correlation function.

Compressibility sum rule

The isothermal compressibility is given by

$$\kappa_T = -\frac{1}{V}\frac{\partial V}{\partial n}\bigg|_T = \frac{1}{n^2}\frac{\partial n}{\partial \mu}\bigg|_T \quad . \tag{10.100}$$

Since a constant potential $v(\boldsymbol{x},t)$ is equivalent to a chemical potential shift, we have

$$\langle \delta n\rangle = \hat{\chi}(\boldsymbol{q}=0,\omega=0)\,\delta\mu \quad \Longrightarrow \quad \kappa_T = \frac{2}{\hbar n^2}\lim_{\boldsymbol{q}\to 0}\int_{-\infty}^{\infty} d\omega\,\frac{S(\boldsymbol{q},\omega)}{\omega} \quad . \tag{10.101}$$

This is known as the *compressibility sum rule*.

Single mode approximation at $T = 0$

For each wavevector $\boldsymbol{q}$, the dynamical structure factor $S(\boldsymbol{q},\omega)$ may be regarded as a distribution function of the frequency ω. The normalized average $n^{\rm th}$ moment $\langle\omega^n\rangle_{\boldsymbol{q}}$ is given by

$$\langle\omega^n\rangle_{\boldsymbol{q}} = \int\limits_{-\infty}^{\infty} d\omega\,\omega^n\,S(\boldsymbol{q},\omega)\Big/\int\limits_{-\infty}^{\infty} d\omega\,S(\boldsymbol{q},\omega) \quad . \tag{10.102}$$

Thus, we can define a set of quantities $\big\{\Omega_{n,m}(\boldsymbol{q})\big\}$, each of which has dimensions of frequency, according to

$$\big[\Omega_{n,m}(\boldsymbol{q})\big]^{n-m} \equiv \frac{\langle\omega^n\rangle_{\boldsymbol{q}}}{\langle\omega^m\rangle_{\boldsymbol{q}}} = \int\limits_{-\infty}^{\infty} d\omega\,\omega^n\,S(\boldsymbol{q},\omega)\Big/\int\limits_{-\infty}^{\infty} d\omega\,\omega^m\,S(\boldsymbol{q},\omega) \quad . \tag{10.103}$$

If, for each wavevector $\boldsymbol{q}$, all the oscillator strength $|\langle\,\mathsf{G}\,|\,\hat{n}_{\boldsymbol{q}}\,|\,n\,\rangle|^2$ in the dsf is saturated by a single mode, then

$$S(\boldsymbol{q},\omega) \approx S_{\sf SMA}(\boldsymbol{q},\omega) \equiv S_{\sf SMA}(\boldsymbol{q})\,\delta\big(\omega - \Omega_{\sf SMA}(\boldsymbol{q})\big) \quad . \tag{10.104}$$

If the SMA is exact, then $S_{\sf SMA}(\boldsymbol{q})$ is the static structure factor $S(\boldsymbol{q})$ from Eqn. 10.99. Within the SMA, the $n^{\rm th}$ moment of the dsf in Eqn. 10.98 is $\big[\Omega_{\sf SMA}(\boldsymbol{q})\big]^n \cdot S_{\sf SMA}(\boldsymbol{q})$, in which case each $\Omega_{n,m}(\boldsymbol{q})$ from Eqn. 10.103 is given by $\Omega_{\sf SMA}(\boldsymbol{q})$. Thus, the SMA frequency may be approximated by $\Omega_{n,m}(\boldsymbol{q})$ for any n and m. For example, if we take $n = +1$ and $m = -1$, we have

$$\Omega^2_{1,-1}(\boldsymbol{q}) = \frac{n\boldsymbol{q}^2}{m\hat{\chi}(\boldsymbol{q})} \quad , \tag{10.105}$$

where

$$\hat{\chi}(\boldsymbol{q}) = \frac{2}{\hbar}\int\limits_0^{\infty} d\omega\,\frac{S(\boldsymbol{q},\omega)}{\omega} \tag{10.106}$$

is the static susceptibility. If instead we were to choose $n = 1$ and $m = 0$, we arrive at

$$\Omega_{1,0}(\boldsymbol{q}) = \frac{n\hbar\boldsymbol{q}^2}{2S(\boldsymbol{q})} \quad . \tag{10.107}$$

Note that within the SMA, we have $\hat{\chi}(\boldsymbol{q}) \approx \hat{\chi}_{\sf SMA}(\boldsymbol{q}) = 2S_{\sf SMA}(\boldsymbol{q})/\hbar\Omega_{\sf SMA}(\boldsymbol{q})$.

10.4.2 Dynamic Structure Factor for the Electron Gas

The dynamic structure factor $S(\boldsymbol{q},\omega)$ tells us about the spectrum of density fluctuations. The density operator $\hat{n}^\dagger_{\boldsymbol{q}} = \sum_i e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i}$ increases the wavevector by $\boldsymbol{q}$. At $T = 0$, in order for $\langle\,n\,|\,\hat{n}^\dagger_{\boldsymbol{q}}\,|\,\mathsf{G}\,\rangle$

to be nonzero (where $|\mathsf{G}\rangle$ is the ground state, *i.e.* the filled Fermi sphere), the state n must correspond to a *particle-hole excitation*. For a given $\boldsymbol{q}$, the maximum excitation frequency is obtained by taking an electron just inside the Fermi sphere, with wavevector $\boldsymbol{k} = k_{\mathsf{F}}\,\hat{\boldsymbol{q}}$ and transferring it to a state outside the Fermi sphere with wavevector $\boldsymbol{k} + \boldsymbol{q}$. For $|\boldsymbol{q}| < 2k_{\mathsf{F}}$, the minimum excitation frequency is zero – one can always form particle-hole excitations with states adjacent to the Fermi sphere. For $|\boldsymbol{q}| > 2k_{\mathsf{F}}$, the minimum excitation frequency is obtained by taking an electron just inside the Fermi sphere with wavevector $\boldsymbol{k} = -k_{\mathsf{F}}\,\hat{\boldsymbol{q}}$ to an unfilled state outside the Fermi sphere with wavevector $\boldsymbol{k} + \boldsymbol{q}$. These cases are depicted in fig. 10.4.

We therefore have

$$\omega_{\mathsf{max}}(q) = \frac{\hbar q^2}{2m} + \frac{\hbar k_{\mathsf{F}} q}{m} \tag{10.108}$$

and

$$\omega_{\mathrm{min}}(q) = \begin{cases} 0 & \text{if } q \leqslant 2k_{\mathsf{F}} \\ \frac{\hbar q^2}{2m} - \frac{\hbar k_{\mathsf{F}} q}{m} & \text{if } q > 2k_{\mathsf{F}} \end{cases} \quad . \tag{10.109}$$

This is depicted in fig. 10.2. Outside of the region bounded by $\omega_{\mathrm{min}}(q)$ and $\omega_{\mathsf{max}}(q)$, there are no single pair excitations. It is of course easy to create *multiple pair* excitations with arbitrary energy and momentum, as depicted in fig. 10.3. However, these multipair states do not couple to the ground state $|\,\mathsf{G}\,\rangle$ through a single application of the density operator $\hat{n}^\dagger_{\boldsymbol{q}'}$, hence they have zero oscillator strength: $\langle\, n\,|\,\hat{n}^\dagger_{\boldsymbol{q}}\,|\,\mathsf{G}\,\rangle = 0$ for any multipair state $|\,n\,\rangle$.

Explicit $T = 0$ calculation

We start with

$$\begin{aligned} 2\pi\, S(\boldsymbol{x}, t) &= \langle n(\boldsymbol{x}, t)\, n(0,0)\rangle \\ &= \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3}\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \sum_{i,j} \left\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i(t)}\, e^{i\boldsymbol{k}'\cdot\boldsymbol{x}_j} \right\rangle \quad . \end{aligned} \tag{10.110}$$

The time evolution of the operator $\boldsymbol{x}_i(t)$ is given by $\boldsymbol{x}_i(t) = \boldsymbol{x}_i + \boldsymbol{p}_i t/m$, where $\boldsymbol{p}_i = -i\hbar\boldsymbol{\nabla}_i$. Using the result

$$e^{A+B} = e^A\, e^B\, e^{-\frac{1}{2}[A,B]} \quad , \tag{10.111}$$

which is valid when $[A,[A,B]] = [B,[A,B]] = 0$, we have

$$e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i(t)} = e^{i\hbar\boldsymbol{k}^2 t/2m}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{p}_i\, t/m} \quad , \tag{10.112}$$

hence

$$2\pi\, S(\boldsymbol{x}, t) = \int \frac{d^3k}{(2\pi)^3} \int \frac{d^3k'}{(2\pi)^3}\, e^{i\hbar\boldsymbol{k}^2 t/2m}\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \sum_{i,j} \left\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i}\, e^{i\boldsymbol{k}\cdot\boldsymbol{p}_i t/m}\, e^{i\boldsymbol{k}'\cdot\boldsymbol{x}_j} \right\rangle \quad . \tag{10.113}$$

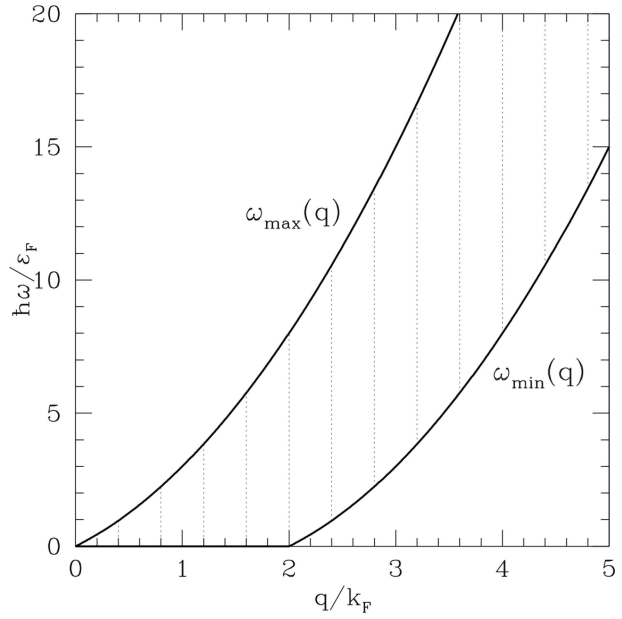


Figure 10.2: Spread of particle-hole excitation frequencies ω in units of $\varepsilon_{\rm F}/\hbar$ *versus* wavevector q in units of $k_{\rm F}$. Outside the hatched areas, there are no *single pair* excitations.

We now break the sum up into diagonal $(i = j)$ and off-diagonal $(i \neq j)$ terms.

For the diagonal terms, with $i = j$, we have

$$\begin{aligned}\left\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{p}_i t/m}\, e^{i\boldsymbol{k}'\cdot\boldsymbol{x}_i}\right\rangle &= e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{k}'t/m}\left\langle e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{x}_i}\, e^{i\boldsymbol{k}\cdot\boldsymbol{p}_i t/m}\right\rangle \\ &= e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{k}'t/m}\,\frac{(2\pi)^3}{NV}\,\delta(\boldsymbol{k}-\boldsymbol{k}')\sum_{\boldsymbol{q}}\Theta(k_{\rm F}-q)\,e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m} \quad ,\end{aligned} \tag{10.114}$$

since the ground state $|\,\mathsf{G}\,\rangle$ is a Slater determinant formed of single particle wavefunctions $\psi_{\boldsymbol{k}}(\boldsymbol{x}) = \exp(i\boldsymbol{q}\cdot\boldsymbol{x})/\sqrt{V}$ with $q < k_{\rm F}$.

For $i \neq j$, we must include exchange effects. We then have

$$\begin{aligned}\left\langle e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{p}_i t/m}\, e^{i\boldsymbol{k}'\cdot\boldsymbol{x}_j}\right\rangle &= \frac{1}{N(N-1)}\sum_{\boldsymbol{q}}\sum_{\boldsymbol{q}'}\Theta(k_{\rm F}-q)\,\Theta(k_{\rm F}-q') \\ &\qquad\times\int\frac{d^3x_i}{V}\int\frac{d^3x_j}{V}\,e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m}\left\{e^{-i\boldsymbol{k}\cdot\boldsymbol{x}_i}\,e^{i\boldsymbol{k}'\boldsymbol{x}_j} - e^{i(\boldsymbol{q}-\boldsymbol{q}'-\boldsymbol{k})\cdot\boldsymbol{x}_i}\,e^{i(\boldsymbol{q}'-\boldsymbol{q}+\boldsymbol{k}')\cdot\boldsymbol{x}_j}\right\} \\ &= \frac{(2\pi)^6}{N(N-1)V^2}\sum_{\boldsymbol{q}}\sum_{\boldsymbol{q}'}\Theta(k_{\rm F}-q)\,\Theta(k_{\rm F}-q') \\ &\qquad\times e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m}\Big\{\delta(\boldsymbol{k})\,\delta(\boldsymbol{k}') - \delta(\boldsymbol{k}-\boldsymbol{k}')\,\delta(\boldsymbol{k}+\boldsymbol{q}'-\boldsymbol{q})\Big\} \quad .\end{aligned} \tag{10.115}$$

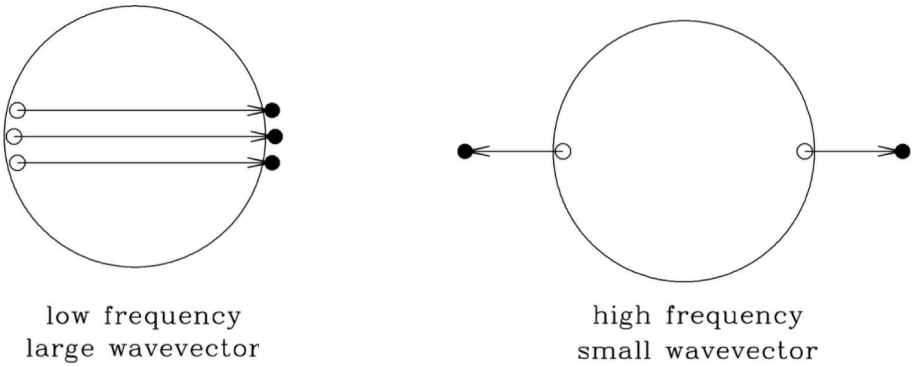


Figure 10.3: With multiple pair excitations, every part of $(\boldsymbol{q},\omega)$ space is accessible. However, these states to not couple to the ground state $\big|\mathsf{G}\big\rangle$ through a *single* application of the density operator $\hat{n}^\dagger_{\boldsymbol{q}}$.

Summing over the $i=j$ terms gives

$$2\pi\,S_{\rm diag}(\boldsymbol{x},t) = \int\frac{d^3k}{(2\pi)^3}\;e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,e^{-i\hbar\boldsymbol{k}^2t/2m}\int\frac{d^3q}{(2\pi)^3}\,\Theta(k_{\rm F}-q)\,e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m} \quad , \tag{10.116}$$

while the off-diagonal terms yield

$$\begin{aligned}2\pi\,S_{\rm off-diag}(\boldsymbol{x},t) &= \int\frac{d^3k}{(2\pi)^3}\;e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\int\frac{d^3q}{(2\pi)^3}\int\frac{d^3q'}{(2\pi)^3}\,\Theta(k_{\rm F}-q)\,\Theta(k_{\rm F}-q') \\ &\qquad\qquad\times(2\pi)^3\Big\{\delta(\boldsymbol{k})-e^{+i\hbar\boldsymbol{k}^2t/2m}\,e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m}\,\delta(\boldsymbol{q}-\boldsymbol{q}'-\boldsymbol{k})\Big\} \\ &= n^2-\int\frac{d^3k}{(2\pi)^3}\;e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,e^{+i\hbar\boldsymbol{k}^2t/2m}\int\frac{d^3q}{(2\pi)^3}\Theta(k_{\rm F}-q)\,\Theta(k_{\rm F}-|\boldsymbol{k}-\boldsymbol{q}|)\,e^{-i\hbar\boldsymbol{k}\cdot\boldsymbol{q}t/m} \quad ,\end{aligned} \tag{10.117}$$

and hence

$$\begin{aligned}2\pi\,S(\boldsymbol{k},\omega) &= n^2\,(2\pi)^4\delta(\boldsymbol{k})\,\delta(\omega)+\int\frac{d^3q}{(2\pi)^3}\,\Theta(k_{\rm F}-q)\Big\{2\pi\,\delta\Big(\omega-\frac{\hbar\boldsymbol{k}^2}{2m}-\frac{\hbar\boldsymbol{k}\cdot\boldsymbol{q}}{m}\Big) \\ &\qquad\qquad-\Theta(k_{\rm F}-|\boldsymbol{k}-\boldsymbol{q}|)\,2\pi\delta\Big(\omega+\frac{\hbar\boldsymbol{k}^2}{2m}-\frac{\hbar\boldsymbol{k}\cdot\boldsymbol{q}}{m}\Big)\Big\} \\ &= (2\pi)^4n^2\delta(\boldsymbol{k})\delta(\omega)+\int\frac{d^3q}{(2\pi)^3}\,\Theta(k_{\rm F}-q)\,\Theta(|\boldsymbol{k}+\boldsymbol{q}|-k_{\rm F})\cdot 2\pi\delta\Big(\omega-\frac{\hbar\boldsymbol{k}^2}{2m}-\frac{\hbar\boldsymbol{k}\cdot\boldsymbol{q}}{m}\Big).\end{aligned} \tag{10.118}$$

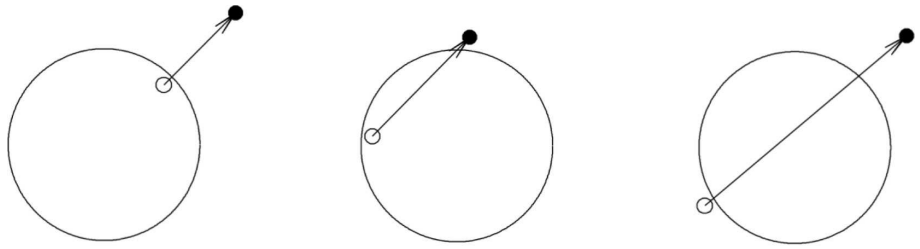


Figure 10.4: Minimum and maximum frequency particle-hole excitations in the free electron gas at $T = 0$. (a) To construct a maximum frequency excitation for a given $\boldsymbol{q}$, create a hole just inside the Fermi sphere at $\boldsymbol{k} = k_{\rm F}\,\hat{\boldsymbol{q}}$ and an electron at $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q}$. (b) For $|\boldsymbol{q}| < 2k_{\rm F}$ the minumum excitation frequency is zero. (c) For $|\boldsymbol{q}| > 2k_{\rm F}$, the minimum excitation frequency is obtained by placing a hole at $\boldsymbol{k} = -k_{\rm F}\,\hat{\boldsymbol{q}}$ and an electron at $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q}$.

For $\boldsymbol{k}, \omega \neq 0$, then,

$$\begin{aligned} 2\pi\, S(\boldsymbol{k},\omega) &= \frac{1}{2\pi}\int\limits_0^{k_{\rm F}} dq\, q^2 \int\limits_{-1}^{1} dx\, \Theta\big(\sqrt{k^2+q^2+2kqx}-k_{\rm F}\big)\,\delta\Big(\omega - \frac{\hbar k^2}{2m} - \frac{\hbar k q}{m}\,x\Big) \\ &= \frac{m}{2\pi\hbar k}\int\limits_0^{k_{\rm F}} dq\, q\,\Theta\Big(\sqrt{q^2+\frac{2m\omega}{\hbar}}-k_{\rm F}\Big)\int\limits_{-1}^{1} dx\,\delta\Big(x+\frac{k}{2q}-\frac{m\omega}{\hbar k q}\Big) \\ &= \frac{m}{4\pi\hbar k}\int\limits_0^{k_{\rm F}^2} du\,\Theta\Big(u+\frac{2m\omega}{\hbar}-k_{\rm F}^2\Big)\,\Theta\Big(u-\Big|\frac{k}{2}-\frac{m\omega}{\hbar k}\Big|^2\Big) \quad . \end{aligned} \tag{10.119}$$

The constraints on u are

$$k_{\rm F}^2 \geqslant u \geqslant \max\left(k_{\rm F}^2 - \frac{2m\omega}{\hbar}\,,\,\Big|\frac{k}{2}-\frac{m\omega}{\hbar k}\Big|^2\right) \quad . \tag{10.120}$$

Clearly $\omega > 0$ is required. There are two cases to consider.

The first case is

$$k_{\rm F}^2 - \frac{2m\omega}{\hbar} \geqslant \Big|\frac{k}{2}-\frac{m\omega}{\hbar k}\Big|^2 \quad\Longrightarrow\quad 0 \leqslant \omega \leqslant \frac{\hbar k_{\rm F} k}{m} - \frac{\hbar k^2}{2m} \quad , \tag{10.121}$$

which in turn requires $k \leqslant 2k_{\rm F}$. In this case, we have

$$2\pi\, S(\boldsymbol{k},\omega) = \frac{m}{4\pi\hbar k}\left\{k_{\rm F}^2 - \Big(k_{\rm F}^2 - \frac{2m\omega}{\hbar}\Big)\right\} = \frac{m^2\omega}{2\pi\hbar^2 k} \quad . \tag{10.122}$$

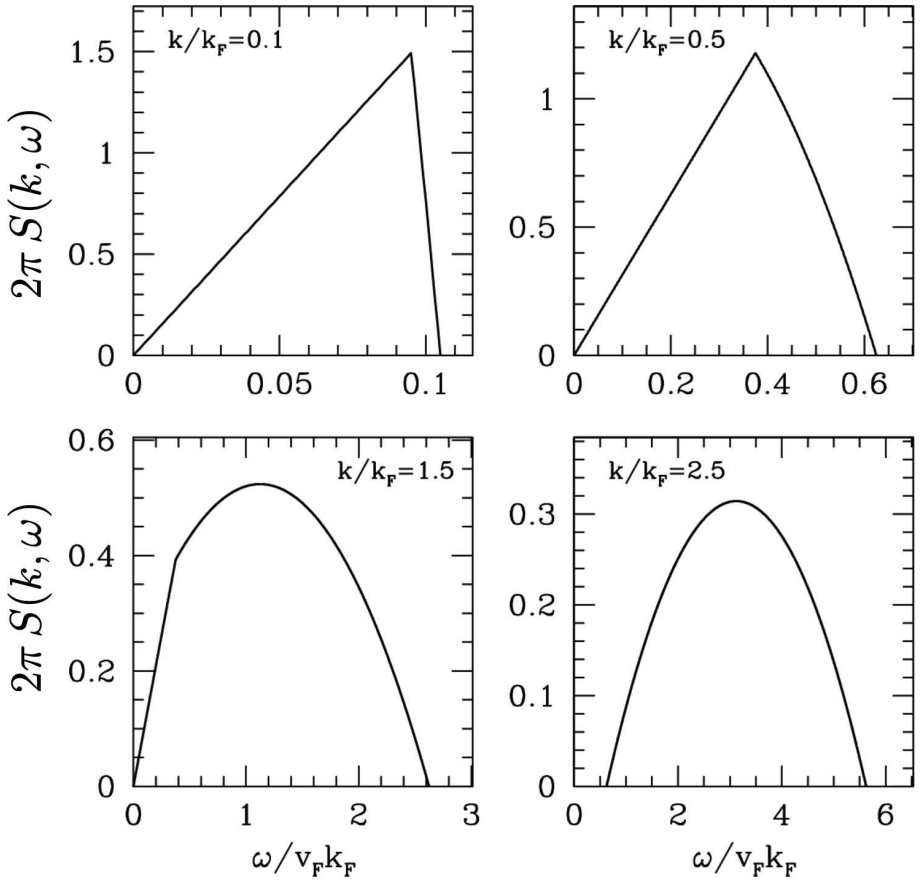


Figure 10.5: The dynamic structure factor $S(k,\omega)$ for the electron gas at various values of $k/k_{\rm F}$.

The second case

$$k_{\rm F}^2 - \frac{2m\omega}{\hbar} \leqslant \left| \frac{k}{2} - \frac{m\omega}{\hbar k} \right|^2 \quad \Longrightarrow \quad \omega \geqslant \frac{\hbar k_{\rm F} k}{m} - \frac{\hbar k^2}{2m} \quad . \tag{10.123}$$

However, we also have that

$$\left| \frac{k}{2} - \frac{m\omega}{\hbar k} \right|^2 \leqslant k_{\rm F}^2 \quad , \tag{10.124}$$

hence ω is restricted to the range

$$\frac{\hbar k}{2m} \left| k - 2k_{\rm F} \right| \leqslant \omega \leqslant \frac{\hbar k}{2m} \left| k + 2k_{\rm F} \right| \quad . \tag{10.125}$$

The integral in (10.119) then gives

$$2\pi\, S(\boldsymbol{k},\omega) = \frac{m}{4\pi\hbar k} \left\{ k_{\rm F}^2 - \left| \frac{k}{2} - \frac{m\omega}{\hbar k} \right|^2 \right\} \quad . \tag{10.126}$$

Putting it all together,

$$2\pi\, S(\boldsymbol{k},\omega) = \begin{cases} \frac{mk_{\rm F}}{\pi^2\hbar^2}\cdot\frac{\pi\omega}{2v_{\rm F}k} & \text{if } 0<\omega\leqslant v_{\rm F}k-\frac{\hbar k^2}{2m} \\ \frac{mk_{\rm F}}{\pi^2\hbar^2}\cdot\frac{\pi k_{\rm F}}{4k}\left[1-\left(\frac{\omega}{v_{\rm F}k}-\frac{k}{2k_{\rm F}}\right)^2\right] & \text{if } \left|v_{\rm F}k-\frac{\hbar k^2}{2m}\right|\leqslant\omega\leqslant v_{\rm F}k+\frac{\hbar k^2}{2m} \\ 0 & \text{if } \omega\geqslant v_{\rm F}k+\frac{\hbar k^2}{2m} \quad . \end{cases} \tag{10.127}$$

Integrating over all frequency gives the static structure factor,

$$S(\boldsymbol{k}) = \frac{1}{V}\left\langle n^\dagger_{\boldsymbol{k}}\, n_{\boldsymbol{k}}\right\rangle = \int\limits_{-\infty}^{\infty} d\omega\; S(\boldsymbol{k},\omega) \quad . \tag{10.128}$$

The result is

$$S(\boldsymbol{k}) = \begin{cases} \left(\frac{3k}{4k_{\rm F}}-\frac{k^3}{16k_{\rm F}^3}\right)n & \text{if } 0<k\leqslant 2k_{\rm F} \\ n & \text{if } k\geqslant 2k_{\rm F} \\ Vn^2 & \text{if } k=0 \quad , \end{cases} \tag{10.129}$$

where $n = k_{\rm F}^3/6\pi^2$ is the density (per spin polarization).

10.5 Charged Systems: Screening and Dielectric Response

10.5.1 Definition of the charge response functions

Consider a many-electron system in the presence of a time-varying external charge density $\rho_{\rm ext}(\boldsymbol{x},t)$. The perturbing Hamiltonian is then

$$\begin{aligned} \hat{H}_1 &= -e\int d^3x \int d^3x'\, \frac{n(\boldsymbol{x})\,\rho_{\rm ext}(\boldsymbol{x}',t)}{\left|\boldsymbol{x}-\boldsymbol{x}'\right|} \\ &= -e\int \frac{d^3k}{(2\pi)^3}\,\frac{4\pi}{\boldsymbol{k}^2}\,\hat{n}(\boldsymbol{k})\,\hat{\rho}_{\rm ext}(-\boldsymbol{k},t) \quad . \end{aligned} \tag{10.130}$$

The induced charge is $-e\,\delta n$, where δn is the induced number density:

$$\delta\hat{n}(\boldsymbol{q},\omega) = \frac{4\pi e}{\boldsymbol{q}^2}\,\hat{\chi}(\boldsymbol{q},\omega)\,\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega) \quad . \tag{10.131}$$

We can use this to determine the dielectric function $\epsilon(\boldsymbol{q},\omega)$:

$$\begin{aligned} \boldsymbol{\nabla}\cdot\boldsymbol{D} &= 4\pi\rho_{\rm ext} \\ \boldsymbol{\nabla}\cdot\boldsymbol{E} &= 4\pi\big(\rho_{\rm ext} - e\left\langle\delta n\right\rangle\big) \quad . \end{aligned} \tag{10.132}$$

In Fourier space,

$$\begin{aligned} i\boldsymbol{q}\cdot\boldsymbol{D}(\boldsymbol{q},\omega) &= 4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega) \\ i\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},\omega) &= 4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega) - 4\pi e\left\langle\delta\hat{n}(\boldsymbol{q},\omega)\right\rangle \quad , \end{aligned} \tag{10.133}$$

so that from $\boldsymbol{D}(\boldsymbol{q},\omega)=\epsilon(\boldsymbol{q},\omega)\,\boldsymbol{E}(\boldsymbol{q},\omega)$ follows

$$\begin{aligned} \frac{1}{\epsilon(\boldsymbol{q},\omega)} = \frac{i\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},\omega)}{i\boldsymbol{q}\cdot\boldsymbol{D}(\boldsymbol{q},\omega)} &= 1 - \frac{\delta\hat{n}(\boldsymbol{q},\omega)}{Z\hat{n}_{\rm ext}(\boldsymbol{q},\omega)} \\ &= 1 - \frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat{\chi}(\boldsymbol{q},\omega) \quad . \end{aligned} \tag{10.134}$$

A system is said to exhibit *perfect screening* if

$$\epsilon(\boldsymbol{q}\to 0,\omega=0)=\infty \quad\Longrightarrow\quad \lim_{\boldsymbol{q}\to 0}\frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat{\chi}(\boldsymbol{q},0)=1 \quad . \tag{10.135}$$

Here, $\hat{\chi}(\boldsymbol{q},\omega)$ is the usual density-density response function,

$$\hat{\chi}(\boldsymbol{q},\omega)=\frac{1}{\hbar V}\sum_j\frac{2\omega_j}{\omega_j^2-(\omega+i\epsilon)^2}\left|\langle\, j\,|\,\hat{n}_{\boldsymbol{q}}\,|\,0\,\rangle\right|^2 \quad , \tag{10.136}$$

where we content ourselves to work at $T=0$, and where $\omega_j\equiv(E_j-E_0)/\hbar$ is the excitation frequency for the state $|\,n\,\rangle$.

From $\boldsymbol{j}_{\rm charge}=\sigma\boldsymbol{E}$ and the continuity equation

$$i\boldsymbol{q}\cdot\langle\hat{\boldsymbol{j}}_{\rm charge}(\boldsymbol{q},\omega)\rangle = -ie\omega\langle\hat{n}(\boldsymbol{q},\omega)\rangle = i\sigma(\boldsymbol{q},\omega)\,\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},\omega) \quad , \tag{10.137}$$

we find

$$\overbrace{\left(4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega)-4\pi e\left\langle\delta\hat{n}(\boldsymbol{q},\omega)\right\rangle\right)}^{i\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},\omega)}\sigma(\boldsymbol{q},\omega) = -i\omega e\left\langle\delta\hat{n}(\boldsymbol{q},\omega)\right\rangle \quad , \tag{10.138}$$

or

$$\frac{4\pi i}{\omega}\,\sigma(\boldsymbol{q},\omega)=\frac{\left\langle\delta\hat{n}(\boldsymbol{q},\omega)\right\rangle}{e^{-1}\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega)-\left\langle\delta\hat{n}(\boldsymbol{q},\omega)\right\rangle}=\frac{1-\epsilon^{-1}(\boldsymbol{q},\omega)}{\epsilon^{-1}(\boldsymbol{q}\omega)}=\epsilon(\boldsymbol{q},\omega)-1 \quad . \tag{10.139}$$

Thus, we arrive at

$$\frac{1}{\epsilon(\boldsymbol{q},\omega)}=1-\frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat{\chi}(\boldsymbol{q},\omega) \qquad , \qquad \epsilon(\boldsymbol{q},\omega)=1+\frac{4\pi i}{\omega}\,\sigma(\boldsymbol{q},\omega) \quad . \tag{10.140}$$

Taken together, these two equations allow us to relate the conductivity and the charge response function,

$$\sigma(\boldsymbol{q},\omega)=-\frac{i\omega}{\boldsymbol{q}^2}\,\frac{e^2\hat{\chi}(\boldsymbol{q},\omega)}{1-\frac{4\pi e^2}{\boldsymbol{q}^2}\hat{\chi}(\boldsymbol{q},\omega)} \quad . \tag{10.141}$$

10.5.2 Static screening: Thomas-Fermi approximation

Imagine a time-independent, slowly varying electrical potential $\phi(\boldsymbol{x})$. We may define the 'local chemical potential' $\widetilde{\mu}(\boldsymbol{x})$ as

$$\mu \equiv \widetilde{\mu}(\boldsymbol{x}) - e\phi(\boldsymbol{x}) \quad , \tag{10.142}$$

where μ is the bulk chemical potential. The local chemical potential is related to the local density by local thermodynamics. At $T = 0$,

$$\begin{aligned} \widetilde{\mu}(\boldsymbol{x}) \equiv \frac{\hbar^2}{2m}\, k_{\rm F}^2(\boldsymbol{x}) &= \frac{\hbar^2}{2m}\Big(3\pi^2 n + 3\pi^2 \delta n(\boldsymbol{x})\Big)^{2/3} \\ &= \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3}\left\{1 + \frac{2}{3}\,\frac{\delta n(\boldsymbol{x})}{n} + \ldots\right\} \quad , \end{aligned} \tag{10.143}$$

hence, to lowest order,

$$\delta n(\boldsymbol{x}) = \frac{3en}{2\mu}\,\phi(\boldsymbol{x}) \quad . \tag{10.144}$$

This makes sense – a positive potential induces an increase in the local electron number density. In Fourier space,

$$\langle \delta\hat{n}(\boldsymbol{q}, \omega = 0)\rangle = \frac{3en}{2\mu}\,\hat{\phi}(\boldsymbol{q}, \omega = 0) \quad . \tag{10.145}$$

Poisson's equation is $-\nabla^2\phi = 4\pi\rho_{\rm tot}$, *i.e.*

$$\begin{aligned} i\boldsymbol{q}\cdot\boldsymbol{E}(\boldsymbol{q},0) = \boldsymbol{q}^2\,\hat{\phi}(\boldsymbol{q},0) &= 4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},0) - 4\pi e\,\langle\delta\hat{n}(\boldsymbol{q},0)\rangle \\ &= 4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},0) - \frac{6\pi n e^2}{\mu}\,\hat{\phi}(\boldsymbol{q},0) \quad , \end{aligned} \tag{10.146}$$

and defining the Thomas-Fermi wavevector $q_{\rm TF}$ by

$$q_{\rm TF}^2 \equiv \frac{6\pi n e^2}{\mu} \quad , \tag{10.147}$$

we have

$$\hat{\phi}(\boldsymbol{q},0) = \frac{4\pi\hat{\rho}_{\rm ext}(\boldsymbol{q},0)}{\boldsymbol{q}^2 + q_{\rm TF}^2} \quad , \tag{10.148}$$

hence

$$e\,\langle\delta\hat{n}(\boldsymbol{q},0)\rangle = \frac{q_{\rm TF}^2}{\boldsymbol{q}^2 + q_{\rm TF}^2}\cdot\hat{\rho}_{\rm ext}(\boldsymbol{q},0) \quad \Longrightarrow \quad \epsilon(\boldsymbol{q},0) = 1 + \frac{q_{\rm TF}^2}{\boldsymbol{q}^2} \quad . \tag{10.149}$$

Note that $\epsilon(\boldsymbol{q}\to 0, \omega = 0) = \infty$, so there is perfect screening.

The Thomas-Fermi wavelength is $\lambda_{\rm TF} = q_{\rm TF}^{-1}$, and may be written as

$$\lambda_{\rm TF} = \left(\frac{\pi}{12}\right)^{1/6}\sqrt{r_s}\,a_{\rm B} \simeq 0.800\,\sqrt{r_s}\,a_{\rm B} \quad , \tag{10.150}$$

where r_s is the dimensionless free electron sphere radius, given in units of the Bohr radius $a_{\rm B} = \hbar^2/me^2 = 0.529$Å, defined by $\frac{4}{3}\pi(r_s a_{\rm B})^3\, n = 1$, hence $r_s \propto n^{-1/3}$. Small r_s corresponds to high density. Since Thomas-Fermi theory is a statistical theory, it can only be valid if there are many particles within a sphere of radius $\lambda_{\rm TF}$, *i.e.* $\frac{4}{3}\pi\lambda_{\rm TF}^3\, n > 1$, or $r_s \lesssim (\pi/12)^{1/3} \simeq 0.640$. TF theory is applicable only in the high density limit.

In the presence of a δ-function external charge density $\rho_{\rm ext}(\boldsymbol{x}) = Ze\,\delta(\boldsymbol{x})$, we have its Fourier transform $\hat\rho_{\rm ext}(\boldsymbol{q}, 0) = Ze$, and

$$\langle \delta\hat n(\boldsymbol{q}, 0)\rangle = \frac{Zq_{\rm TF}^2}{\boldsymbol{q}^2 + q_{\rm TF}^2} \quad \Longrightarrow \quad \langle \delta n(\boldsymbol{x})\rangle = \frac{Z\, e^{-r/\lambda_{\rm TF}}}{4\pi r} \tag{10.151}$$

Note the decay on the scale of $\lambda_{\rm TF}$. Note also the perfect screening:

$$e\,\langle \delta\hat n(\boldsymbol{q}\to 0, \omega = 0)\rangle = \hat\rho_{\rm ext}(\boldsymbol{q}\to 0, \omega = 0) = Ze \quad . \tag{10.152}$$

10.5.3 High frequency behavior of $\epsilon(\boldsymbol{q}, \omega)$

We have

$$\epsilon^{-1}(\boldsymbol{q}, \omega) = 1 - \frac{4\pi e^2}{\boldsymbol{q}^2}\,\hat\chi(\boldsymbol{q}, \omega) \tag{10.153}$$

and, at $T = 0$,

$$\hat\chi(\boldsymbol{q}, \omega) = \frac{1}{\hbar V}\sum_j \big|\langle\, j\,|\,\hat n_{\boldsymbol{q}}^\dagger\,|\,0\,\rangle\big|^2 \left\{\frac{1}{\omega + \omega_j + i\epsilon} - \frac{1}{\omega - \omega_j + i\epsilon}\right\} \quad , \tag{10.154}$$

where the number density operator is

$$\hat n_{\boldsymbol{q}}^\dagger = \begin{cases} \sum_i e^{i\boldsymbol{q}\cdot\boldsymbol{x}_i} & (1^{\rm st}\ \text{quantized}) \\ \sum_{\boldsymbol{k}} \psi_{\boldsymbol{k}+\boldsymbol{q}}^\dagger\,\psi_{\boldsymbol{k}} & (2^{\rm nd}\ \text{quantized: } \{\psi_{\boldsymbol{k}}, \psi_{\boldsymbol{k}'}^\dagger\} = \delta_{\boldsymbol{k}\boldsymbol{k}'}) \end{cases} \quad . \tag{10.155}$$

Taking the limit $\omega \to \infty$, we find

$$\hat\chi(\boldsymbol{q}, \omega\to\infty) = -\frac{2}{\hbar V\omega^2}\sum_j \big|\langle\, j\,|\,\hat n_{\boldsymbol{q}}^\dagger\,|\,0\,\rangle\big|^2\,\omega_j = -\frac{2}{\hbar\omega^2}\int\limits_{-\infty}^{\infty}\frac{d\omega'}{2\pi}\,\omega'\,S(\boldsymbol{q}, \omega') \quad . \tag{10.156}$$

Invoking the f-sum rule, the above integral is $n\hbar\boldsymbol{q}^2/2m$, hence

$$\hat\chi(\boldsymbol{q}, \omega\to\infty) = -\frac{n\boldsymbol{q}^2}{m\omega^2} \quad , \tag{10.157}$$

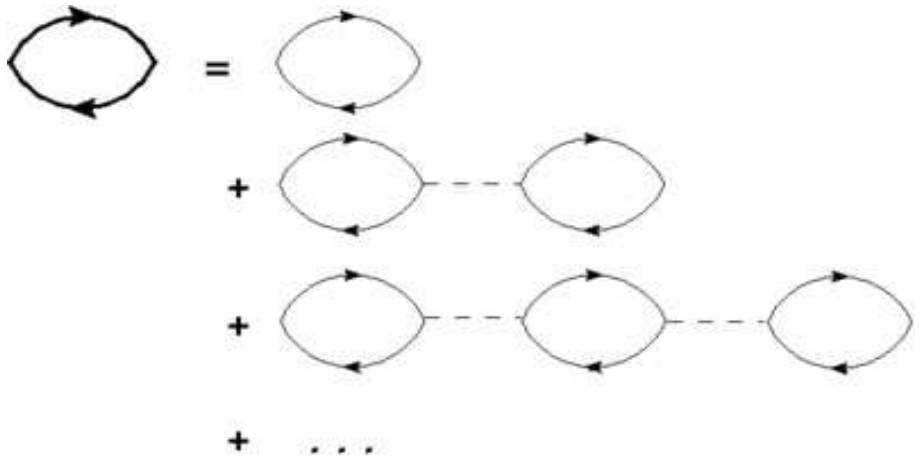


Figure 10.6: Perturbation expansion for RPA susceptibility bubble. Each bare bubble contributes a factor $\chi^0(\boldsymbol{q},\omega)$ and each wavy interaction line $\hat{v}(\boldsymbol{q})$. The infinite series can be summed, yielding eqn. 10.162.

and

$$\epsilon^{-1}(\boldsymbol{q},\omega\to\infty) = 1 + \frac{\omega_{\rm p}^2}{\omega^2} \quad , \tag{10.158}$$

where

$$\omega_{\rm p} \equiv \sqrt{\frac{4\pi n e^2}{m}} \tag{10.159}$$

is the *plasma frequency*.

10.5.4 Random phase approximation (RPA)

The electron charge appears nowhere in the free electron gas response function $\chi^0(\boldsymbol{q},\omega)$. An interacting electron gas certainly does know about electron charge, since the Coulomb repulsion between electrons is part of the Hamiltonian. The idea behind the RPA is to obtain an approximation to the interacting $\hat{\chi}(\boldsymbol{q},\omega)$ from the noninteracting $\chi^0(\boldsymbol{q},\omega)$ by self-consistently adjusting the charge so that the perturbing charge density is not $\rho_{\rm ext}(\boldsymbol{x})$, but rather $\rho_{\rm ext}(\boldsymbol{x},t) - e\,\langle\delta n(\boldsymbol{x},t)\rangle$. Thus, we write

$$\begin{aligned} e\,\langle\delta\hat{n}(\boldsymbol{q},\omega)\rangle &= \frac{4\pi e^2}{\boldsymbol{q}^2}\,\chi^{\rm RPA}(\boldsymbol{q},\omega)\,\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega) \\ &= \frac{4\pi e^2}{\boldsymbol{q}^2}\,\chi^0(\boldsymbol{q},\omega)\left\{\hat{\rho}_{\rm ext}(\boldsymbol{q},\omega) - e\,\langle\delta\hat{n}(\boldsymbol{q},\omega)\rangle\right\} \quad , \end{aligned} \tag{10.160}$$

which gives

$$\chi^{\rm RPA}(\boldsymbol{q},\omega) = \frac{\chi^0(\boldsymbol{q},\omega)}{1 + \frac{4\pi e^2}{\boldsymbol{q}^2}\chi^0(\boldsymbol{q},\omega)} \quad . \tag{10.161}$$

Several comments are in order.

1. If the electron-electron interaction were instead given by a general $\hat{v}(\boldsymbol{q})$ rather than the specific Coulomb form $\hat{v}(\boldsymbol{q}) = 4\pi e^2/\boldsymbol{q}^2$, we would obtain

$$\chi^{\rm RPA}(\boldsymbol{q},\omega) = \frac{\chi^0(\boldsymbol{q},\omega)}{1+\hat{v}(\boldsymbol{q})\,\chi^0(\boldsymbol{q},\omega)} \quad . \tag{10.162}$$

2. Within the RPA, there is perfect screening:

$$\lim_{\boldsymbol{q}\to 0} \frac{4\pi e^2}{\boldsymbol{q}^2}\,\chi^{\rm RPA}(\boldsymbol{q},\omega) = 1 \quad . \tag{10.163}$$

3. The RPA expression may be expanded in an infinite series,

$$\chi^{\rm RPA} = \chi^0 - \chi^0\,\hat{v}\,\chi^0 + \chi^0\,\hat{v}\,\chi^0\,\hat{v}\,\chi^0 - \ldots \quad , \tag{10.164}$$

which has a diagrammatic interpretation, depicted in fig. 10.6. The perturbative expansion in the interaction $\hat{v}$ may be resummed to yield the RPA result.

4. The RPA dielectric function takes the simple form

$$\epsilon^{\rm RPA}(\boldsymbol{q},\omega) = 1 + \frac{4\pi e^2}{\boldsymbol{q}^2}\,\chi^0(\boldsymbol{q},\omega) \quad . \tag{10.165}$$

5. Explicitly,

$$\begin{aligned}\mathsf{Re}\;\epsilon^{\rm RPA}(\boldsymbol{q},\omega) = 1 + \frac{q_{\rm TF}^2}{q^2}\Bigg\{\frac{1}{2} + \frac{k_{\rm F}}{4q}\Bigg[&\left(1 - \frac{(\omega-\hbar q^2/2m)^2}{(v_{\rm F}q)^2}\right)\ln\left|\frac{\omega - v_{\rm F}q - \hbar q^2/2m}{\omega + v_{\rm F}q - \hbar q^2/2m}\right| \\ &+ \left(1 - \frac{(\omega-\hbar q^2/2m)^2}{(v_{\rm F}q)^2}\right)\ln\left|\frac{\omega - v_{\rm F}q + \hbar q^2/2m}{\omega + v_{\rm F}q + \hbar q^2/2m}\right|\Bigg]\Bigg\}\end{aligned} \tag{10.166}$$

and

$$\mathsf{Im}\;\epsilon^{\rm RPA}(\boldsymbol{q},\omega) = \begin{cases} \frac{\pi\omega}{2v_{\rm F}q}\cdot\frac{q_{\rm TF}^2}{q^2} & \text{if } 0 \leqslant \omega \leqslant v_{\rm F}q - \hbar q^2/2m \\ \frac{\pi k_{\rm F}}{4q}\left(1 - \frac{(\omega-\hbar q^2/2m)^2}{(v_{\rm F}q)^2}\right)\frac{q_{\rm TF}^2}{q^2} & \text{if } v_{\rm F}q - \hbar q^2/2m \leqslant \omega \leqslant v_{\rm F}q + \hbar q^2/2m \\ 0 & \text{if } \omega > v_{\rm F}q + \hbar q^2/2m \end{cases} \tag{10.167}$$

6. Note that
$$\epsilon^{\rm RPA}(\boldsymbol{q},\omega\to\infty) = 1 - \frac{\omega_{\rm p}^2}{\omega^2} \quad , \tag{10.168}$$
in agreement with the f-sum rule, and
$$\epsilon^{\rm RPA}(\boldsymbol{q}\to 0,\omega=0) = 1 + \frac{q_{\rm TF}^2}{q^2} \quad , \tag{10.169}$$
in agreement with Thomas-Fermi theory.

7. When $\omega = 0$ we have
$$\epsilon^{\rm RPA}(\boldsymbol{q},0) = 1 + \frac{q_{\rm TF}^2}{q^2}\left\{\frac{1}{2} + \frac{k_{\rm F}}{2q}\left(1-\frac{q^2}{4k_{\rm F}^2}\right)\ln\left|\frac{q+2k_{\rm F}}{q-2k_{\rm F}}\right|\right\} \quad , \tag{10.170}$$
which is real and which has a singularity at $q = 2k_{\rm F}$. This means that the long-distance behavior of $\langle\delta n(\boldsymbol{x})\rangle$ must oscillate. For a local charge perturbation, $\rho_{\rm ext}(\boldsymbol{x}) = Ze\,\delta(\boldsymbol{x})$, we have
$$\langle\delta n(\boldsymbol{x})\rangle = \frac{Z}{2\pi^2 r}\int\limits_0^\infty dq\, q\sin(qr)\left\{1-\frac{1}{\epsilon(\boldsymbol{q},0)}\right\} \quad , \tag{10.171}$$
and within the RPA one finds for long distances
$$\langle\delta n(\boldsymbol{x})\rangle \sim \frac{Z\cos(2k_{\rm F}r)}{r^3} \quad , \tag{10.172}$$
rather than the Yukawa form familiar from Thomas-Fermi theory.

10.5.5 Plasmons

The RPA response function diverges when $\hat{v}(\boldsymbol{q})\,\chi^0(\boldsymbol{q},\omega) = -1$. For a given value of $\boldsymbol{q}$, this occurs for a specific value (or for a discrete set of values) of ω, *i.e.* it defines a dispersion relation $\omega = \Omega(\boldsymbol{q})$. The poles of $\chi^{\rm RPA}$ and are identified with elementary excitations of the electron gas known as *plasmons*.

To find the plasmon dispersion, we first derive a result for $\chi^0(\boldsymbol{q},\omega)$, starting with
$$\begin{aligned}\chi^0(\boldsymbol{q},t) &= \frac{i}{\hbar V}\left\langle\big[\hat{n}(\boldsymbol{q},t),\hat{n}(-\boldsymbol{q},0)\big]\right\rangle\Theta(t)\\ &= \frac{i}{\hbar V}\left\langle\Big[\sum_{\boldsymbol{k}\sigma}\psi^\dagger_{\boldsymbol{k},\sigma}\,\psi_{\boldsymbol{k}+\boldsymbol{q},\sigma}, \sum_{\boldsymbol{k}',\sigma'}\psi^\dagger_{\boldsymbol{k}',\sigma'}\,\psi_{\boldsymbol{k}'-\boldsymbol{q},\sigma'}\Big]\right\rangle e^{i(\varepsilon(\boldsymbol{k})-\varepsilon(\boldsymbol{k}+\boldsymbol{q}))t/\hbar}\,\Theta(t) \quad ,\end{aligned} \tag{10.173}$$
where $\varepsilon(\boldsymbol{k})$ is the noninteracting electron dispersion. For a free electron gas, $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$. Next, using
$$\big[AB,CD\big] = A\big\{B,C\big\}D - \big\{A,C\big\}BD + CA\big\{B,D\big\} - C\big\{A,D\big\}B \tag{10.174}$$

we obtain

$$\chi^0(\boldsymbol{q},t) = \frac{i}{\hbar V}\sum_{\boldsymbol{k}\sigma}(f_{\boldsymbol{k}} - f_{\boldsymbol{k}+\boldsymbol{q}})\,e^{i(\varepsilon(\boldsymbol{k})-\varepsilon(\boldsymbol{k}+\boldsymbol{q}))t/\hbar}\,\Theta(t) \quad , \tag{10.175}$$

and therefore

$$\chi^0(\boldsymbol{q},\omega) = 2\int\frac{d^3k}{(2\pi)^3}\;\frac{f_{\boldsymbol{k}+\boldsymbol{q}} - f_{\boldsymbol{k}}}{\hbar\omega - \varepsilon(\boldsymbol{k}+\boldsymbol{q}) + \varepsilon(\boldsymbol{k}) + i\epsilon} \quad . \tag{10.176}$$

Here,

$$f_{\boldsymbol{k}} = \frac{1}{e^{(\varepsilon(\boldsymbol{k})-\mu)/k_{\rm B}T}+1} \tag{10.177}$$

is the Fermi distribution. At $T=0$, $f_{\boldsymbol{k}} = \Theta(k_{\rm F}-k)$, and for $\omega \gg v_{\rm F}q$ we can expand $\chi^0(\boldsymbol{q},\omega)$ in powers of ω^{-2}, yielding

$$\chi^0(\boldsymbol{q},\omega) = -\frac{k_{\rm F}^3}{3\pi^2}\cdot\frac{q^2}{m\omega^2}\left\{1+\frac{3}{5}\left(\frac{\hbar k_{\rm F}q}{m\omega}\right)^2+\ldots\right\} \quad , \tag{10.178}$$

so the resonance condition becomes

$$\begin{aligned} 0 &= 1 + \frac{4\pi e^2}{q^2}\,\chi^0(\boldsymbol{q},\omega) \\ &= 1 - \frac{\omega_{\rm p}^2}{\omega^2}\left\{1+\frac{3}{5}\left(\frac{v_{\rm F}q}{\omega}\right)^2+\ldots\right\} \quad . \end{aligned} \tag{10.179}$$

This gives the dispersion

$$\omega = \omega_{\rm p}\left\{1+\frac{3}{10}\left(\frac{v_{\rm F}q}{\omega_{\rm p}}\right)^2+\ldots\right\} \quad . \tag{10.180}$$

Recall that the particle-hole continuum frequencies are bounded by $\omega_{\rm min}(q)$ and $\omega_{\rm max}(q)$, which are given in eqs. 10.109 and 10.108. Eventually the plasmon penetrates the particle-hole continuum, at which point it becomes heavily damped since it can decay into particle-hole excitations.

10.5.6 Jellium

Finally, consider an electron gas in the presence of a neutralizing ionic background. We assume one species of ion with mass $M_{\rm i}$ and charge $+Z_{\rm i}e$, and we smear the ionic charge into a continuum as an approximation. This nonexistent material is known in the business of many-body physics as *jellium*. Let the ion number density be $n_{\rm i}$, and the electron number density be $n_{\rm e}$. Then Laplace's equation says

$$\nabla^2\phi = -4\pi\rho_{\rm charge} = -4\pi e\big(n_{\rm i} - n_{\rm e} + n_{\rm ext}\big) \quad , \tag{10.181}$$

where $n_{\rm ext} = \rho_{\rm ext}/e$, where $\rho_{\rm ext}$ is the test charge density, regarded as a perturbation to the jellium. The ions move according to

$$M_{\rm i}\,\frac{d\boldsymbol{v}}{dt} = Z_{\rm i}\,e\boldsymbol{E} = -Z_{\rm i}\,e\,\boldsymbol{\nabla}\phi \quad . \tag{10.182}$$

They also satisfy continuity, which to lowest order in small quantities is governed by the equation

$$n_{\rm i}^0 \, \boldsymbol{\nabla}\cdot\boldsymbol{v} + \frac{\partial n_{\rm i}}{\partial t} = 0 \quad , \tag{10.183}$$

where n_i^0 is the average ionic number density. Taking the time derivative of the above equation, and invoking Newton's law for the ion's as well as Laplace, we find

$$-\frac{\partial^2 n_{\rm i}(\boldsymbol{x},t)}{\partial t^2} = \frac{4\pi n_{\rm i}^0 Z_{\rm i} e^2}{M_{\rm i}} \Big(n_{\rm i}(\boldsymbol{x},t) + n_{\rm ext}(\boldsymbol{x},t) - n_{\rm e}(\boldsymbol{x},t) \Big) \quad . \tag{10.184}$$

In Fourier space,

$$\omega^2 \, \hat{n}_{\rm i}(\boldsymbol{q},\omega) = \Omega_{\rm p,i}^2 \Big(\hat{n}_{\rm i}(\boldsymbol{q},\omega) + \hat{n}_{\rm ext}(\boldsymbol{q},\omega) - \hat{n}_{\rm e}(\boldsymbol{q},\omega) \Big) \quad , \tag{10.185}$$

where

$$\Omega_{\rm p,i} = \sqrt{\frac{4\pi n_{\rm i}^0 Z_{\rm i} e^2}{M_{\rm i}}} \tag{10.186}$$

is the ionic plasma frequency. Typically $\Omega_{\rm p,i} \approx 10^{13}\,{\rm s}^{-1}$.

Since the ionic mass $M_{\rm i}$ is much greater than the electron mass, the ionic plasma frequency is much greater than the electron plasma frequency. We assume that the ions may be regarded as 'slow' and that the electrons respond according to Eqn. 10.131, *viz.*

$$\hat{n}_{\rm e}(\boldsymbol{q},\omega) = \frac{4\pi e}{\boldsymbol{q}^2} \, \chi_{\rm e}(\boldsymbol{q},\omega) \, \Big(\hat{n}_{\rm i}(\boldsymbol{q},\omega) + \hat{n}_{\rm ext}(\boldsymbol{q},\omega) \Big) \quad . \tag{10.187}$$

We then have

$$\frac{\omega^2}{\Omega_{\rm p,i}^2} \, \hat{n}_{\rm i}(\boldsymbol{q},\omega) = \frac{\hat{n}_{\rm i}(\boldsymbol{q},\omega) + \hat{n}_{\rm ext}(\boldsymbol{q},\omega)}{\epsilon_{\rm e}(\boldsymbol{q},\omega)} \quad . \tag{10.188}$$

From this equation, we obtain $\hat{n}_{\rm i}(\boldsymbol{q},\omega)$ and then $n_{\rm tot} \equiv n_{\rm i} - n_{\rm e} + n_{\rm ext}$. We thereby obtain

$$\hat{n}_{\rm tot}(\boldsymbol{q},\omega) = \frac{\hat{n}_{\rm ext}(\boldsymbol{q},\omega)}{\epsilon_{\rm e}(\boldsymbol{q},\omega) - \omega^{-2}\,\Omega_{\rm p,i}^2} \quad . \tag{10.189}$$

Finally, the dielectric function of the jellium system is given by

$$\begin{aligned} \epsilon(\boldsymbol{q},\omega) &= \frac{\hat{n}_{\rm ext}(\boldsymbol{q},\omega)}{\hat{n}_{\rm tot}(\boldsymbol{q},\omega)} \\ &= \epsilon_{\rm e}(\boldsymbol{q},\omega) - \frac{\omega^2}{\Omega_{\rm p,i}^2} \quad . \end{aligned} \tag{10.190}$$

At frequencies low compared to the electron plasma frequency, we approximate $\epsilon_{\rm e}(\boldsymbol{q},\omega)$ by the Thomas-Fermi form, $\epsilon_{\rm e}(\boldsymbol{q},\omega) \approx \big(q^2 + q_{\rm TF}^2\big)/q^2$. Then

$$\epsilon(\boldsymbol{q},\omega) \approx 1 + \frac{q_{\rm TF}^2}{q^2} - \frac{\Omega_{\rm p,i}^2}{\omega^2} \quad . \tag{10.191}$$

The zeros of this function, given by $\epsilon(\boldsymbol{q},\omega_{\boldsymbol{q}})=0$, occur for

$$\omega_{\boldsymbol{q}} = \frac{\Omega_{\rm p,i}\,q}{\sqrt{q^2+q_{\rm TF}^2}} \quad . \tag{10.192}$$

This allows us to write

$$\frac{4\pi e^2}{\boldsymbol{q}^2}\,\frac{1}{\epsilon(\boldsymbol{q},\omega)} = \frac{4\pi e^2}{\boldsymbol{q}^2+q_{\rm TF}^2}\cdot\frac{\omega^2}{\omega^2-\omega_{\boldsymbol{q}}^2} \quad . \tag{10.193}$$

This is interpreted as the effective interaction between charges in the jellium model, arising from both electronic and ionic screening. Note that the interaction is negative, *i.e.* attractive, for $\omega^2<\omega_{\boldsymbol{q}}^2$. At frequencies high compared to $\omega_{\boldsymbol{q}}$, but low compared to the electronic plasma frequency, the effective potential is of the Yukawa form. Only the electrons then participate in screening, because the phonons are too slow.

10.6 Electromagnetic Response

Consider an interacting system consisting of electrons of charge $-e$ in the presence of a time-varying electromagnetic field. The electromagnetic field is given in terms of the 4-potential $A^\mu=(A^0,\boldsymbol{A})$:

$$\boldsymbol{E} = -\boldsymbol{\nabla}A^0 - \frac{1}{c}\frac{\partial\boldsymbol{A}}{\partial t} \qquad , \qquad \boldsymbol{B}=\boldsymbol{\nabla}\times\boldsymbol{A} \quad . \tag{10.194}$$

The Hamiltonian for an N-particle system is

$$\begin{aligned}\hat H(A^\mu) &= \sum_{i=1}^N\left\{\frac{1}{2m}\Big(\boldsymbol{p}_i+\frac{e}{c}\boldsymbol{A}(\boldsymbol{x}_i,t)\Big)^2 - eA^0(\boldsymbol{x}_i,t)+v_{\rm ext}(\boldsymbol{x}_i)\right\}+\sum_{i<j}u(\boldsymbol{x}_i-\boldsymbol{x}_j)\\ &= \hat H(0)-\frac{1}{c}\int d^3x\;j^{\rm p}_\mu(\boldsymbol{x})\,A^\mu(\boldsymbol{x},t)+\frac{e^2}{2mc^2}\int d^3x\;n(\boldsymbol{x})\,\boldsymbol{A}^2(\boldsymbol{x},t) \quad ,\end{aligned} \tag{10.195}$$

where we have defined

$$\begin{aligned} n(\boldsymbol{x}) &\equiv \sum_{i=1}^N\delta(\boldsymbol{x}-\boldsymbol{x}_i)\\ j^{\rm p}_0(\boldsymbol{x}) &\equiv c\,e\,n(\boldsymbol{x})\\ \boldsymbol{j}^{\rm p}(\boldsymbol{x}) &\equiv -\frac{e}{2m}\sum_{i=1}^N\Big\{\boldsymbol{p}_i\,\delta(\boldsymbol{x}-\boldsymbol{x}_i)+\delta(\boldsymbol{x}-\boldsymbol{x}_i)\,\boldsymbol{p}_i\Big\} \quad .\end{aligned} \tag{10.196}$$

Throughout this discussion we invoke covariant/contravariant notation, using the metric

$$g_{\mu\nu}=g^{\mu\nu}=\begin{pmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix} \quad , \tag{10.197}$$

so that

$$\begin{aligned} j^\mu &= (j^0, j^1, j^2, j^3) \equiv (j^0, \boldsymbol{j}) \\ j_\mu &= g_{\mu\nu} j^\nu = (-j^0, j^1, j^2, j^3) \\ j_\mu A^\mu &= j^\mu g_{\mu\nu}\, A^\nu = -j^0 A^0 + \boldsymbol{j} \cdot \boldsymbol{A} \equiv j \cdot A \end{aligned} \tag{10.198}$$

The quantity $j^{\rm p}_\mu(\boldsymbol{x})$ is known as the *paramagnetic current density*. The physical current density $j_\mu(\boldsymbol{x})$ also contains a *diamagnetic* contribution:

$$\begin{aligned} j_\mu(\boldsymbol{x}) &= -c\,\frac{\delta \hat{H}}{\delta A^\mu(\boldsymbol{x})} = j^{\rm p}_\mu(\boldsymbol{x}) + j^{\rm d}_\mu(\boldsymbol{x}) \\ j^{\rm d}_0(\boldsymbol{x}) &= 0 \\ \boldsymbol{j}^{\rm d}(\boldsymbol{x}) &= -\frac{e^2}{mc}\, n(\boldsymbol{x}) \boldsymbol{A}(\boldsymbol{x}) = -\frac{e}{mc^2}\, j^{\rm p}_0(\boldsymbol{x})\, \boldsymbol{A}(\boldsymbol{x}) \quad . \end{aligned} \tag{10.199}$$

The electromagnetic response tensor $K_{\mu\nu}$ is defined via

$$\big\langle j_\mu(\boldsymbol{x}, t) \big\rangle = -\frac{c}{4\pi} \int dt' \int d^3x'\; K_{\mu\nu}(\boldsymbol{x}t; \boldsymbol{x}'t')\, A^\nu(\boldsymbol{x}', t') \quad , \tag{10.200}$$

valid to first order in the external 4-potential A^μ. From

$$\begin{aligned} \big\langle j^{\rm p}_\mu(\boldsymbol{x}, t) \big\rangle &= \frac{i}{\hbar c} \int dt' \int d^3x' \, \big\langle \big[j^{\rm p}_\mu(\boldsymbol{x}, t), j^{\rm p}_\nu(\boldsymbol{x}', t') \big] \big\rangle \, \Theta(t - t')\, A^\nu(\boldsymbol{x}', t') \\ \big\langle j^{\rm d}_\mu(\boldsymbol{x}, t) \big\rangle &= -\frac{e}{mc^2} \, \big\langle j^{\rm p}_0(\boldsymbol{x}, t) \big\rangle \, A^\mu(\boldsymbol{x}, t)\, (1 - \delta_{\mu 0}) \quad , \end{aligned} \tag{10.201}$$

we conclude

$$\begin{aligned} K_{\mu\nu}(\boldsymbol{x}t; \boldsymbol{x}'t') = \frac{4\pi}{i\hbar c^2} \Big\langle \big[j^{\rm p}_\mu(\boldsymbol{x}, t), j^{\rm p}_\nu(\boldsymbol{x}', t') \big] \Big\rangle \, \Theta(t - t') \\ + \frac{4\pi e}{mc^2} \, \big\langle j^{\rm p}_0(\boldsymbol{x}, t) \big\rangle \, \delta(\boldsymbol{x} - \boldsymbol{x}')\, \delta(t - t')\, \delta_{\mu\nu}\, (1 - \delta_{\mu 0}) \quad . \end{aligned} \tag{10.202}$$

The first term is sometimes known as the *paramagnetic response kernel*,

$$K^{\rm p}_{\mu\nu}(x; x') = \frac{4\pi}{i\hbar c^2} \, \big\langle \big[j^{\rm p}_\mu(x), j^{\rm p}_\nu(x') \big] \big\rangle \, \Theta(t - t') \quad , \tag{10.203}$$

is not directly calculable by diagrammatic perturbation theory. Rather, one obtains the time-ordered response function $K^{\rm p,T}_{\mu\nu}(x; x') = (4\pi/i\hbar c^2) \, \big\langle \hat{\mathcal{T}} j^{\rm p}_\mu(x)\, j^{\rm p}_\nu(x') \big\rangle$, where $x^\mu \equiv (ct, \boldsymbol{x})$.

Second quantized notation

In the presence of an electromagnetic field described by the 4-potential $A^\mu = (c\phi, \boldsymbol{A})$, the Hamiltonian of an interacting electron system takes the form

$$\begin{aligned}\hat{H} = \sum_\sigma \int d^3x\, \psi^\dagger_\sigma(\boldsymbol{x}) \left\{ \frac{1}{2m}\Big(\frac{\hbar}{i}\boldsymbol{\nabla} + \frac{e}{c}\boldsymbol{A}\Big)^2 - eA^0(\boldsymbol{x}) + v_{\text{ext}}(\boldsymbol{x}) \right\} \psi_\sigma(\boldsymbol{x}) \\ + \frac{1}{2}\sum_{\sigma,\sigma'} \int d^3x \int d^3x'\, \psi^\dagger_\sigma(\boldsymbol{x})\, \psi^\dagger_{\sigma'}(\boldsymbol{x}')\, u(\boldsymbol{x}-\boldsymbol{x}')\, \psi_{\sigma'}(\boldsymbol{x}')\, \psi_\sigma(\boldsymbol{x}) \quad ,\end{aligned} \tag{10.204}$$

where $v(\boldsymbol{x}-\boldsymbol{x}')$ is a two-body interaction, *e.g.* $e^2/|\boldsymbol{x}-\boldsymbol{x}'|$, and $U(\boldsymbol{x})$ is the external scalar potential. Expanding in powers of A^μ,

$$\hat{H}(A^\mu) = \hat{H}(0) - \frac{1}{c}\int d^3x\, j^{\rm p}_\mu(\boldsymbol{x})\, A^\mu(\boldsymbol{x}) + \frac{e^2}{2mc^2}\sum_\sigma \int d^3x\, \psi^\dagger_\sigma(\boldsymbol{x})\, \psi_\sigma(\boldsymbol{x})\, \boldsymbol{A}^2(\boldsymbol{x}) \quad , \tag{10.205}$$

where the paramagnetic current density $j^{\rm p}_\mu(\boldsymbol{x})$ is defined by

$$\begin{aligned} j^{\rm p}_0(\boldsymbol{x}) &= c\, e \sum_\sigma \psi^\dagger_\sigma(\boldsymbol{x})\, \psi_\sigma(\boldsymbol{x}) \\ \boldsymbol{j}^{\rm p}(\boldsymbol{x}) &= \frac{ie\hbar}{2m}\sum_\sigma \Big\{ \psi^\dagger_\sigma(\boldsymbol{x})\, \boldsymbol{\nabla}\, \psi_\sigma(\boldsymbol{x}) - \Big(\boldsymbol{\nabla}\psi^\dagger_\sigma(\boldsymbol{x})\Big)\, \psi_\sigma(\boldsymbol{x}) \Big\} \quad . \end{aligned} \tag{10.206}$$

10.6.1 Gauge invariance and charge conservation

In Fourier space, with $q^\mu = (\omega/c, \boldsymbol{q})$, we have, for homogeneous systems,

$$\big\langle j_\mu(q) \big\rangle = -\frac{c}{4\pi}\, K_{\mu\nu}(q)\, A^\nu(q) \quad . \tag{10.207}$$

Note our convention on Fourier transforms:

$$\begin{aligned} H(x) &= \int \frac{d^4k}{(2\pi)^4}\; \hat{H}(k)\, e^{+ik\cdot x} \\ \hat{H}(k) &= \int d^4x\, H(x)\, e^{-ik\cdot x} \quad , \end{aligned} \tag{10.208}$$

where $k\cdot x \equiv k_\mu x^\mu = \boldsymbol{k}\cdot\boldsymbol{x} - \omega t$. Under a gauge transformation, $A^\mu \to A^\mu + \partial^\mu\Lambda$, *i.e.*

$$A^\mu(q) \to A^\mu(q) + i\Lambda(q)\, q^\mu \quad , \tag{10.209}$$

where Λ is an arbitrary scalar function. Since the physical current must be unchanged by a gauge transformation, we conclude that $K_{\mu\nu}(q)\,q^\nu = 0$. We also have the continuity equation, $\partial^\mu j_\mu = 0$, the Fourier space version of which says $q^\mu\,j_\mu(q) = 0$, which in turn requires $q^\mu\,K_{\mu\nu}(q) = 0$. Therefore,

$$\sum_\mu q^\mu\,K_{\mu\nu}(q) = \sum_\nu K_{\mu\nu}(q)\,q^\nu = 0 \quad . \tag{10.210}$$

In fact, the above conditions are identical owing to the reciprocity relations,

$$\begin{aligned} \mathsf{Re}\ K_{\mu\nu}(q) &= +\mathsf{Re}\ K_{\nu\mu}(-q) \\ \mathsf{Im}\ K_{\mu\nu}(q) &= -\mathsf{Im}\ K_{\nu\mu}(-q) \quad , \end{aligned} \tag{10.211}$$

which follow from the spectral representation of $K_{\mu\nu}(q)$. Thus,

$$\text{gauge invariance} \iff \text{charge conservation} \quad . \tag{10.212}$$

10.6.2 A sum rule

If we work in a gauge where $A^0 = 0$, then $\boldsymbol{E} = -c^{-1}\dot{\boldsymbol{A}}$, hence $\boldsymbol{E}(q) = iq^0\boldsymbol{A}(q)$, and

$$\begin{aligned} \big\langle j_i(q)\big\rangle &= -\frac{c}{4\pi}\,K_{ij}(q)\,A^j(q) \\ &= -\frac{c}{4\pi}\,K_{ij}(q)\,\frac{c}{i\omega}\,E^j(q) \equiv \sigma_{ij}(q)\,E^j(q) \quad . \end{aligned}$$

Thus, the conductivity tensor is given by

$$\sigma_{ij}(\boldsymbol{q},\omega) = \frac{ic^2}{4\pi\omega}\,K_{ij}(\boldsymbol{q},\omega) \quad . \tag{10.213}$$

If, in the $\omega \to 0$ limit, the conductivity is to remain finite, then we must have

$$\int_0^\infty dt \int d^3x\,\Big\langle \big[j_i^{\rm p}(\boldsymbol{x},t), j_j^{\rm p}(0,0)\big]\Big\rangle\,e^{+i\omega t} = \frac{ie^2n}{m}\,\delta_{ij} \quad , \tag{10.214}$$

where n is the electron number density. This relation is spontaneously violated in a superconductor, where $\sigma(\omega) \propto \omega^{-1}$ as $\omega \to 0$.

10.6.3 Longitudinal and transverse response

In an isotropic system, the spatial components of $K_{\mu\nu}$ may be resolved into longitudinal and transverse components, since the only preferred spatial vector is $\boldsymbol{q}$ itself. Thus, we may write

$$K_{ij}(\boldsymbol{q},\omega) = K_\parallel(\boldsymbol{q},\omega)\,\hat{q}_i\,\hat{q}_j + K_\perp(\boldsymbol{q},\omega)\,\big(\delta_{ij} - \hat{q}_i\,\hat{q}_j\big) \quad , \tag{10.215}$$

where $\hat{q}_i \equiv q_i/|\boldsymbol{q}|$. We now invoke current conservation, which says $q^\mu\, K_{\mu\nu}(q) = 0$. When $\nu = j$ is a spatial index,

$$q^0\, K_{0j} + q^i\, K_{ij} = \frac{\omega}{c}\, K_{0j} + K_{\parallel}\, q_j \quad , \tag{10.216}$$

which yields

$$K_{0j}(\boldsymbol{q},\omega) = -\frac{c}{\omega}\, q^j\, K_{\parallel}(\boldsymbol{q},\omega) = K_{j0}(\boldsymbol{q},\omega) \quad . \tag{10.217}$$

In other words, the three components of $K_{0j}(q)$ are in fact completely determined by $K_{\parallel}(q)$ and q itself. When $\nu = 0$,

$$0 = q^0\, K_{00} + q^i\, K_{i0} = \frac{\omega}{c}\, K_{00} - \frac{c}{\omega}\, \boldsymbol{q}^2\, K_{\parallel} \quad , \tag{10.218}$$

which says

$$K_{00}(\boldsymbol{q},\omega) = \frac{c^2}{\omega^2}\, \boldsymbol{q}^2\, K_{\parallel}(\boldsymbol{q},\omega) \quad . \tag{10.219}$$

Thus, of the 10 freedoms of the symmetric 4×4 tensor $K_{\mu\nu}(q)$, there are only two independent ones – the functions $K_{\parallel}(q)$ and $K_{\perp}(q)$.

10.6.4 Neutral systems

In neutral systems, we define the number density and number current density as

$$\begin{aligned} n(\boldsymbol{x}) &= \sum_{i=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}_i) \\ \boldsymbol{j}(\boldsymbol{x}) &= \frac{1}{2m} \sum_{i=1}^{N} \Big\{ \boldsymbol{p}_i\, \delta(\boldsymbol{x} - \boldsymbol{x}_i) + \delta(\boldsymbol{x} - \boldsymbol{x}_i)\, \boldsymbol{p}_i \Big\} \quad . \end{aligned} \tag{10.220}$$

The charge and current susceptibilities are then given by

$$\begin{aligned} \chi(\boldsymbol{x},t) &= \frac{i}{\hbar} \left\langle \big[n(\boldsymbol{x},t), n(0,0) \big] \right\rangle \Theta(t) \\ \chi_{ij}(\boldsymbol{x},t) &= \frac{i}{\hbar} \left\langle \big[j_i(\boldsymbol{x},t), j_j(0,0) \big] \right\rangle \Theta(t) \quad . \end{aligned} \tag{10.221}$$

We define the longitudinal and transverse susceptibilities for homogeneous systems according to

$$\chi_{ij}(\boldsymbol{q},\omega) = \chi_{\parallel}(\boldsymbol{q},\omega)\, \hat{q}_i\, \hat{q}_j + \chi_{\perp}(\boldsymbol{q},\omega)\, (\delta_{ij} - \hat{q}_i\, \hat{q}_j) \quad . \tag{10.222}$$

From the continuity equation,

$$\boldsymbol{\nabla} \cdot \boldsymbol{j} + \frac{\partial n}{\partial t} = 0 \tag{10.223}$$

follows the relation

$$\chi_{\parallel}(\boldsymbol{q},\omega) = \frac{n}{m} + \frac{\omega^2}{\boldsymbol{q}^2}\, \hat{\chi}(\boldsymbol{q},\omega) \quad . \tag{10.224}$$

EXERCISE: Derive eqn. (10.224).

The relation between $K_{\mu\nu}(q)$ and the neutral susceptibilities defined above is then

$$\begin{aligned} K_{00}(\boldsymbol{x},t) &= -4\pi e^2\,\chi(\boldsymbol{x},t) \\ K_{ij}(\boldsymbol{x},t) &= \frac{4\pi e^2}{c^2}\left\{\frac{n}{m}\,\delta(\boldsymbol{x})\,\delta(t) - \chi_{ij}(\boldsymbol{x},t)\right\} \quad , \end{aligned} \tag{10.225}$$

and therefore

$$\begin{aligned} K_{\parallel}(\boldsymbol{q},\omega) &= \frac{4\pi e^2}{c^2}\left\{\frac{n}{m} - \chi_{\parallel}(\boldsymbol{q},\omega)\right\} \\ K_{\perp}(\boldsymbol{q},\omega) &= \frac{4\pi e^2}{c^2}\left\{\frac{n}{m} - \chi_{\perp}(\boldsymbol{q},\omega)\right\} \quad . \end{aligned} \tag{10.226}$$

10.6.5 The Meissner effect and superfluid density

Suppose we apply an electromagnetic field $\boldsymbol{E}$. We adopt a gauge in which $A^0 = 0$, $\boldsymbol{E} = -c^{-1}\dot{\boldsymbol{A}}$, and $\boldsymbol{B} = \boldsymbol{\nabla}\times\boldsymbol{A}$. To satisfy Maxwell's equations[5], we have $\boldsymbol{q}\cdot\boldsymbol{A}(\boldsymbol{q},\omega) = 0$, *i.e.* $\boldsymbol{A}(\boldsymbol{q},\omega)$ is purely transverse. But then

$$\big\langle \boldsymbol{j}(\boldsymbol{q},\omega)\big\rangle = -\frac{c}{4\pi}\,K_{\perp}(\boldsymbol{q},\omega)\,\boldsymbol{A}(\boldsymbol{q},\omega) \quad . \tag{10.227}$$

This leads directly to the Meissner effect whenever $\lim_{\boldsymbol{q}\to 0} K_{\perp}(\boldsymbol{q},0)$ is finite. To see this, we write

$$\begin{aligned} \boldsymbol{\nabla}\times\boldsymbol{B} &= \boldsymbol{\nabla}\,(\boldsymbol{\nabla}\cdot\boldsymbol{A}) - \boldsymbol{\nabla}^2\boldsymbol{A} \\ &= \frac{4\pi}{c}\,\boldsymbol{j} + \frac{1}{c}\frac{\partial\boldsymbol{E}}{\partial t} \\ &= \frac{4\pi}{c}\Big(-\frac{c}{4\pi}\Big)\,K_{\perp}(-i\boldsymbol{\nabla}, i\,\partial_t)\,\boldsymbol{A} - \frac{1}{c^2}\frac{\partial^2\boldsymbol{A}}{\partial t^2} \quad , \end{aligned} \tag{10.228}$$

which yields

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\boldsymbol{A} = K_{\perp}(-i\boldsymbol{\nabla}, i\,\partial_t)\boldsymbol{A} \quad . \tag{10.229}$$

In the static limit, $\boldsymbol{\nabla}^2\boldsymbol{A} = K_{\perp}(-i\boldsymbol{\nabla},0)\boldsymbol{A}$, and we define

$$\frac{1}{\lambda_{\rm L}^2} \equiv \lim_{\boldsymbol{q}\to 0} K_{\perp}(\boldsymbol{q},0) \quad . \tag{10.230}$$

$\lambda_{\rm L}$ is the *London penetration depth*, which is related to the *superfluid density* $n_{\rm s}$ by

$$n_{\rm s} \equiv \frac{mc^2}{4\pi e^2\lambda_{\rm L}^2} = n - m\lim_{\boldsymbol{q}\to 0}\chi_{\perp}(\boldsymbol{q},0) \quad . \tag{10.231}$$

[5] *I.e.* if $\boldsymbol{\nabla}\cdot\boldsymbol{E} = 0$, which pertains when there is local charge neutrality.

Since the function $K_\perp(\boldsymbol{q},0)$ has dimensions of L^{-2}, assuming spatial isotropy, we may define another length scale, λ, by

$$\lambda = \frac{2}{\pi}\int\limits_0^\infty \frac{dq}{q^2 + K_\perp(q,0)} \quad . \tag{10.232}$$

Note that if $K_\perp(q,0)$ is q-independent then $\lambda = \lambda_{\rm L}$.

Ideal Bose gas

We start from

$$\begin{aligned}\chi_{ij}(\boldsymbol{q},t) &= \frac{i}{\hbar V}\left\langle\left[j_i(\boldsymbol{q},t), j_j(-\boldsymbol{q},0)\right]\right\rangle\Theta(t)\\ j_i(\boldsymbol{q}) &= \frac{\hbar}{2m}\sum_{\boldsymbol{k}}(2k_i+q_i)\,\psi^\dagger_{\boldsymbol{k}}\,\psi_{\boldsymbol{k}+\boldsymbol{q}} \quad .\end{aligned} \tag{10.233}$$

For the free Bose gas, with dispersion $\omega_{\boldsymbol{k}} = \hbar\boldsymbol{k}^2/2m$,

$$\begin{aligned}j_i(\boldsymbol{q},t) &= \frac{\hbar}{2m}\sum_{\boldsymbol{k}}(2k_i+q_i)\,e^{i(\omega_{\boldsymbol{k}}-\omega_{\boldsymbol{k}+\boldsymbol{q}})t}\,\psi^\dagger_{\boldsymbol{k}}\,\psi_{\boldsymbol{k}+\boldsymbol{q}}\\ \left[j_i(\boldsymbol{q},t), j_j(-\boldsymbol{q},0)\right] &= \frac{\hbar^2}{4m^2}\sum_{\boldsymbol{k},\boldsymbol{k}'}(2k_i+q_i)(2k'_j-q_j)\,e^{i(\omega_{\boldsymbol{k}}-\omega_{\boldsymbol{k}+\boldsymbol{q}})t}\left[\psi^\dagger_{\boldsymbol{k}}\,\psi_{\boldsymbol{k}+\boldsymbol{q}}, \psi^\dagger_{\boldsymbol{k}'}\,\psi_{\boldsymbol{k}'-\boldsymbol{q}}\right]\end{aligned} \tag{10.234}$$

Using

$$[AB, CD] = A\,[B,C]\,D + AC\,[B,D] + C\,[A,D]\,B + [A,C]\,DB \quad , \tag{10.235}$$

we obtain

$$\left[j_i(\boldsymbol{q},t), j_j(-\boldsymbol{q},0)\right] = \frac{\hbar^2}{4m^2}\sum_{\boldsymbol{k}}(2k_i+q_i)(2k_j+q_j)\,e^{i(\omega_{\boldsymbol{k}}-\omega_{\boldsymbol{k}+\boldsymbol{q}})t}\left\{n^0(\omega_{\boldsymbol{k}}) - n^0(\omega_{\boldsymbol{k}+\boldsymbol{q}})\right\} \quad , \tag{10.236}$$

where $n^0(\omega)$ is the equilibrium Bose distribution[6],

$$n^0(\omega) = \frac{1}{e^{\beta\hbar\omega}\,e^{-\beta\mu} - 1} \quad . \tag{10.237}$$

Thus,

$$\chi_{ij}(\boldsymbol{q},\omega) = \frac{\hbar}{4m^2V}\sum_{\boldsymbol{k}}(2k_i+q_i)(2k_j+q_j)\,\frac{n^0(\omega_{\boldsymbol{k}+\boldsymbol{q}}) - n^0(\omega_{\boldsymbol{k}})}{\omega + \omega_{\boldsymbol{k}} - \omega_{\boldsymbol{k}+\boldsymbol{q}} + i\epsilon} \tag{10.238}$$

$$= \frac{\hbar n_0}{4m^2}\left\{\frac{1}{\omega+\omega_{\boldsymbol{q}}+i\epsilon} - \frac{1}{\omega-\omega_{\boldsymbol{q}}+i\epsilon}\right\}q_i\,q_j + \frac{\hbar}{m^2}\int\frac{d^3k}{(2\pi)^3}\,\frac{n^0(\omega_{\boldsymbol{k}+\boldsymbol{q}/2}) - n^0(\omega_{\boldsymbol{k}-\boldsymbol{q}/2})}{\omega + \omega_{\boldsymbol{k}-\boldsymbol{q}/2} - \omega_{\boldsymbol{k}+\boldsymbol{q}/2} + i\epsilon}\,k_i\,k_j \quad ,$$

[6]Recall that $\mu = 0$ in the condensed phase.

where $n_0 = N_0/V$ is the condensate number density. Taking the $\omega = 0$, $\boldsymbol{q} \to 0$ limit yields

$$\chi_{ij}(\boldsymbol{q} \to 0, 0) = \frac{n_0}{m}\,\hat{q}_i\,\hat{q}_j + \frac{n'}{m}\,\delta_{ij} \quad , \tag{10.239}$$

where n' is the density of uncondensed bosons. From this we read off

$$\chi_{\parallel}(\boldsymbol{q} \to 0, 0) = \frac{n}{m} \qquad , \qquad \chi_{\perp}(\boldsymbol{q} \to 0, 0) = \frac{n'}{m} \quad , \tag{10.240}$$

where $n = n_0 + n'$ is the total boson number density. The superfluid density, according to (10.231), is $n_{\rm s} = n_0(T)$.

In fact, the ideal Bose gas is *not* a superfluid. Its excitation spectrum is too 'soft' - any superflow is unstable toward decay into single particle excitations.

Chapter 11

Phenomenological Theories of Superconductivity

11.1 Basic Phenomenology of Superconductors

The superconducting state is a phase of matter, as is ferromagnetism, metallicity, *etc.* The phenomenon was discovered in the Spring of 1911 by the Dutch physicist H. Kamerlingh Onnes, who observed an abrupt vanishing of the resistivity of solid mercury at $T = 4.15\,\mathrm{K}$[1]. Under ambient pressure, there are 33 elemental superconductors[2], all of which have a metallic phase at higher temperatures, and hundreds of compounds and alloys which exhibit the phenomenon. A timeline of superconductors and their critical temperatures is provided in Fig. 11.1. The related phenomenon of superfluidity was first discovered in liquid helium below $T = 2.17\,\mathrm{K}$, at atmospheric pressure, independently in 1937 by P. Kapitza (Moscow) and by J. F. Allen and A. D. Misener (Cambridge). At some level, a superconductor may be considered as a charged superfluid – we will elaborate on this statement later on. Here we recite the basic phenomenology of superconductors:

- *Vanishing electrical resistance* : The DC electrical resistance at zero magnetic field vanishes in the superconducting state. This is established in some materials to better than one part in 10^{15} of the normal state resistance. Above the critical temperature T_c, the DC resistivity at $H = 0$ is finite. The AC resistivity remains zero up to a critical frequency, $\omega_c = 2\Delta/\hbar$, where Δ is the gap in the electronic excitation spectrum. The frequency threshold is 2Δ because the superconducting condensate is made up of electron *pairs*, so breaking a pair results in two *quasiparticles*, each with energy Δ or greater. For *weak coupling* superconductors, which are described by the famous BCS theory (1957),

[1] Coincidentally, this just below the temperature at which helium liquefies under atmospheric pressure.
[2] An additional 23 elements are superconducting under high pressure.

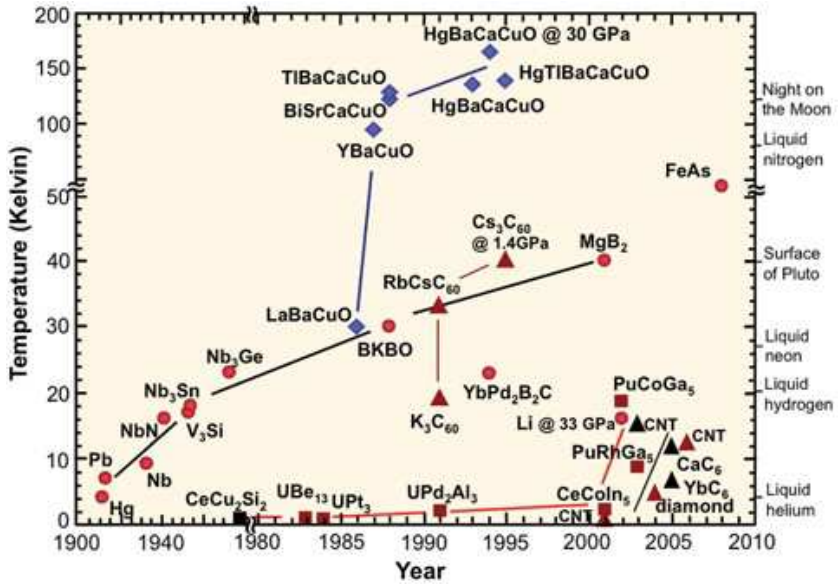


Figure 11.1: Timeline of superconductors and their transition temperatures (from Wikipedia).

there is a relation between the gap energy and the superconducting transition temperature, $2\Delta_0 = 3.5\, k_{\rm B} T_{\rm c}$, which we derive when we study the BCS model. The gap $\Delta(T)$ is temperature-dependent and vanishes at $T_{\rm c}$.

- *Flux expulsion* : In 1933 it was descovered by Meissner and Ochsenfeld that magnetic fields in superconducting tin and lead to not penetrate into the bulk of a superconductor, but rather are confined to a surface layer of thickness λ, called the *London penetration depth*. Typically λ in on the scale of tens to hundreds of nanometers.

  It is important to appreciate the difference between a superconductor and a perfect metal. If we set $\sigma = \infty$ then from $\boldsymbol{j} = \sigma \boldsymbol{E}$ we must have $\boldsymbol{E} = 0$, hence Faraday's law $\boldsymbol{\nabla} \times \boldsymbol{E} = -c^{-1} \partial_t \boldsymbol{B}$ yields $\partial_t \boldsymbol{B} = 0$, which says that $\boldsymbol{B}$ remains *constant* in a perfect metal. Yet Meissner and Ochsenfeld found that below $T_{\rm c}$ flux was *expelled* from the bulk of the superconductor, *i.e.* $\boldsymbol{B} = 0$. If, however, the superconducting sample is not simply connected, *i.e.* if it has voids, such as in the case of a superconducting ring, then in the Meissner phase flux may be trapped in the voids. Such trapped flux is quantized in integer units of the superconducting fluxoid $\phi_{\rm L} = hc/2e = 2.07 \times 10^{-7}\ {\rm G\,cm^2}$ (see Fig. 11.2).

  In the Meissner phase, Maxwell's equations imply the existence of a *screening current* at the surface of the superconductor, which extends a distance λ into the bulk. The effective screening current density is given by $\boldsymbol{K} = c \hat{\boldsymbol{n}}_{\rm s} \times \boldsymbol{H}_{\rm s}/4\pi$, where $\boldsymbol{H}_{\rm s}$ is the magnetic field outside the surface, and $\hat{\boldsymbol{n}}_{\rm s}$ is the surface normal[3]. Surface currents are also present in any

[3]If we refer to $\boldsymbol{H}$ as the *magnetic field*, we should take care to refer to $\boldsymbol{B}$ in other terms, such as the *magnetic induction*, as in volume 8 of Landau and Lifshitz. However, even authoritative sources typically are careless about this and write things like, "a magnetic field does not penetrate into the bulk of a superconductor" (as we have done above).

magnetized body, with

$$\boldsymbol{K} = \frac{c}{4\pi}\hat{\boldsymbol{n}}_{\rm s} \times (\boldsymbol{H}_{\rm s} - \boldsymbol{B}) = \frac{c}{4\pi}(\mu^{-1} - 1)\,\hat{\boldsymbol{n}}_{\rm s} \times \boldsymbol{B} \quad , \tag{11.1}$$

where $\boldsymbol{B} = \mu\boldsymbol{H}$ is the magnetic induction in the body and μ the magnetic permeability.

- *Critical field(s)* : The Meissner state exists for $T < T_{\rm c}$ only when the applied magnetic field H is smaller than the *critical field* $H_{\rm c}(T)$, with

$$H_{\rm c}(T) \simeq H_{\rm c}(0)\left(1 - \frac{T^2}{T_{\rm c}^2}\right) \quad . \tag{11.2}$$

  In so-called type-I superconductors, the system goes normal[4] for $H > H_{\rm c}(T)$. For most elemental type-I materials (*e.g.*, Hg, Pb, Nb, Sn) one has $H_{\rm c}(0) \leqslant 1\,{\rm kG}$. In type-II materials, there are two critical fields, $H_{\rm c1}(T)$ and $H_{\rm c2}(T)$. For $H < H_{\rm c1}$, we have flux expulsion, and the system is in the Meissner phase. For $H > H_{\rm c2}$, we have uniform flux penetration and the system is normal. For $H_{\rm c1} < H < H_{\rm c2}$, the system in a *mixed state* in which quantized vortices of flux $\phi_{\rm L}$ penetrate the system (see Fig. 11.3). There is a depletion of what we shall describe as the superconducting order parameter $\Psi(\boldsymbol{r})$ in the vortex cores over a length scale ξ, which is the *coherence length* of the superconductor. The upper critical field is set by the condition that the vortex cores start to overlap: $H_{\rm c2} = \phi_{\rm L}/2\pi\xi^2$. The vortex cores can be pinned by disorder. Vortices also interact with each other out to a distance λ, and at low temperatures in the absence of disorder the vortices order into a (typically triangular) *Abrikosov vortex lattice* (see Fig. 11.4). Typically one has $H_{\rm c2} = \sqrt{2}\kappa\,H_{\rm c1}$, where $\kappa = \lambda/\xi$ is a ratio of the two fundamental length scales. Type-II materials exist when $H_{\rm c2} > H_{\rm c1}$, *i.e.* when $\kappa > \frac{1}{\sqrt{2}}$. Type-II behavior tends to occur in superconducting alloys, such as Nb-Sn.

- *Persistent currents* : We have already mentioned that a metallic ring in the presence of an external magnetic field may enclosed a quantized trapped flux $n\phi_{\rm L}$ when cooled below its superconducting transition temperature. If the field is now decreased to zero, the trapped flux remains, and is generated by a *persistent current* which flows around the ring. In thick rings, such currents have been demonstrated to exist undiminished for years, and may be stable for astronomically long times.

- *Specific heat jump* : The heat capacity of metals behaves as $c_V \equiv C_V/V = \frac{\pi^2}{3}\,k_{\rm B}^2 T g(\varepsilon_{\rm F})$, where $g(\varepsilon_{\rm F})$ is the density of states at the Fermi level. In a superconductor, once one subtracts the low temperature phonon contribution $c_V^{\rm phonon} = AT^3$, one is left for $T < T_{\rm c}$ with an electronic contribution behaving as $c_V^{\rm elec} \propto e^{-\Delta/k_{\rm B}T}$. There is also a jump in the specific heat at $T = T_{\rm c}$, the magnitude of which is generally about three times the normal specific heat just above $T_{\rm c}$. This jump is consistent with a second order transition with critical exponent $\alpha = 0$.

[4]Here and henceforth, "normal" is an abbreviation for "normal metal".

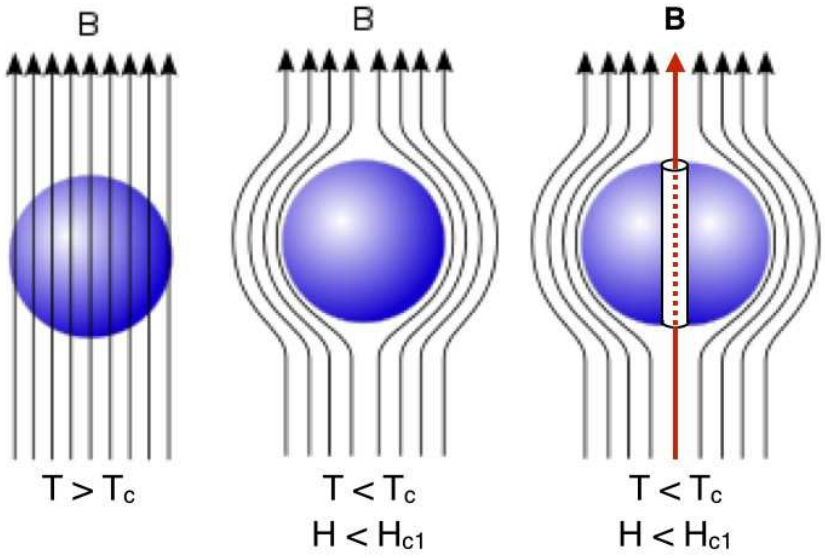


Figure 11.2: Flux expulsion from a superconductor in the Meissner state. In the right panel, quantized trapped flux penetrates a void in the sample.

- *Tunneling and Josephson effect* : The energy gap in superconductors can be measured by electron tunneling between a superconductor and a normal metal, or between two superconductors separated by an insulating layer. In the case of a weak link between two superconductors, current can flow at zero bias voltage, a situation known as the *Josephson effect*.

11.2 Thermodynamics of Superconductors

The differential free energy density of a magnetic material is given by

$$df = -s\,dT + \frac{1}{4\pi}\,\boldsymbol{H}\cdot d\boldsymbol{B} \quad , \tag{11.3}$$

which says that $f = f(T, \boldsymbol{B})$. Here s is the entropy density, and $\boldsymbol{B}$ the magnetic field. The quantity $\boldsymbol{H}$ is called the *magnetizing field* and is thermodynamically conjugate to $\boldsymbol{B}$:

$$s = -\left(\frac{\partial f}{\partial T}\right)_{\boldsymbol{B}} \quad , \quad \boldsymbol{H} = 4\pi\left(\frac{\partial f}{\partial \boldsymbol{B}}\right)_{T} \quad . \tag{11.4}$$

In the Ampère-Maxwell equation, $\nabla\times\boldsymbol{H} = 4\pi c^{-1}\boldsymbol{j}_{\rm ext} + c^{-1}\partial_t\boldsymbol{D}$, the sources of $\boldsymbol{H}$ appear on the RHS[5]. Usually $c^{-1}\partial_t\boldsymbol{D}$ is negligible, in which case $\boldsymbol{H}$ is generated by external sources such as magnetic solenoids. The magnetic field $\boldsymbol{B}$ is given by $\boldsymbol{B} = \boldsymbol{H} + 4\pi\boldsymbol{M} \equiv \mu\boldsymbol{H}$, where $\boldsymbol{M}$ is

[5]Throughout these notes, RHS/LHS will be used to abbreviate "right/left hand side".

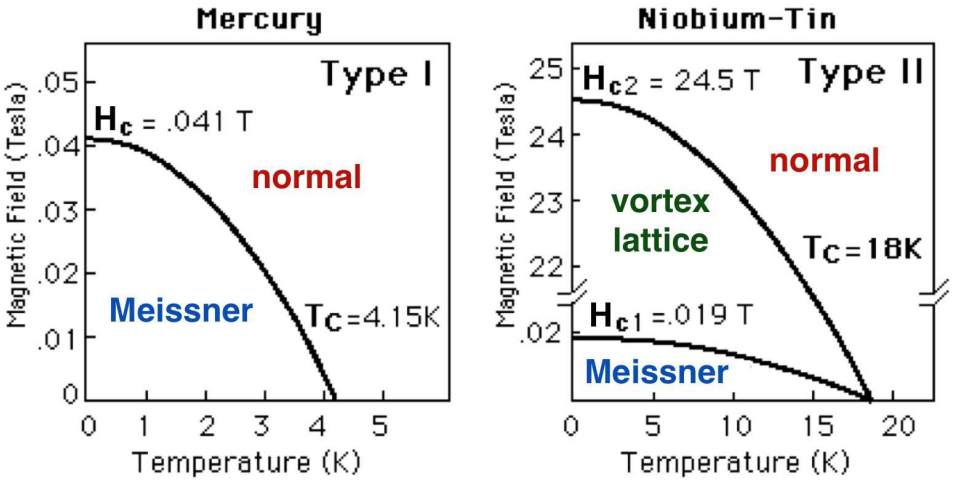


Figure 11.3: Phase diagrams for type I and type II superconductors in the (T, H) plane.

the magnetization density. We therefore have no direct control over $\boldsymbol{B}$, and it is necessary to discuss the thermodynamics in terms of the Gibbs free energy density, $g(T, \boldsymbol{H})$:

$$\begin{aligned} g(T, \boldsymbol{H}) &= f(T, \boldsymbol{B}) - \frac{1}{4\pi} \boldsymbol{B} \cdot \boldsymbol{H} \\ dg &= -s\, dT - \frac{1}{4\pi}\, \boldsymbol{B} \cdot d\boldsymbol{H} \quad . \end{aligned} \tag{11.5}$$

Thus,

$$s = -\left(\frac{\partial g}{\partial T}\right)_{\boldsymbol{H}} \quad , \quad \boldsymbol{B} = -4\pi \left(\frac{\partial g}{\partial \boldsymbol{H}}\right)_{T} \quad . \tag{11.6}$$

Assuming a bulk sample which is isotropic, we then have

$$g(T, H) = g(T, 0) - \frac{1}{4\pi} \int_0^H dH'\, B(H') \quad . \tag{11.7}$$

In a normal metal, $\mu \approx 1$ (cgs units), which means $B \approx H$, which yields

$$g_{\rm n}(T, H) = g_{\rm n}(T, 0) - \frac{H^2}{8\pi} \quad . \tag{11.8}$$

In the Meissner phase of a superconductor, $B = 0$, so

$$g_{\rm s}(T, H) = g_{\rm s}(T, 0) \quad . \tag{11.9}$$

For a type-I material, the free energies cross at $H = H_{\rm c}$, so

$$g_{\rm s}(T, 0) = g_{\rm n}(T, 0) - \frac{H_{\rm c}^2}{8\pi} \quad , \tag{11.10}$$

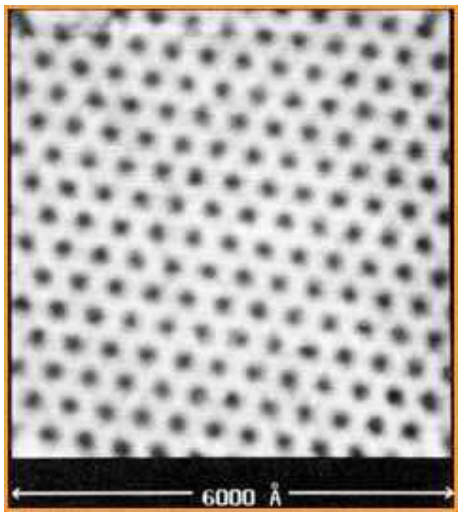


Figure 11.4: STM image of a vortex lattice in $NbSe_2$ at $H = 1\,\mathrm{T}$ and $T = 1.8\,\mathrm{K}$. From H. F. Hess *et al.*, *Phys. Rev. Lett.* **62**, 214 (1989).

which says that there is a negative *condensation energy density* $\varepsilon_{\rm cond}(T) = -H_{\rm c}^2(T)/8\pi$ which stabilizes the superconducting phase. We may now write

$$g_{\rm s}(T,H) - g_{\rm n}(T,H) = \frac{1}{8\pi}\Big(H^2 - H_{\rm c}^2(T)\Big) \quad , \tag{11.11}$$

so the superconductor is the equilibrium state for $H < H_{\rm c}$. Taking the derivative with respect to temperature, the entropy difference is given by

$$s_{\rm s}(T,H) - s_{\rm n}(T,H) = \frac{1}{4\pi}\,H_{\rm c}(T)\,\frac{dH_{\rm c}(T)}{dT} < 0 \quad , \tag{11.12}$$

since $H_{\rm c}(T)$ is a decreasing function of temperature. Note that the entropy difference is independent of the external magnetizing field H. As we see from Fig. 11.3, the derivative $H_{\rm c}'(T)$ changes discontinuously at $T = T_{\rm c}$. The latent heat $\ell = T\,\Delta s$ vanishes because $H_{\rm c}(T_{\rm c})$ itself vanishes, but the specific heat is discontinuous:

$$c_{\rm s}(T_{\rm c},H=0) - c_{\rm n}(T_{\rm c},H=0) = \frac{T_{\rm c}}{4\pi}\left(\frac{dH_{\rm c}(T)}{dT}\right)^2_{T_{\rm c}} \quad , \tag{11.13}$$

and from the phenomenological relation of Eqn. 11.2, we have $H_{\rm c}'(T_{\rm c}) = -2H_{\rm c}(0)/T_{\rm c}$, hence

$$\Delta c \equiv c_{\rm s}(T_{\rm c},H=0) - c_{\rm n}(T_{\rm c},H=0) = \frac{H_{\rm c}^2(0)}{\pi T_{\rm c}} \quad . \tag{11.14}$$

We can appeal to Eqn. 11.12 to compute the difference $\Delta c(T,H)$ for general $T < T_{\rm c}$:

$$\Delta c(T,H) = \frac{T}{8\pi}\frac{d^2}{dT^2}\,H_{\rm c}^2(T) \quad . \tag{11.15}$$

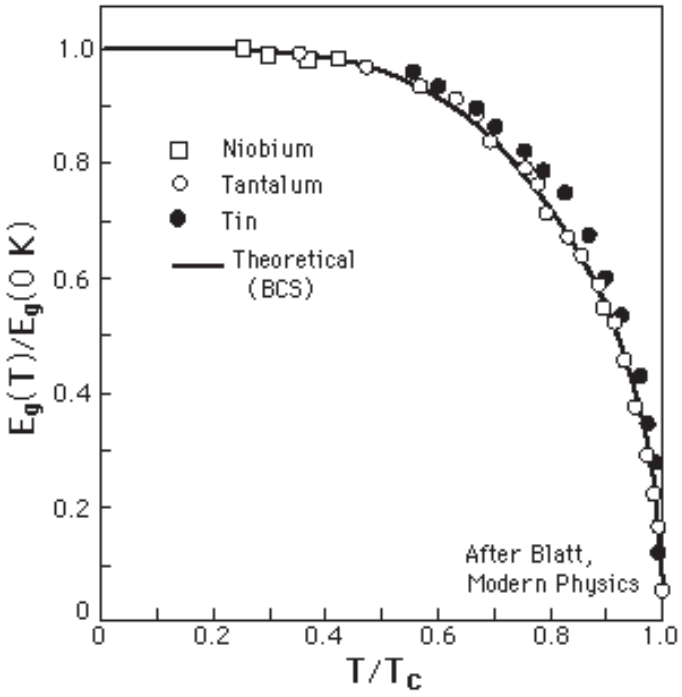


Figure 11.5: Dimensionless energy gap $\Delta(T)/\Delta_0$ in niobium, tantalum, and tin. The solid curve is the prediction from BCS theory, derived in chapter 3 below.

With the approximation of Eqn. 11.2, we obtain

$$c_{\rm s}(T,H) - c_{\rm n}(T,H) \simeq \frac{TH_{\rm c}^2(0)}{2\pi T_{\rm c}^2}\left\{3\left(\frac{T}{T_{\rm c}}\right)^2 - 1\right\} \quad . \tag{11.16}$$

In the limit $T \to 0$, we expect $c_{\rm s}(T)$ to vanish exponentially as $e^{-\Delta/k_{\rm B}T}$, hence we have $\Delta c(T \to 0) = -\gamma T$, where γ is the coefficient of the linear T term in the metallic specific heat. Thus, we expect $\gamma \simeq H_{\rm c}^2(0)/2\pi T_{\rm c}^2$. Note also that this also predicts the ratio $\Delta c(T_{\rm c},0)/c_{\rm n}(T_{\rm c},0) = 2$. In fact, within BCS theory, as we shall later show, this ratio is approximately 1.43. BCS also yields the low temperature form

$$H_{\rm c}(T) = H_{\rm c}(0)\left\{1 - \alpha\left(\frac{T}{T_{\rm c}}\right)^2 + \mathcal{O}\left(e^{-\Delta/k_{\rm B}T}\right)\right\} \tag{11.17}$$

with $\alpha \simeq 1.07$. Thus, $H_{\rm c}^{\rm BCS}(0) = \left(2\pi\gamma T_{\rm c}^2/\alpha\right)^{1/2}$.

11.3 London Theory

Fritz and Heinz London in 1935 proposed a two fluid model for the macroscopic behavior of superconductors. The two fluids are: (i) the normal fluid, with electron number density $n_{\rm n}$, which has finite resistivity, and (ii) the superfluid, with electron number density $n_{\rm s}$, and which moves with zero resistance. The associated velocities are $v_{\rm n}$ and $v_{\rm s}$, respectively. Thus, the total number density and current density are

$$\begin{aligned} n &= n_{\rm n} + n_{\rm s} \\ \boldsymbol{j} &= \boldsymbol{j}_{\rm n} + \boldsymbol{j}_{\rm s} = -e\big(n_{\rm n}\boldsymbol{v}_{\rm n} + n_{\rm s}\boldsymbol{v}_{\rm s}\big) \quad . \end{aligned} \tag{11.18}$$

The normal fluid is dissipative, hence $\boldsymbol{j}_{\rm n} = \sigma_{\rm n}\boldsymbol{E}$, but the superfluid obeys $F = m\boldsymbol{a}$, *i.e.*

$$m\,\frac{d\boldsymbol{v}_{\rm s}}{dt} = -e\boldsymbol{E} \quad \Rightarrow \quad \frac{d\boldsymbol{j}_{\rm s}}{dt} = \frac{n_{\rm s}e^2}{m}\,\boldsymbol{E} \quad . \tag{11.19}$$

In the presence of an external magnetic field, the superflow satisfies

$$\begin{aligned} \frac{d\boldsymbol{v}_{\rm s}}{dt} &= -\frac{e}{m}\big(\boldsymbol{E} + c^{-1}\boldsymbol{v}_{\rm s}\times\boldsymbol{B}\big) \\ &= \frac{\partial\boldsymbol{v}_{\rm s}}{\partial t} + (\boldsymbol{v}_{\rm s}\cdot\boldsymbol{\nabla})\boldsymbol{v}_{\rm s} = \frac{\partial\boldsymbol{v}_{\rm s}}{\partial t} + \boldsymbol{\nabla}\big(\tfrac{1}{2}\boldsymbol{v}_{\rm s}^2\big) - \boldsymbol{v}_{\rm s}\times(\boldsymbol{\nabla}\times\boldsymbol{v}_{\rm s}) \quad . \end{aligned} \tag{11.20}$$

We then have

$$\frac{\partial\boldsymbol{v}_{\rm s}}{\partial t} + \frac{e}{m}\,\boldsymbol{E} + \boldsymbol{\nabla}\big(\tfrac{1}{2}\boldsymbol{v}_{\rm s}^2\big) = \boldsymbol{v}_{\rm s}\times\left(\boldsymbol{\nabla}\times\boldsymbol{v}_{\rm s} - \frac{e\boldsymbol{B}}{mc}\right) \quad . \tag{11.21}$$

Taking the curl, and invoking Faraday's law $\boldsymbol{\nabla}\times\boldsymbol{E} = -c^{-1}\partial_t\boldsymbol{B}$, we obtain

$$\frac{\partial}{\partial t}\left(\boldsymbol{\nabla}\times\boldsymbol{v}_{\rm s} - \frac{e\boldsymbol{B}}{mc}\right) = \boldsymbol{\nabla}\times\left\{\boldsymbol{v}_{\rm s}\times\left(\boldsymbol{\nabla}\times\boldsymbol{v}_{\rm s} - \frac{e\boldsymbol{B}}{mc}\right)\right\} \quad , \tag{11.22}$$

which may be written as

$$\frac{\partial\boldsymbol{Q}}{\partial t} = \boldsymbol{\nabla}\times(\boldsymbol{v}_{\rm s}\times\boldsymbol{Q}) \quad , \tag{11.23}$$

where $\boldsymbol{Q} \equiv \boldsymbol{\nabla}\times\boldsymbol{v}_{\rm s} - (e/mc)\boldsymbol{B}$. Now $\boldsymbol{v}_{\rm s} = -\boldsymbol{j}_{\rm s}/n_{\rm s}e$, and if we assume $n_{\rm s}$ is uniform, upon invoking the Maxwell-Ampère law we obtain

$$\boldsymbol{v}_{\rm s} = \frac{c}{4\pi n_{\rm s}e}\left(\frac{\epsilon}{c}\frac{\partial\boldsymbol{E}}{\partial t} - \boldsymbol{\nabla}\times\boldsymbol{B}\right) \qquad \Rightarrow \qquad \boldsymbol{Q} = \frac{e\lambda_{\rm L}^2}{mc}\left(\nabla^2\boldsymbol{B} - \lambda_{\rm L}^{-2}\boldsymbol{B} - \frac{\epsilon}{c^2}\,\frac{\partial^2\boldsymbol{B}}{\partial t^2}\right) \quad , \tag{11.24}$$

where ϵ is the dielectric constant and $\lambda_{\rm L} = \sqrt{mc^2/4\pi n_{\rm s}e^2}$ is the *London penetration depth*. Typically we can neglect the time-dependence of $\boldsymbol{E}$.

Eqn. 11.23 says that if $\boldsymbol{Q} = 0$, it remains zero for all time. Assumption: the equilibrium state has $\boldsymbol{Q} = 0$, which entails $\nabla^2 \boldsymbol{B} = \lambda_{\rm L}^{-2} \boldsymbol{B}$. The magnetic field can only penetrate up to a distance on the order of $\lambda_{\rm L}$ inside the superconductor. This is the essence of the Meissner effect.

Note that

$$\boldsymbol{\nabla} \times \boldsymbol{j}_{\rm s} = -\frac{c}{4\pi\lambda_{\rm L}^2}\, \boldsymbol{B} \tag{11.25}$$

and the definition $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A}$ licenses us to write

$$\boldsymbol{j}_{\rm s} = -\frac{c}{4\pi\lambda_{\rm L}^2}\, \boldsymbol{A} \quad , \tag{11.26}$$

provided an appropriate gauge choice for $\boldsymbol{A}$ is taken. Since $\boldsymbol{\nabla} \cdot \boldsymbol{j}_{\rm s} = 0$ in steady state, we conclude $\boldsymbol{\nabla} \cdot \boldsymbol{A} = 0$ is the proper gauge; this is called the Coulomb gauge. The normal component of the superfluid at the boundary, $\boldsymbol{j}_{{\rm s},\perp}$, must vanish, hence $\boldsymbol{A}_\perp = 0$ as well. However, we may still introduce a 'little gauge transformation' $\boldsymbol{A} \to \boldsymbol{A} + \boldsymbol{\nabla}\chi$ provided $\nabla^2\chi = 0$. Consider now an isolated body which is simply connected, *i.e.* any closed loop drawn within the body is continuously contractable to a point. Therefore $\partial_\perp \chi$ must also vanish everywhere on the boundary, which says that χ is determined up to a global constant.

If the superconductor is multiply connected, though, the condition $\partial_\perp \chi = 0$ allows for non-constant solutions for χ. The line integral of $\boldsymbol{A}$ around a closed loop surrounding a void $\mathcal{D}$ in the superconductor is, by Stokes' theorem, the magnetic flux through the loop:

$$\oint\limits_{\partial\mathcal{D}} d\boldsymbol{l} \cdot \boldsymbol{A} = \int\limits_{\mathcal{D}} dS\, \hat{\boldsymbol{n}} \cdot \boldsymbol{B} = \Phi_{\mathcal{D}} \quad . \tag{11.27}$$

On the other hand, within the interior of the superconductor, since $\boldsymbol{B} = \boldsymbol{\nabla} \times \boldsymbol{A} = 0$, we can write $\boldsymbol{A} = \boldsymbol{\nabla}\chi$, which says that the trapped flux $\Phi_{\mathcal{D}}$ is given by $\Phi_{\mathcal{D}} = \Delta\chi$, then change in the gauge function as one proceeds counterclockwise around the loop. F. London argued that if the gauge transformation $\boldsymbol{A} \to \boldsymbol{A} + \boldsymbol{\nabla}\chi$ is associated with a quantum mechanical wavefunction associated with a charge e object, then the flux $\Phi_{\mathcal{D}}$ will be quantized in units of the Dirac flux quantum $\phi_0 = hc/e = 4.137 \times 10^{-7}\,{\rm G\,cm^2}$. The argument is simple. The transformation of the wavefunction $\Psi \to \Psi\, e^{-i\alpha}$ is cancelled by the replacement $\boldsymbol{A} \to \boldsymbol{A} + (\hbar c/e)\boldsymbol{\nabla}\alpha$. Thus, we have $\chi = \alpha\phi_0/2\pi$, and single-valuedness requires $\Delta\alpha = 2\pi n$ around a loop, hence $\Phi_{\mathcal{D}} = \Delta\chi = n\phi_0$.

The above argument is almost correct. The final piece was put in place by Lars Onsager in 1953. Onsager pointed out that if the particles described by the superconducting wavefunction Ψ were of charge $e^* = 2e$, then, *mutatis mutandis*, one would conclude the quantization condition is $\Phi_{\mathcal{D}} = n\phi_{\rm L}$, where $\phi_{\rm L} = hc/2e$ is the London flux quantum, which is half the size of the Dirac flux quantum. This suggestion was confirmed in subsequent experiments by Deaver and Fairbank, and by Doll and Näbauer, both in 1961.

11.3.1 De Gennes' derivation of London Theory

De Gennes writes the total free energy of the superconductor as

$$\begin{aligned} F &= \int d^3x\ f_0 + E_{\rm kinetic} + E_{\rm field} \\ E_{\rm kinetic} &= \int d^3x\ \tfrac{1}{2} m n_{\rm s} \boldsymbol{v}_{\rm s}^2(\boldsymbol{x}) = \int d^3x\ \frac{m}{2n_{\rm s}e^2}\,\boldsymbol{j}_{\rm s}^2(\boldsymbol{x}) \\ E_{\rm field} &= \int d^3x\ \frac{\boldsymbol{B}^2(\boldsymbol{x})}{8\pi} \quad . \end{aligned} \tag{11.28}$$

Here f_0 is the free energy density of the *metallic* state, in which no currents flow. What makes this a model of a superconductor is the assumption that a current $\boldsymbol{j}_{\rm s}$ flows in the presence of a magnetic field. Thus, under steady state conditions $\boldsymbol{\nabla}\times\boldsymbol{B} = 4\pi c^{-1}\boldsymbol{j}_{\rm s}$, so

$$F = \int d^3x \left\{ f_0 + \frac{\boldsymbol{B}^2}{8\pi} + \lambda_{\rm L}^2\,\frac{(\boldsymbol{\nabla}\times\boldsymbol{B})^2}{8\pi} \right\} \quad . \tag{11.29}$$

Taking the functional variation and setting it to zero,

$$4\pi\,\frac{\delta F}{\delta \boldsymbol{B}} = \boldsymbol{B} + \lambda_{\rm L}^2\,\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{B}) = \boldsymbol{B} - \lambda_{\rm L}^2\,\nabla^2\boldsymbol{B} = 0 \quad . \tag{11.30}$$

11.3.2 Pippard's nonlocal extension

The London equation $\boldsymbol{j}_{\rm s}(\boldsymbol{x}) = -c\boldsymbol{A}(\boldsymbol{x})/4\pi\lambda_{\rm L}^2$ says that the supercurrent is perfectly yoked to the vector potential, and on arbitrarily small length scales. This is unrealistic. A. B. Pippard undertook a phenomenological generalization of the (phenomenological) London equation, writing[6]

$$\begin{aligned} j_{\rm s}^\alpha(\boldsymbol{x}) &= -\frac{c}{4\pi\lambda_{\rm L}^2}\int d^3r\ K^{\alpha\beta}(\boldsymbol{r})\,A_\beta(\boldsymbol{x}+\boldsymbol{r}) \\ &= -\frac{c}{4\pi\lambda_{\rm L}^2}\cdot\frac{3}{4\pi\xi}\int d^3r\ \frac{e^{-r/\xi}}{r^2}\,\hat r^\alpha\,\hat r^\beta\,A_\beta(\boldsymbol{x}+\boldsymbol{r}) \quad . \end{aligned} \tag{11.31}$$

Note that the kernel $K^{\alpha\beta}(\boldsymbol{r}) = 3\,e^{-r/\xi}\,\hat r^\alpha\hat r^\beta/4\pi\xi r^2$ is normalized so that

$$\int d^3r\ K^{\alpha\beta}(\boldsymbol{r}) = \frac{3}{4\pi\xi}\int d^3r\ \frac{e^{-r/\xi}}{r^2}\,\hat r^\alpha\,\hat r^\beta = \overbrace{\frac{1}{\xi}\int_0^\infty dr\ e^{-r/\xi}}^{1}\cdot\overbrace{3\int\frac{d\hat{\boldsymbol{r}}}{4\pi}\,\hat r^\alpha\,\hat r^\beta}^{\delta^{\alpha\beta}} = \delta^{\alpha\beta} \quad . \tag{11.32}$$

[6]See A. B. Pippard, *Proc. Roy. Soc. Lond.* **A216**, 547 (1953).

The exponential factor means that $K^{\alpha\beta}(\boldsymbol{r})$ is negligible for $r \gg \xi$. If the vector potential is constant on the scale ξ, then we may pull $A_\beta(\boldsymbol{x})$ out of the integral in Eqn. 11.32, in which case we recover the original London equation. Invoking continuity in the steady state, $\boldsymbol{\nabla}\cdot\boldsymbol{j} = 0$ requires

$$\frac{3}{4\pi\xi^2}\int d^3r\,\frac{e^{-r/\xi}}{r^2}\,\hat{\boldsymbol{r}}\cdot\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{r}) = 0 \quad , \tag{11.33}$$

which is to be regarded as a gauge condition on the vector potential. One can show that this condition is equivalent to $\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$, the original Coulomb gauge.

Assuming $\boldsymbol{\nabla}\cdot\boldsymbol{A} = 0$ (London gauge), the Fourier transform of Pippard's equation yields

$$\boldsymbol{j}_{\rm s}(\boldsymbol{q}) = -\frac{c}{4\pi}\,K(\boldsymbol{q})\,\boldsymbol{A}(\boldsymbol{q}) \tag{11.34}$$

with

$$K(\boldsymbol{q}) = \frac{3}{2\lambda_{\rm L}^2}\,\frac{1}{q\xi}\left\{\frac{\tan^{-1}(q\xi)}{(q\xi)^2}\big[1+(q\xi)^2\big] - \frac{1}{q\xi}\right\} = \frac{3}{2\lambda_{\rm L}^2}\times\begin{cases}1 & q\xi \ll 1\\ \frac{1}{2}\pi/q\xi & q\xi \gg 1\end{cases} \quad . \tag{11.35}$$

In disordered superconductors, Pippard took

$$K^{\alpha\beta}(\boldsymbol{r}) = \frac{3}{4\pi\xi_0}\,\frac{e^{-r/\xi}}{r^2}\,\hat{r}^\alpha\,\hat{r}^\beta \quad , \tag{11.36}$$

with

$$\frac{1}{\xi} = \frac{1}{\xi_0} + \frac{1}{a\,\ell} \quad , \tag{11.37}$$

where ℓ is the metallic elastic mean free path, and a is a dimensionless constant on the order of unity. Note that $\int d^3r\,K^{\alpha\beta}(\boldsymbol{r}) = (\xi/\xi_0)\,\delta^{\alpha\beta}$. Thus, for $\lambda_{\rm L} \gg \xi$, one obtains an effective penetration depth $\lambda = (\xi_0/\xi)^{1/2}\lambda_{\rm L}$, where $\lambda_{\rm L} = \sqrt{mc^2/4\pi n_{\rm s}e^2}$. In the opposite limit, where $\lambda_{\rm L} \ll \xi$, Pippard found $\lambda = (3/4\pi^2)^{1/6}\big(\xi_0\lambda_{\rm L}^2\big)^{1/3}$. For strongly type-I superconductors, $\xi \gg \lambda_{\rm L}$. Since $\boldsymbol{j}_{\rm s}(\boldsymbol{x})$ is averaging the vector potential over a region of size $\xi \gg \lambda_{\rm L}$, the screening currents near the surface of the superconductor are weaker, which means the magnetic field penetrates deeper than $\lambda_{\rm L}$. The physical penetration depth is λ, where, according to Pippard, $\lambda/\lambda_{\rm L} \propto \big(\xi_0/\lambda_{\rm L}\big)^{1/3} \gg 1$.

11.4 Ginzburg-Landau Theory

The basic idea behind Ginzburg-Landau theory is to write the free energy as a simple functional of the *order parameter(s)* of a thermodynamic system and their derivatives. In ^{4}He, the order parameter $\Psi(\boldsymbol{x}) = \langle\psi(\boldsymbol{x})\rangle$ is the quantum and thermal average of the field operator $\psi(\boldsymbol{x})$ which destroys a helium atom at position $\boldsymbol{x}$. When Ψ is nonzero, we have Bose condensation with condensate density $n_0 = |\Psi|^2$. Above the lambda transition, one has $n_0(T > T_\lambda) = 0$.

In an s-wave superconductor, the order parameter field is given by

$$\Psi(\boldsymbol{x}) \propto \big\langle \psi_\uparrow(\boldsymbol{x})\,\psi_\downarrow(\boldsymbol{x}) \big\rangle \quad , \tag{11.38}$$

where $\psi_\sigma(\boldsymbol{x})$ destroys a conduction band electron of spin σ at position $\boldsymbol{x}$. Owing to the anticommuting nature of the fermion operators, the fermion field $\psi_\sigma(\boldsymbol{x})$ itself cannot condense, and it is only the *pair field* $\Psi(\boldsymbol{x})$ (and other products involving an even number of fermion field operators) which can take a nonzero value.

11.4.1 Landau theory for superconductors

The superconducting order parameter $\Psi(\boldsymbol{x})$ is thus a complex scalar, as in a superfluid. As we shall see, the difference is that the superconductor is *charged*. In the absence of magnetic fields, the Landau free energy density is approximated as

$$f = a\,|\Psi|^2 + \tfrac{1}{2}b\,|\Psi|^4 \quad . \tag{11.39}$$

The coefficients a and b are real and temperature-dependent but otherwise constant in a spatially homogeneous system. The sign of a is negotiable, but $b > 0$ is necessary for thermodynamic stability. The free energy has an $\mathrm{O}(2)$ symmetry, *i.e.* it is invariant under the substitution $\Psi \to \Psi\, e^{i\alpha}$. For $a < 0$ the free energy is minimized by writing

$$\Psi = \sqrt{-\frac{a}{b}}\; e^{i\phi} \quad , \tag{11.40}$$

where ϕ, the phase of the superconductor, is a constant. The system spontaneously breaks the $\mathrm{O}(2)$ symmetry and chooses a direction in Ψ space in which to point.

In our formulation here, the free energy of the normal state, *i.e.* when $\Psi = 0$, is $f_{\rm n} = 0$ at all temperatures, and that of the superconducting state is $f_{\rm s} = -a^2/2b$. From thermodynamic considerations, therefore, we have

$$f_{\rm s}(T) - f_{\rm n}(T) = -\frac{H_{\rm c}^2(T)}{8\pi} \quad \Rightarrow \quad \frac{a^2(T)}{b(T)} = \frac{H_{\rm c}^2(T)}{4\pi} \quad . \tag{11.41}$$

Furthermore, from London theory we have that $\lambda_{\rm L}^2 = mc^2/4\pi n_{\rm s} e^2$, and if we normalize the order parameter according to

$$\big|\Psi\big|^2 = \frac{n_{\rm s}}{n} \quad , \tag{11.42}$$

where $n_{\rm s}$ is the number density of superconducting electrons and n the total number density of conduction band electrons, then

$$\frac{\lambda_{\rm L}^2(0)}{\lambda_{\rm L}^2(T)} = \big|\Psi(T)\big|^2 = -\frac{a(T)}{b(T)} \quad . \tag{11.43}$$

Here we have taken $n_{\rm s}(T=0)=n$, so $|\Psi(0)|^2=1$. Putting this all together, we find

$$a(T) = -\frac{H_{\rm c}^2(T)}{4\pi}\cdot\frac{\lambda_{\rm L}^2(T)}{\lambda_{\rm L}^2(0)} \qquad , \qquad b(T) = \frac{H_{\rm c}^2(T)}{4\pi}\cdot\frac{\lambda_{\rm L}^4(T)}{\lambda_{\rm L}^4(0)} \tag{11.44}$$

Close to the transition, $H_{\rm c}(T)$ vanishes in proportion to $\lambda_{\rm L}^{-2}(T)$, so $a(T_{\rm c})=0$ while $b(T_{\rm c})>0$ remains finite at $T_{\rm c}$. Later on below, we shall relate the penetration depth $\lambda_{\rm L}$ to a stiffness parameter in the Ginzburg-Landau theory.

We may now compute the specific heat discontinuity from $c=-T\frac{\partial^2 f}{\partial T^2}$. It is left as an exercise to the reader to show

$$\Delta c = c_{\rm s}(T_{\rm c}) - c_{\rm n}(T_{\rm c}) = \frac{T_{\rm c}\left[a'(T_{\rm c})\right]^2}{b(T_{\rm c})} \quad , \tag{11.45}$$

where $a'(T)=da/dT$. Of course, $c_n(T)$ isn't zero! Rather, here we are accounting only for the specific heat due to that part of the free energy associated with the condensate. The Ginzburg-Landau description completely ignores the metal, and doesn't describe the physics of the normal state Fermi surface, which gives rise to $c_{\rm n}=\gamma T$. The discontinuity Δc is a mean field result. It works extremely well for superconductors, where, as we shall see, the Ginzburg criterion is satisfied down to extremely small temperature variations relative to $T_{\rm c}$. In ^{4}He, one sees an cusp-like behavior with an apparent weak divergence at the lambda transition. Recall that in the language of critical phenomena, $c(T)\propto|T-T_{\rm c}|^{-\alpha}$. For the O(2) model in $d=3$ dimensions, the exponent α is very close to zero, which is close to the mean field value $\alpha=0$. The order parameter exponent is $\beta=\frac{1}{2}$ at the mean field level; the exact value is closer to $\frac{1}{3}$. One has, for $T<T_{\rm c}$,

$$\big|\Psi(T<T_{\rm c})\big| = \sqrt{-\frac{a(T)}{b(T)}} = \sqrt{\frac{a'(T_{\rm c})}{b(T_{\rm c})}}\,(T_{\rm c}-T)^{1/2}+\ldots \quad . \tag{11.46}$$

11.4.2 Ginzburg-Landau free energy

The Landau free energy is minimized by setting $|\Psi|^2=-a/b$ for $a<0$. The phase of Ψ is therefore free to vary, and indeed free to vary independently everywhere in space. Phase fluctuations should cost energy, so we posit an augmented free energy functional,

$$F\big[\Psi,\Psi^*\big] = \int d^dx\,\Big\{a\,\big|\Psi(\boldsymbol{x})\big|^2 + \tfrac{1}{2}\,b\,\big|\Psi(\boldsymbol{x})\big|^4 + K\,\big|\boldsymbol{\nabla}\Psi(\boldsymbol{x})\big|^2+\ldots\Big\} \quad . \tag{11.47}$$

Here K is a stiffness with respect to spatial variation of the order parameter $\Psi(\boldsymbol{x})$. From K and a, we can form a length scale, $\xi=\sqrt{K/|a|}$, known as the *coherence length*. This functional in fact is very useful in discussing properties of neutral superfluids, such as ^{4}He, but superconductors are *charged*, and we have instead

$$F\big[\Psi,\Psi^*,\boldsymbol{A}\big] = \int d^dx\,\Big\{a\,\big|\Psi(\boldsymbol{x})\big|^2 + \tfrac{1}{2}\,b\,\big|\Psi(\boldsymbol{x})\big|^4 + K\,\Big|\big(\boldsymbol{\nabla}+\tfrac{ie^*}{\hbar c}\,\boldsymbol{A}\big)\Psi(\boldsymbol{x})\Big|^2 + \tfrac{1}{8\pi}\,(\boldsymbol{\nabla}\times\boldsymbol{A})^2+\ldots\Big\} \quad . \tag{11.48}$$

Here $q = -e^* = -2e$ is the *charge* of the condensate. We assume $\boldsymbol{E} = 0$, so $\boldsymbol{A}$ is not time-dependent.

Under a local transformation $\Psi(\boldsymbol{x}) \to \Psi(\boldsymbol{x})\, e^{i\alpha(\boldsymbol{x})}$, we have

$$\left(\boldsymbol{\nabla} + \tfrac{ie^*}{\hbar c}\,\boldsymbol{A}\right)\left(\Psi\, e^{i\alpha}\right) = e^{i\alpha}\left(\boldsymbol{\nabla} + i\boldsymbol{\nabla}\alpha + \tfrac{ie^*}{\hbar c}\boldsymbol{A}\right)\Psi \quad , \tag{11.49}$$

which, upon making the gauge transformation $\boldsymbol{A} \to \boldsymbol{A} - \frac{\hbar c}{e^*}\boldsymbol{\nabla}\alpha$, reverts to its original form. Thus, the free energy is unchanged upon replacing $\Psi \to \Psi e^{i\alpha}$ and $\boldsymbol{A} \to \boldsymbol{A} - \frac{\hbar c}{e^*}\boldsymbol{\nabla}\alpha$. Since gauge transformations result in no physical consequences, we conclude that the *longitudinal* phase fluctuations of a charged order parameter do not really exist. More on this later when we discuss the Anderson-Higgs mechanism.

11.4.3 Equations of motion

Varying the free energy in Eqn. 11.48 with respect to Ψ^* and $\boldsymbol{A}$, respectively, yields

$$\begin{aligned} 0 &= \frac{\delta F}{\delta \Psi^*} = a\,\Psi + b\,|\Psi|^2\,\Psi - K\left(\boldsymbol{\nabla} + \tfrac{ie^*}{\hbar c}\,\boldsymbol{A}\right)^2\Psi \\ 0 &= \frac{\delta F}{\delta \boldsymbol{A}} = \frac{2Ke^*}{\hbar c}\left[\frac{1}{2i}\left(\Psi^*\boldsymbol{\nabla}\Psi - \Psi\boldsymbol{\nabla}\Psi^*\right) + \frac{e^*}{\hbar c}\,|\Psi|^2\boldsymbol{A}\right] + \frac{1}{4\pi}\,\boldsymbol{\nabla}\times\boldsymbol{B} \quad . \end{aligned} \tag{11.50}$$

The second of these equations is the Ampère-Maxwell law, $\boldsymbol{\nabla}\times\boldsymbol{B} = 4\pi c^{-1}\boldsymbol{j}$, with

$$\boldsymbol{j} = -\frac{2Ke^*}{\hbar^2}\left[\frac{\hbar}{2i}\left(\Psi^*\boldsymbol{\nabla}\Psi - \Psi\boldsymbol{\nabla}\Psi^*\right) + \frac{e^*}{c}\,|\Psi|^2\boldsymbol{A}\right] \quad . \tag{11.51}$$

If we set Ψ to be constant, we obtain $\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{B}) + \lambda_{\rm L}^{-2}\boldsymbol{B} = 0$, with

$$\lambda_{\rm L}^{-2} = 8\pi K\left(\frac{e^*}{\hbar c}\right)^2|\Psi|^2 \quad . \tag{11.52}$$

Thus we recover the relation $\lambda_{\rm L}^{-2} \propto |\Psi|^2$. Note that $|\Psi|^2 = |a|/b$ in the ordered phase, hence

$$\lambda_{\rm L}^{-1} = \left[\frac{8\pi a^2}{b}\cdot\frac{K}{|a|}\right]^{1/2}\frac{e^*}{\hbar c} = \frac{\sqrt{2}\,e^*}{\hbar c}H_{\rm c}\,\xi \quad , \tag{11.53}$$

which says

$$H_{\rm c} = \frac{\phi_{\rm L}}{\sqrt{8}\,\pi\,\xi\lambda_{\rm L}} \quad . \tag{11.54}$$

At a superconductor-vacuum interface, we should have

$$\hat{\boldsymbol{n}} \cdot \Big(\frac{\hbar}{i} \boldsymbol{\nabla} + \frac{e^*}{c} \boldsymbol{A} \Big) \Psi \big|_{\partial\Omega} = 0 \quad , \tag{11.55}$$

where Ω denotes the superconducting region and $\hat{\boldsymbol{n}}$ denotes the surface normal. This guarantees $\hat{\boldsymbol{n}} \cdot \boldsymbol{j}\big|_{\partial\Omega} = 0$, since

$$\boldsymbol{j} = -\frac{2Ke^*}{\hbar^2} \, \mathrm{Re} \left(\frac{\hbar}{i} \, \Psi^* \boldsymbol{\nabla} \Psi + \frac{e^*}{c} \, |\Psi|^2 \boldsymbol{A} \right) \quad . \tag{11.56}$$

Note that $\hat{\boldsymbol{n}} \cdot \boldsymbol{j} = 0$ also holds if

$$\hat{\boldsymbol{n}} \cdot \Big(\frac{\hbar}{i} \boldsymbol{\nabla} + \frac{e^*}{c} \boldsymbol{A} \Big) \Psi \big|_{\partial\Omega} = ir\Psi \quad , \tag{11.57}$$

with r a real constant. This boundary condition is appropriate at a junction with a normal metal.

11.4.4 Critical current

Consider the case where $\Psi = \Psi_0$. The free energy density is

$$f = a \, |\Psi_0|^2 + \tfrac{1}{2} \, b \, |\Psi_0|^4 + K \left(\frac{e^*}{\hbar c} \right)^2 \boldsymbol{A}^2 \, |\Psi_0|^2 \quad . \tag{11.58}$$

If $a > 0$ then f is minimized for $\Psi_0 = 0$. What happens for $a < 0$, *i.e.* when $T < T_{\rm c}$? Minimizing with respect to $|\Psi_0|$, we find

$$|\Psi_0|^2 = \frac{|a| - K(e^*/\hbar c)^2 \boldsymbol{A}^2}{b} \quad . \tag{11.59}$$

The current density is then

$$\boldsymbol{j} = -2cK \left(\frac{e^*}{\hbar c} \right)^2 \left(\frac{|a| - K(e^*/\hbar c)^2 \boldsymbol{A}^2}{b} \right) \boldsymbol{A} \quad . \tag{11.60}$$

Let $A = |\boldsymbol{A}|$. Clearly we must have $0 \leqslant A \leqslant A_{\rm max}$, where

$$A_{\rm max} = \frac{\hbar c}{e^*} \cdot \sqrt{\frac{|a|}{K}} = \frac{\phi_{\rm L}}{2\pi\xi} \tag{11.61}$$

The maximum current density occurs at $A = A_{\rm c}$, where $j'(A_{\rm c}) = 0$, the solution to which is $A_{\rm c} = A_{\rm max}/\sqrt{3}$. The value of the current $j_{\rm c} \equiv j(A_{\rm c})$ is the maximum possible current, known as the *critical current density*, and is given by

$$j_{\rm c} = \frac{4}{3\sqrt{3}} \, \frac{c \, K^{1/2} \, |a|^{3/2}}{b} = \frac{2}{3\sqrt{3}} \cdot \frac{e^*}{h} \cdot H_{\rm c}^2 \, \xi \quad . \tag{11.62}$$

Physically, what is happening is this. When the kinetic energy density in the superflow exceeds the condensation energy density $H_{\rm c}^2/8\pi = a^2/2b$, the system goes normal. Note that the critical current density vanishes at the critical temperature as $j_{\rm c}(T) \propto (T_{\rm c} - T)^{3/2}$.

Should we feel bad about using a gauge-covariant variable like $\boldsymbol{A}$ in the above analysis? Not really, because when we write $\boldsymbol{A}$, what we really mean is the gauge-*invariant* combination $\boldsymbol{A} + \frac{\hbar c}{e^*}\boldsymbol{\nabla}\varphi$, where $\varphi = \arg(\Psi)$ is the phase of the order parameter.

London limit

In the so-called *London limit*, we write $\Psi = \sqrt{n_0}\, e^{i\varphi}$, with n_0 constant. Then

$$\boldsymbol{j} = -\frac{2Ke^*n_0}{\hbar}\Big(\boldsymbol{\nabla}\varphi + \frac{e^*}{\hbar c}\boldsymbol{A}\Big) = -\frac{c}{4\pi\lambda_{\rm L}^2}\left(\frac{\phi_{\rm L}}{2\pi}\,\boldsymbol{\nabla}\varphi + \boldsymbol{A}\right) \quad . \tag{11.63}$$

Thus,

$$\begin{aligned}\boldsymbol{\nabla}\times\boldsymbol{j} &= \frac{c}{4\pi}\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\boldsymbol{B}) \\ &= -\frac{c}{4\pi\lambda_{\rm L}^2}\,\boldsymbol{B} - \frac{c}{4\pi\lambda_{\rm L}^2}\,\frac{\phi_{\rm L}}{2\pi}\,\boldsymbol{\nabla}\times\boldsymbol{\nabla}\varphi \quad ,\end{aligned} \tag{11.64}$$

which says

$$\lambda_{\rm L}^2\,\nabla^2\boldsymbol{B} = \boldsymbol{B} + \frac{\phi_{\rm L}}{2\pi}\,\boldsymbol{\nabla}\times\boldsymbol{\nabla}\varphi \quad . \tag{11.65}$$

If we assume $\boldsymbol{B} = B\hat{\boldsymbol{z}}$ and the phase field φ has singular vortex lines of topological index $n_i \in \mathbb{Z}$ located at position $\boldsymbol{\rho}_i$ in the (x, y) plane, we have

$$\lambda_{\rm L}^2\,\nabla^2 B = B + \phi_{\rm L}\sum_i n_i\,\delta(\boldsymbol{\rho} - \boldsymbol{\rho}_i) \quad . \tag{11.66}$$

Taking the Fourier transform, we solve for $\hat{B}(\boldsymbol{q})$, where $\boldsymbol{k} = (\boldsymbol{q}, k_z)$:

$$\hat{B}(\boldsymbol{q}) = -\frac{\phi_{\rm L}}{1 + \boldsymbol{q}^2\lambda_{\rm L}^2}\sum_i n_i\,e^{-i\boldsymbol{q}\cdot\boldsymbol{\rho}_i} \quad , \tag{11.67}$$

whence

$$B(\boldsymbol{\rho}) = -\frac{\phi_{\rm L}}{2\pi\lambda_{\rm L}^2}\sum_i n_i\,K_0\left(\frac{|\boldsymbol{\rho} - \boldsymbol{\rho}_i|}{\lambda_{\rm L}}\right) \quad , \tag{11.68}$$

where $K_0(z)$ is the modified Bessel function, whose asymptotic behaviors are given by

$$K_0(z) \sim \begin{cases} -{\rm C} - \ln(z/2) & (z \to 0) \\ (\pi/2z)^{1/2}\exp(-z) & (z \to \infty) \end{cases} \quad , \tag{11.69}$$

where $\mathrm{C} = 0.57721566\ldots$ is the Euler-Mascheroni constant. The logarithmic divergence as $\rho \to 0$ is an artifact of the London limit. Physically, the divergence should be cut off when $|\boldsymbol{\rho} - \boldsymbol{\rho}_i| \sim \xi$. The current density for a single vortex at the origin is

$$\boldsymbol{j}(\boldsymbol{r}) = \frac{nc}{4\pi}\,\boldsymbol{\nabla} \times \boldsymbol{B} = -\frac{c}{4\pi\lambda_{\rm L}} \cdot \frac{\phi_{\rm L}}{2\pi\lambda_{\rm L}^2}\, K_1\big(\rho/\lambda_{\rm L}\big)\,\hat{\boldsymbol{\varphi}} \quad , \tag{11.70}$$

where $n \in \mathbb{Z}$ is the vorticity, and $K_1(z) = -K_0'(z)$ behaves as z^{-1} as $z \to 0$ and $\exp(-z)/\sqrt{2\pi z}$ as $z \to \infty$. Note the $i^{\rm th}$ vortex carries magnetic flux $n_i\,\phi_{\rm L}$.

11.4.5 Ginzburg criterion

Consider fluctuations in $\Psi(\boldsymbol{x})$ above $T_{\rm c}$. If $|\Psi| \ll 1$, we may neglect quartic terms and write

$$F = \int d^dx\,\Big(a\,|\Psi|^2 + K\,|\boldsymbol{\nabla}\Psi|^2\Big) = \sum_{\boldsymbol{k}} \big(a + K\boldsymbol{k}^2\big)|\hat{\Psi}(\boldsymbol{k})|^2 \quad , \tag{11.71}$$

where we have expanded

$$\Psi(\boldsymbol{x}) = \frac{1}{\sqrt{V}} \sum_{\boldsymbol{k}} \hat{\Psi}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{x}} \quad . \tag{11.72}$$

The Helmholtz free energy $A(T)$ is given by

$$e^{-A/k_{\rm B}T} = \int D[\Psi,\Psi^*]\; e^{-F/T} = \prod_{\boldsymbol{k}} \left(\frac{\pi k_{\rm B}T}{a + K\boldsymbol{k}^2}\right) \quad , \tag{11.73}$$

which is to say

$$A(T) = k_{\rm B}T \sum_{\boldsymbol{k}} \ln\left(\frac{\pi k_{\rm B}T}{a + K\boldsymbol{k}^2}\right) \quad . \tag{11.74}$$

We write $a(T) = \alpha t$ with $t = (T - T_{\rm c})/T_{\rm c}$ the reduced temperature. We now compute the singular contribution to the specific heat $C_V = -TA''(T)$, which only requires we differentiate with respect to T as it appears in $a(T)$. Dividing by $N_{\rm s}k_{\rm B}$, where $N_{\rm s} = V/\mathsf{a}^d$ is the number of lattice sites, we obtain the dimensionless heat capacity per unit cell,

$$c = \frac{\alpha^2\mathsf{a}^d}{K^2} \int^{\Lambda\xi} \frac{d^dk}{(2\pi)^d}\, \frac{1}{(\xi^{-2} + \boldsymbol{k}^2)^2} \quad , \tag{11.75}$$

where $\Lambda \sim \mathsf{a}^{-1}$ is an ultraviolet cutoff on the order of the inverse lattice spacing, and as before $\xi = (K/a)^{1/2} \propto |t|^{-1/2}$. We define $R_* \equiv (K/\alpha)^{1/2}$, in which case $\xi = R_*\,|t|^{-1/2}$, and

$$c = R_*^{-4}\,\mathsf{a}^d\,\xi^{4-d} \int^{\Lambda\xi} \frac{d^d\bar{q}}{(2\pi)^d}\, \frac{1}{(1+\bar{q}^2)^2}\;, \tag{11.76}$$

where $\bar{\boldsymbol{q}} \equiv \boldsymbol{q}\xi$. Thus,

$$c(t) \sim \begin{cases} \text{const.} & \text{if } d > 4 \\ -\ln t & \text{if } d = 4 \\ t^{\frac{d}{2}-2} & \text{if } d < 4 \, . \end{cases} \tag{11.77}$$

For $d > 4$, mean field theory is qualitatively accurate, with finite corrections. In dimensions $d \leqslant 4$, the mean field result is overwhelmed by fluctuation contributions as $t \to 0^+$ (*i.e.* as $T \to T_{\rm c}^+$). We see that the Ginzburg-Landau mean field theory is sensible provided the fluctuation contributions are small, *i.e.* provided

$$R_*^{-4} \, \mathsf{a}^d \, \xi^{4-d} \ll 1 \, , \tag{11.78}$$

which entails $t \gg t_{\rm G}$, where

$$t_{\rm G} = \left(\frac{\mathsf{a}}{R_*} \right)^{\frac{2d}{4-d}} \tag{11.79}$$

is the *Ginzburg reduced temperature*. The criterion for the sufficiency of mean field theory, namely $t \gg t_{\rm G}$, is known as the *Ginzburg criterion*. The region $|t| < t_{\rm G}$ is known as the *critical region*.

In a lattice ferromagnet, as we have seen, $R_* \sim \mathsf{a}$ is on the scale of the lattice spacing itself, hence $t_{\rm G} \sim 1$ and the critical regime is very large. Mean field theory then fails quickly as $T \to T_{\rm c}$. In a (conventional) three-dimensional superconductor, R_* is on the order of the Cooper pair size, and $R_*/\mathsf{a} \sim 10^2 - 10^3$, hence $t_{\rm G} = (a/R_*)^6 \sim 10^{-18} - 10^{-12}$ is negligibly narrow. The mean field theory of the superconducting transition – BCS theory – is then valid essentially all the way to $T = T_{\rm c}$.

Another way to think about it is as follows. In dimensions $d > 2$, for $|\boldsymbol{r}|$ fixed and $\xi \to \infty$, one has[7]

$$\big\langle \Psi^*(\boldsymbol{r}) \Psi(0) \big\rangle \simeq \frac{C_d}{k_{\rm B} T \, R_*^2} \, \frac{e^{-r/\xi}}{r^{d-2}} \quad , \tag{11.80}$$

where C_d is a dimensionless constant. If we compute the ratio of fluctuations to the mean value over a patch of linear dimension ξ, we have

$$\begin{aligned} \frac{\text{fluctuations}}{\text{mean}} &= \frac{\int\limits^{\xi} d^d r \, \big\langle \Psi^*(\boldsymbol{r}) \, \Psi(0) \big\rangle}{\int\limits^{\xi} d^d r \, \big\langle |\Psi(\boldsymbol{r})|^2 \big\rangle} \\ &\propto \frac{1}{R_*^2 \, \xi^d \, |\Psi|^2} \int\limits^{\xi} d^d r \, \frac{e^{-r/\xi}}{r^{d-2}} \propto \frac{1}{R_*^2 \, \xi^{d-2} \, |\Psi|^2} \quad . \end{aligned} \tag{11.81}$$

Close to the critical point we have $\xi \propto R_* \, |t|^{-\nu}$ and $|\Psi| \propto |t|^\beta$, with $\nu = \frac{1}{2}$ and $\beta = \frac{1}{2}$ within mean field theory. Setting the ratio of fluctuations to mean to be small, we recover the Ginzburg criterion.

[7] Exactly at $T = T_{\rm c}$, the correlations behave as $\big\langle \Psi^*(\boldsymbol{r}) \, \Psi(0) \big\rangle \propto r^{-(d-2+\eta)}$, where η is a critical exponent.

11.4.6 Domain wall solution

Consider first the simple case of the neutral superfluid. The additional parameter K provides us with a new length scale, $\xi = \sqrt{K/|a|}$, which is called the coherence length. Varying the free energy with respect to $\Psi^*(\boldsymbol{x})$, one obtains

$$\frac{\delta F}{\delta \Psi^*(\boldsymbol{x})} = a\,\Psi(\boldsymbol{x}) + b\,\big|\Psi(\boldsymbol{x})\big|^2\Psi(\boldsymbol{x}) - K\nabla^2\Psi(\boldsymbol{x}) \quad . \tag{11.82}$$

Rescaling, we write $\Psi \equiv \big(|a|/b\big)^{1/2}\psi$, and setting the above functional variation to zero, we obtain

$$-\xi^2\nabla^2\psi + \,\mathsf{sgn}(T - T_{\rm c})\,\psi + |\psi|^2\psi = 0 \quad . \tag{11.83}$$

Consider the case of a domain wall when $T < T_{\rm c}$. We assume all spatial variation occurs in the x-direction, and we set $\psi(x = 0) = 0$ and $\psi(x = \infty) = 1$. Furthermore, we take $\psi(x) = f(x)\,e^{i\alpha}$ where α is a constant[8]. We then have $-\xi^2 f''(x) - f + f^3 = 0$, which may be recast as

$$\xi^2\frac{d^2f}{dx^2} = \frac{\partial}{\partial f}\left[\tfrac{1}{4}\big(1 - f^2\big)^2\right] \quad . \tag{11.84}$$

This looks just like $F = ma$ if we regard f as the coordinate, x as time, and $-V(f) = \frac{1}{4}\big(1 - f^2\big)^2$. Thus, the potential describes an *inverted* double well with symmetric minima at $f = \pm 1$. The solution to the equations of motion is then that the 'particle' rolls starts at 'time' $x = -\infty$ at 'position' $f = +1$ and 'rolls' down, eventually passing the position $f = 0$ exactly at time $x = 0$. Multiplying the above equation by $f'(x)$ and integrating once, we have

$$\xi^2\left(\frac{df}{dx}\right)^2 = \tfrac{1}{2}\big(1 - f^2\big)^2 + C \quad , \tag{11.85}$$

where C is a constant, which is fixed by setting $f(x \to \infty) = +1$, which says $f'(\infty) = 0$, hence $C = 0$. Integrating once more,

$$f(x) = \tanh\left(\frac{x - x_0}{\sqrt{2}\,\xi}\right) \quad , \tag{11.86}$$

where x_0 is the second constant of integration. This, too, may be set to zero upon invoking the boundary condition $f(0) = 0$. Thus, the width of the domain wall is $\xi(T)$. This solution is valid provided that the local magnetic field averaged over scales small compared to ξ, *i.e.* $\boldsymbol{b} = \big\langle \boldsymbol{\nabla} \times \boldsymbol{A} \big\rangle$, is negligible.

[8]Remember that for a superconductor, phase fluctuations of the order parameter are nonphysical since they are eliiminable by a gauge transformation.

The energy per unit area of the domain wall is given by $\tilde{\sigma}$, where

$$\begin{aligned}\tilde{\sigma} &= \int_0^\infty dx \left\{ K \left| \frac{d\Psi}{dx} \right|^2 + a\,|\Psi|^2 + \tfrac{1}{2}\,b\,|\Psi|^4 \right\} \\ &= \frac{a^2}{b} \int_0^\infty dx \left\{ \xi^2 \left(\frac{df}{dx} \right)^2 - f^2 + \tfrac{1}{2}\,f^4 \right\} \quad .\end{aligned} \tag{11.87}$$

Now we ask: is domain wall formation energetically favorable in the superconductor? To answer, we compute the difference in surface energy between the domain wall state and the uniform superconducting state. We call the resulting difference σ, the true domainwall energy relative to the superconducting state:

$$\begin{aligned}\sigma &= \tilde{\sigma} - \int_0^\infty dx \left(-\frac{H_{\rm c}^2}{8\pi} \right) \\ &= \frac{a^2}{b} \int_0^\infty dx \left\{ \xi^2 \left(\frac{df}{dx} \right)^2 + \tfrac{1}{2} \left(1 - f^2 \right)^2 \right\} \equiv \frac{H_{\rm c}^2}{8\pi}\,\delta \quad ,\end{aligned} \tag{11.88}$$

where we have used $H_{\rm c}^2 = 4\pi a^2/b$. Invoking the previous result $f' = (1-f^2)/\sqrt{2}\,\xi$, the parameter δ is given by

$$\delta = 2 \int_0^\infty dx \left(1 - f^2 \right)^2 = 2 \int_0^1 df\, \frac{\left(1 - f^2 \right)^2}{f'} = \frac{4\sqrt{2}}{3}\,\xi(T) \quad . \tag{11.89}$$

Had we permitted a field to penetrate over a distance $\lambda_{\rm L}(T)$ in the domain wall state, we'd have obtained

$$\delta(T) = \frac{4\sqrt{2}}{3}\,\xi(T) - \lambda_{\rm L}(T) \quad . \tag{11.90}$$

Detailed calculations show

$$\delta = \begin{cases} \frac{4\sqrt{2}}{3}\,\xi \approx 1.89\,\xi & \text{if } \xi \gg \lambda_{\rm L} \\ 0 & \text{if } \xi = \sqrt{2}\,\lambda_{\rm L} \\ -\frac{8(\sqrt{2}-1)}{3}\,\lambda_{\rm L} \approx -1.10\,\lambda_{\rm L} & \text{if } \lambda_{\rm L} \gg \xi \quad . \end{cases} \tag{11.91}$$

Accordingly, we define the Ginzburg-Landau parameter $\kappa \equiv \lambda_{\rm L}/\xi$, which is temperature-dependent near $T = T_{\rm c}$, as we'll soon show.

So the story is as follows. In type-I materials, the positive $(\delta > 0)$ N-S surface energy keeps the sample spatially homogeneous for all $H < H_{\rm c}$. In type-II materials, the negative surface energy causes the system to break into domains, which are vortex structures, as soon as H exceeds the lower critical field $H_{\rm c1}$. This is known as the *mixed state*.

11.4.7 Scaled Ginzburg-Landau equations

For $T < T_{\rm c}$, we write

$$\Psi = \sqrt{\frac{|a|}{b}}\,\psi \quad , \quad \boldsymbol{x} = \lambda_{\rm L}\,\boldsymbol{r} \quad , \quad \boldsymbol{A} = \sqrt{2}\,\lambda_{\rm L}\,H_{\rm c}\,\boldsymbol{a} \quad , \quad \boldsymbol{H} = \sqrt{2}H_{\rm c}\boldsymbol{h} \quad , \tag{11.92}$$

as well $\boldsymbol{\partial} \equiv \lambda_{\rm L}\boldsymbol{\nabla}$. Recall the GL parameter, which is dimensionless, is given by

$$\kappa = \frac{\lambda_{\rm L}}{\xi} = \frac{\sqrt{2}\,e^*}{\hbar c}\,H_{\rm c}\,\lambda_{\rm L}^2 \quad , \tag{11.93}$$

where from Eqn. 11.54 we have $H_{\rm c} = \phi_{\rm L}/\sqrt{8}\pi\xi\lambda_{\rm L}$. The Gibbs free energy is then

$$G = \frac{H_{\rm c}^2\lambda_{\rm L}^3}{4\pi}\int d^3r\,\Big\{ -|\psi|^2 + \tfrac{1}{2}\,|\psi|^2 + \big|(\kappa^{-1}\boldsymbol{\partial} + i\boldsymbol{a})\,\psi\big|^2 + (\boldsymbol{\partial}\times\boldsymbol{a})^2 - 2\boldsymbol{h}\cdot\boldsymbol{\partial}\times\boldsymbol{a}\Big\} \quad . \tag{11.94}$$

Setting $\delta G = 0$, we obtain

$$\begin{aligned} (\kappa^{-1}\boldsymbol{\partial} + i\boldsymbol{a})^2\,\psi + \psi - |\psi|^2\psi &= 0 \\ \boldsymbol{\partial}\times(\boldsymbol{\partial}\times\boldsymbol{a} - \boldsymbol{h}) + |\psi|^2\boldsymbol{a} - \frac{i}{2\kappa}\big(\psi^*\,\boldsymbol{\partial}\,\psi - \psi\,\boldsymbol{\partial}\,\psi^*\big) &= 0 \quad . \end{aligned} \tag{11.95}$$

The condition that no current flow through the boundary is

$$\hat{\boldsymbol{n}}\cdot\big(\boldsymbol{\partial} + i\kappa\boldsymbol{a}\big)\psi\Big|_{\partial\Omega} = 0 \quad . \tag{11.96}$$

Note that the dimensionless difference in superconducting and normal state energy densities is given by

$$g_{\rm s} - g_{\rm n} = -|\psi|^2 + \tfrac{1}{2}\,|\psi|^4 + \psi^*\boldsymbol{\pi}^2\psi + (\boldsymbol{b}-\boldsymbol{h})^2 \quad , \tag{11.97}$$

where

$$\boldsymbol{\pi} = \frac{1}{i\kappa}\,\boldsymbol{\partial} + \boldsymbol{a} \quad . \tag{11.98}$$

11.5 Applications of Ginzburg-Landau Theory

The applications of GL theory are numerous. Here we run through some examples.

11.5.1 Domain wall energy

Consider a domain wall interpolating between a normal metal at $x \to -\infty$ and a superconductor at $x \to +\infty$. The difference between the Gibbs free energies is

$$\begin{aligned}\Delta G = G_{\rm s} - G_{\rm n} &= \int d^3x \left\{ a\,|\Psi|^2 + \tfrac{1}{2}b\,|\Psi|^4 + K\left|(\boldsymbol{\nabla} + \tfrac{ie^*}{\hbar c}\,\boldsymbol{A})\,\Psi\right|^2 + \frac{(B-H)^2}{8\pi}\right\} \\ &= \frac{H_{\rm c}^2\,\lambda_{\rm L}^3}{4\pi}\int d^3r\,\left[-|\psi|^2 + \tfrac{1}{2}|\psi|^4 + \left|(\kappa^{-1}\boldsymbol{\partial} + i\boldsymbol{a})\,\psi\right|^2 + (\boldsymbol{b}-\boldsymbol{h})^2\right] \quad ,\end{aligned} \tag{11.99}$$

with $|\boldsymbol{b}| = B/\sqrt{2}\,H_{\rm c}$ and $|\boldsymbol{h}| = H/\sqrt{2}\,H_{\rm c}$. Henceforth we will write $b = |\boldsymbol{b}|$ and $h = |\boldsymbol{h}|$[9]. We define

$$\Delta G(T, H_{\rm c}) \equiv \frac{H_{\rm c}^2}{8\pi}\cdot A\,\lambda_{\rm L}\cdot\delta \quad , \tag{11.100}$$

as we did above in Eqn. 11.88, except here δ is rendered dimensionless by scaling it by $\lambda_{\rm L}$. Here A is the cross-sectional area, so δ is a dimensionless domain wall energy per unit area. Integrating by parts and appealing to the Euler-Lagrange equations, we have

$$\int d^3r\,\left[-|\psi|^2+|\psi|^4+\left|(\kappa^{-1}\boldsymbol{\partial}+i\boldsymbol{a})\,\psi\right|^2\right] = \int d^3r\,\psi^*\left[-\psi+|\psi|^2\psi-(\kappa^{-1}\boldsymbol{\partial}+i\boldsymbol{a})^2\,\psi\right] = 0 \quad , \tag{11.101}$$

and therefore

$$\delta = \int\limits_{-\infty}^{\infty} dx\,\left[-|\psi|^4 + 2\,(b-h)^2\right] \quad . \tag{11.102}$$

Deep in the metal, as $x \to -\infty$, we expect $\psi \to 0$ and $b \to h$. Deep in the superconductor, as $x \to +\infty$, we expect $|\psi| \to 1$ and $b \to 0$. The bulk energy contribution then vanishes for $h = h_{\rm c} = \frac{1}{\sqrt{2}}$, which means δ is finite, corresponding to the domain wall free energy per unit area.

We take $\psi = f \in \mathbb{R}$, $\boldsymbol{a} = a(x)\,\hat{\boldsymbol{y}}$, so $\boldsymbol{b} = b(x)\,\hat{\boldsymbol{z}}$ with $b(x) = a'(x)$. Thus, $\boldsymbol{\partial}\times\boldsymbol{b} = -a''(x)\,\hat{\boldsymbol{y}}$, and the Euler-Lagrange equations are

$$\begin{aligned}\frac{1}{\kappa^2}\,\frac{d^2f}{dx^2} &= \big(a^2-1\big)f + f^3 \\ \frac{d^2a}{dx^2} &= af^2 \quad .\end{aligned} \tag{11.103}$$

These equations must be solved simultaneously to obtain the full solution. They are equivalent to a nonlinear dynamical system of dimension $N = 4$, where the phase space coordinates are

[9] Henceforth in context it should be clear when we are talking about $b = |\boldsymbol{b}|$ and when we are talking about the Ginzburg-Landau coefficient.

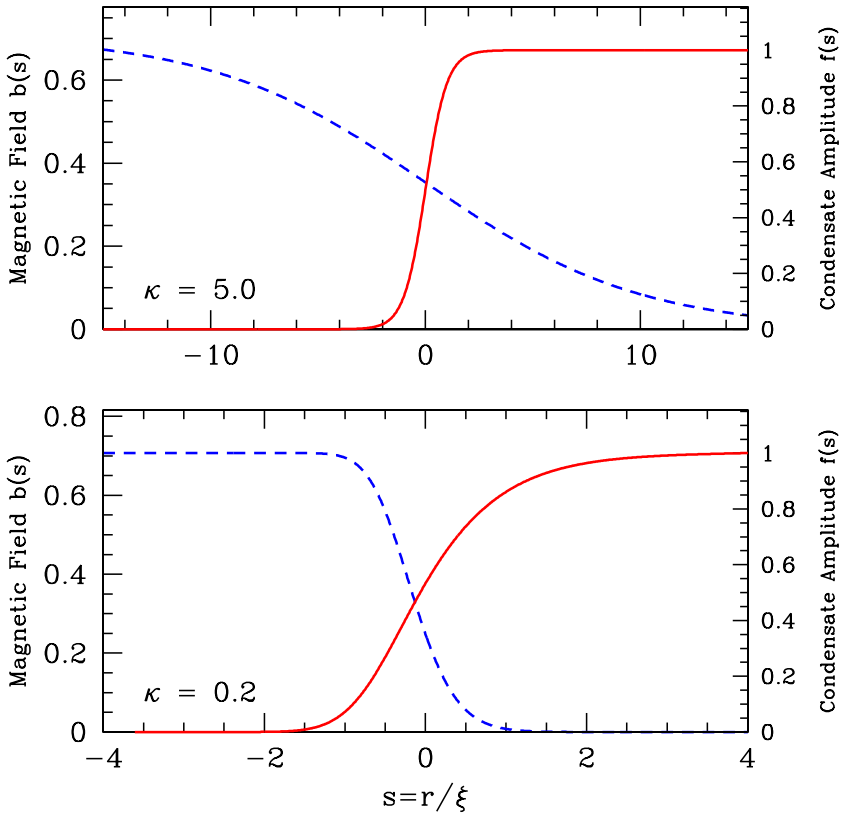


Figure 11.6: Numerical solution to a Ginzburg-Landau domain wall interpolating between normal metal ($x \to -\infty$) and superconducting ($x \to +\infty$) phases, for $H = H_{c2}$. Upper panel corresponds to $\kappa = 5$, and lower panel to $\kappa = 0.2$. Condensate amplitude $f(s)$ is shown in red, and dimensionless magnetic field $b(s) = B(s)/\sqrt{2}\,H_c$ in dashed blue.

(f, f', a, a'), *i.e.*

$$\frac{d}{dx}\begin{pmatrix} f \\ f' \\ a \\ a' \end{pmatrix} = \begin{pmatrix} f' \\ \kappa^2(a^2-1)f + \kappa^2 f^3 \\ a' \\ af^2 \end{pmatrix} \quad . \tag{11.104}$$

Four boundary conditions must be provided, which we can take to be

$$f(-\infty) = 0 \quad , \quad a'(-\infty) = \frac{1}{\sqrt{2}} \quad , \quad f(+\infty) = 1 \quad , \quad a'(+\infty) = 0 \quad . \tag{11.105}$$

Usually with dynamical systems, we specify N boundary conditions at some initial value $x = x_0$ and then integrate to the final value, using a Runge-Kutta method. Here we specify $\frac{1}{2}N$ boundary conditions at each of the two ends, which requires we use something such as the *shooting method* to solve the coupled ODEs, which effectively converts the boundary value problem to an initial value problem. In Fig. 11.6, we present such a numerical solution to the above system, for $\kappa = 0.2$ (type-I) and for $\kappa = 5$ (type-II).

Vortex solution

To describe a vortex line of strength $n \in \mathbb{Z}$, we choose cylindrical coordinates (ρ, φ, z), and assume no variation in the vertical (z) direction. We write $\psi(\boldsymbol{r}) = f(\rho)\, e^{in\varphi}$ and $\boldsymbol{a}(\boldsymbol{r}) = a(\rho)\, \hat{\boldsymbol{\varphi}}$. which says $\boldsymbol{b}(\boldsymbol{r}) = b(\rho)\, \hat{\boldsymbol{z}}$ with $b(\rho) = \frac{\partial a}{\partial \rho} + \frac{a}{\rho}$. We then obtain

$$\begin{aligned} \frac{1}{\kappa^2}\left(\frac{d^2 f}{d\rho^2} + \frac{1}{\rho}\,\frac{df}{d\rho}\right) &= \left(\frac{n}{\kappa\rho} + a\right)^2 f - f + f^3 \\ \frac{d^2 a}{d\rho^2} + \frac{1}{\rho}\,\frac{da}{d\rho} &= \frac{a}{\rho^2} + \left(\frac{n}{\kappa\rho} + a\right) f^2 \quad . \end{aligned} \tag{11.106}$$

As in the case of the domain wall, this also corresponds to an $N = 4$ dynamical system boundary value problem, which may be solved numerically using the shooting method.

11.5.2 Thin type-I films : critical field strength

Consider a thin extreme type-I (*i.e.* $\kappa \ll 1$) film. Let the finite dimension of the film be along $\hat{\boldsymbol{x}}$, and write $f = f(x)$, $\boldsymbol{a} = a(x)\, \hat{\boldsymbol{y}}$, so $\boldsymbol{\partial} \times \boldsymbol{a} = b(x)\, \hat{\boldsymbol{z}} = a'(x)\, \hat{\boldsymbol{z}}$. We assume $f(x) \in \mathbb{R}$. Now $\boldsymbol{\partial} \times \boldsymbol{b} = -a''(x)\, \hat{\boldsymbol{y}}$, so we have from the second of Eqs. 11.95 that

$$\frac{d^2 a}{dx^2} = a f^2 \quad , \tag{11.107}$$

while the first of Eqs. 11.95 yields

$$\frac{1}{\kappa^2}\,\frac{d^2 f}{dx^2} + (1 - a^2) f - f^3 = 0 \quad . \tag{11.108}$$

We require $f'(x) = 0$ on the boundaries, which we take to lie at $x = \pm\frac{1}{2}d$. For $\kappa \ll 1$, we have, to a first approximation, $f''(x) = 0$ with $f'(\pm\frac{1}{2}d) = 0$. This yields $f = f_0$, a constant, in which case $a''(x) = f_0^2\, a(x)$, yielding

$$a(x) = \frac{h_0 \sinh(f_0 x)}{f_0 \cosh(\frac{1}{2} f_0 d)} \qquad , \qquad b(x) = \frac{h_0 \cosh(f_0 x)}{\cosh(\frac{1}{2} f_0 d)} \quad , \tag{11.109}$$

with $h_0 = H_0/\sqrt{2}\, H_{\rm c}$ the scaled field outside the superconductor. Note $b(\pm\frac{1}{2}d) = h_0$. To determine the constant f_0, we set $f = f_0 + f_1$ and solve for f_1:

$$-\frac{d^2 f_1}{dx^2} = \kappa^2 \Big[\big(1 - a^2(x)\big) f_0 - f_0^3\Big] \quad . \tag{11.110}$$

In order for a solution to exist, the RHS must be orthogonal to the zeroth order solution[10], *i.e.* we demand

$$\int\limits_{-d/2}^{d/2} dx \left[1 - a^2(x) - f_0^2\right] \equiv 0 \quad , \tag{11.111}$$

which requires

$$h_0^2 = \frac{2f_0^2\,(1-f_0^2)\cosh^2(\frac{1}{2}f_0 d)}{\big[\sinh(f_0 d)/f_0 d\,\big] - 1} \quad , \tag{11.112}$$

which should be considered an implicit relation for $f_0(h_0)$. The magnetization is

$$m = \frac{1}{4\pi d}\int\limits_{-d/2}^{d/2} dx\, b(x) - \frac{h_0}{4\pi} = \frac{h_0}{4\pi}\left[\frac{\tanh(\frac{1}{2}f_0 d)}{\frac{1}{2}f_0 d} - 1\right] \quad . \tag{11.113}$$

Note that for $f_0 d \gg 1$, we recover the complete Meissner effect, $h_0 = -4\pi m$. In the opposite limit $f_0 d \ll 1$, we find

$$m \simeq -\frac{f_0^2 d^2 h_0}{48\pi} \quad , \quad h_0^2 \simeq \frac{12(1-f_0^2)}{d^2} \quad \Rightarrow \quad m \simeq -\frac{h_0 d^2}{8\pi}\left(1 - \frac{h_0^2 d^2}{12}\right) \quad . \tag{11.114}$$

Next, consider the free energy difference,

$$\begin{aligned} G_{\rm s} - G_{\rm n} &= \frac{H_{\rm c}^2 \lambda_{\rm L}^3}{4\pi}\int\limits_{-d/2}^{d/2} dx\,\Big[-f^2 + \tfrac{1}{2}f^4 + (b-h_0)^2 + \big|(\kappa^{-1}\boldsymbol{\partial} + i\boldsymbol{a})\,f\big|^2\Big] \\ &= \frac{H_{\rm c}^2\lambda_{\rm L}^3 d}{4\pi}\left[\left(1 - \frac{\tanh(f_0 d/2)}{f_0 d/2}\right)h_0^2 - f_0^2 + \tfrac{1}{2}f_0^4\right] \quad . \end{aligned} \tag{11.115}$$

The critical field $h_0 = h_{\rm c}$ occurs when $G_{\rm s} = G_{\rm n}$, hence

$$h_{\rm c}^2 = \frac{f_0^2\,(1-\frac{1}{2}f_0^2)}{\Big[1 - \frac{\tanh(f_0 d/2)}{f_0 d/2}\Big]} = \frac{2f_0^2\,(1-f_0^2)\cosh^2(f_0 d/2)}{\big[\sinh(f_0 d)/f_0 d\,\big] - 1} \quad . \tag{11.116}$$

We must eliminate f_0 to determine $h_{\rm c}(d)$.

When the film is thick we can write $f_0 = 1 - \varepsilon$ with $\varepsilon \ll 1$. Then $df_0 = d(1-\varepsilon) \gg 1$ and we have $h_{\rm c}^2 \simeq 2d\varepsilon$ and $\varepsilon = h_{\rm c}^2/2d \ll 1$. We also have

$$h_{\rm c}^2 \approx \frac{\frac{1}{2}}{1 - \frac{2}{d}} \approx \tfrac{1}{2}\Big(1 + \frac{2}{d}\Big) \quad , \tag{11.117}$$

[10] If $\hat{L}f_1 = R$, then $\langle\, f_0 \,|\, R\,\rangle = \langle\, f_0 \,|\, \hat{L} \,|\, f_1\,\rangle = \langle\, \hat{L}^\dagger f_0 \,|\, f_1\,\rangle$. Assuming $\hat{L}$ is self-adjoint, and that $\hat{L}f_0 = 0$, we obtain $\langle\, f_0 \,|\, R\,\rangle = 0$. In our case, the operator $\hat{L}$ is given by $\hat{L} = -d^2/dx^2$.

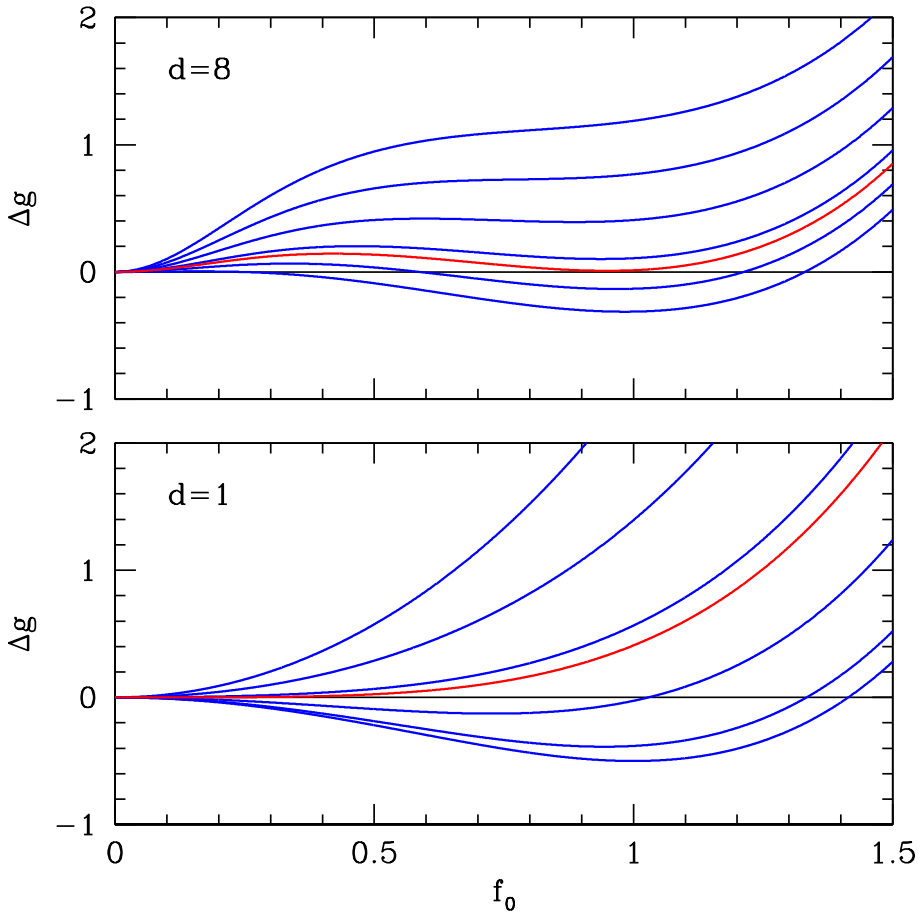


Figure 11.7: Difference in dimensionless free energy density Δg between superconducting and normal state for a thin extreme type-I film of thickness $d\lambda_{\rm L}$. Free energy curves are shown as a function of the amplitude f_0 for several values of the applied field $h_0 = H/\sqrt{2}\,H_{\rm c}(\infty)$ (upper curves correspond to larger h_0 values). Top panel: $d = 8$ curves, with the critical field (in red) at $h_{\rm c} \approx 0.827$ and a first order transition. Lower panel: $d = 1$ curves, with $h_{\rm c} = \sqrt{12} \approx 3.46$ (in red) and a second order transition. The critical thickness is $d_{\rm c} = \sqrt{5}$.

which says

$$h_{\rm c}(d) = \tfrac{1}{\sqrt{2}}\big(1 + d^{-1}\big) \quad \Rightarrow \quad H_{\rm c}(d) = H_{\rm c}(\infty)\Big(1 + \frac{\lambda_{\rm L}}{d}\Big) \quad , \tag{11.118}$$

where in the very last equation we restore dimensionful units for d.

For a thin film, we have $f_0 \approx 0$, in which case

$$h_{\rm c} = \frac{2\sqrt{3}}{d}\sqrt{1 - f_0^2} \quad , \tag{11.119}$$

and expanding the hyperbolic tangent, we find

$$h_{\rm c}^2 = \frac{12}{d^2}\big(1 - \tfrac{1}{2}f_0^2\big) \quad . \tag{11.120}$$

This gives

$$f_0 \approx 0 \quad , \quad h_{\rm c} \approx \frac{2\sqrt{3}}{d} \quad \Rightarrow \quad H_{\rm c}(d) = 2\sqrt{6}\, H_{\rm c}(\infty)\, \frac{\lambda_{\rm L}}{d} \quad . \tag{11.121}$$

Note for d large we have $f_0 \approx 1$ at the transition (first order), while for d small we have $f_0 \approx 0$ at the transition (second order). We can see this crossover from first to second order by plotting

$$g = \frac{4\pi}{d\lambda_{\rm L}^3 H_{\rm c}^3}\big(G_{\rm s} - G_{\rm n}\big) = \left(1 - \frac{\tanh(\frac{1}{2}f_0 d)}{\frac{1}{2}f_0 d}\right) h_0^2 - f_0^2 + \tfrac{1}{2}f_0^4 \tag{11.122}$$

as a function of f_0 for various values of h_0 and d. Setting $dg/df_0 = 0$ and $d^2g/df_0^2 = 0$ and $f_0 = 0$, we obtain $d_{\rm c} = \sqrt{5}$. See Fig. 11.7. For consistency, we must have $d \ll \kappa^{-1}$.

11.5.3 Critical current of a wire

Consider a wire of radius R and let the total current carried be I. The magnetizing field $\boldsymbol{H}$ is azimuthal, and integrating around the surface of the wire, we obtain

$$2\pi R H_0 = \oint\limits_{r=R} d\boldsymbol{l} \cdot \boldsymbol{H} = \int d\boldsymbol{S} \cdot \boldsymbol{\nabla} \times \boldsymbol{H} = \frac{4\pi}{c}\int d\boldsymbol{S} \cdot \boldsymbol{j} = \frac{4\pi I}{c} \quad . \tag{11.123}$$

Thus,

$$H_0 = H(R) = \frac{2I}{cR} \quad . \tag{11.124}$$

We work in cylindrical coordinates (ρ, φ, z), taking $\boldsymbol{a} = a(\rho)\,\hat{\boldsymbol{z}}$ and $f = f(\rho)$. The scaled GL equations give

$$\big(\kappa^{-1}\boldsymbol{\partial} + i\boldsymbol{a}\big)^2 f + f - f^3 = 0 \tag{11.125}$$

with[11]

$$\boldsymbol{\partial} = \hat{\boldsymbol{\rho}}\,\frac{\partial}{\partial\rho} + \frac{\hat{\boldsymbol{\varphi}}}{\rho}\,\frac{\partial}{\partial\varphi} + \hat{\boldsymbol{z}}\,\frac{\partial}{\partial z} \quad . \tag{11.126}$$

Thus,

$$\frac{1}{\kappa^2}\,\frac{\partial^2 f}{\partial\rho^2} + \big(1 - a^2\big) f - f^3 = 0 \quad , \tag{11.127}$$

with $f'(R) = 0$. From $\boldsymbol{\partial} \times \boldsymbol{b} = -\big(\kappa^{-1}\boldsymbol{\partial}\theta + \boldsymbol{a}\big)\,|\psi|^2$, where $\arg(\psi) = \theta$, we have $\psi = f \in \mathbb{R}$ hence $\theta = 0$, and therefore

$$\frac{\partial^2 a}{\partial\rho^2} + \frac{1}{\rho}\,\frac{\partial a}{\partial\rho} = a f^2 \quad . \tag{11.128}$$

The magnetic field is

$$\boldsymbol{b} = \boldsymbol{\partial} \times a(\rho)\,\hat{\boldsymbol{z}} = -\frac{\partial a}{\partial\rho}\,\hat{\boldsymbol{\varphi}} \quad , \tag{11.129}$$

[11] Though we don't need to invoke these results, it is good to recall $\frac{\partial\hat{\boldsymbol{\rho}}}{\partial\varphi} = \hat{\boldsymbol{\varphi}}$ and $\frac{\partial\hat{\boldsymbol{\varphi}}}{\partial\varphi} = -\hat{\boldsymbol{\rho}}$.

hence $b(\rho) = -\frac{\partial a}{\partial \rho}$, with

$$b(R) = \frac{H(R)}{\sqrt{2}\,H_{\rm c}} = \frac{\sqrt{2}\,I}{cRH_{\rm c}} \quad . \tag{11.130}$$

Again, we assume $\kappa \ll 1$, hence $f = f_0$ is the leading order solution to Eqn. 11.127. The vector potential and magnetic field, accounting for boundary conditions, are then given by

$$a(\rho) = -\frac{b(R)\,I_0(f_0\,\rho)}{f_0\,I_1(f_0 R)} \quad , \qquad b(\rho) = \frac{b(R)\,I_1(f_0\,\rho)}{I_1(f_0 R)} \quad , \tag{11.131}$$

where $I_n(z)$ is a modified Bessel function. As in §11.5.2, we determine f_0 by writing $f = f_0 + f_1$ and demanding that f_1 be orthogonal to the uniform solution. This yields the condition

$$\int_0^R d\rho\,\rho\left(1 - f_0^2 - a^2(\rho)\right) = 0 \quad , \tag{11.132}$$

which gives

$$b^2(R) = \frac{f_0^2\,(1-f_0^2)\,I_1^2(f_0 R)}{I_0^2(f_0 R) - I_1^2(f_0 R)} \quad . \tag{11.133}$$

Thin wire : $R \ll 1$

When $R \ll 1$, we expand the Bessel functions, using

$$I_n(z) = \left(\tfrac{1}{2}z\right)^n \sum_{k=0}^{\infty} \frac{(\frac{1}{4}z^2)^k}{k!\,(k+n)!} \quad . \tag{11.134}$$

Thus

$$\begin{aligned} I_0(z) &= 1 + \tfrac{1}{4}z^2 + \ldots \\ I_1(z) &= \tfrac{1}{2}z + \tfrac{1}{16}z^3 + \ldots \end{aligned} \quad , \tag{11.135}$$

and therefore

$$b^2(R) = \tfrac{1}{4}f_0^4\big(1 - f_0^2\big)R^2 + \mathcal{O}(R^4) \quad . \tag{11.136}$$

To determine the critical current, we demand that the maximum value of $b(\rho)$ take place at $\rho = R$, yielding

$$\frac{\partial(b^2)}{\partial f_0} = \big(f_0^3 - \tfrac{3}{2}f_0^5\big)R^2 \equiv 0 \quad \Rightarrow \quad f_{0,\rm max} = \sqrt{\tfrac{2}{3}} \quad . \tag{11.137}$$

From $f_{0,\rm max}^2 = \frac{2}{3}$, we then obtain

$$b(R) = \frac{R}{3\sqrt{3}} = \frac{\sqrt{2}\,I_{\rm c}}{cRH_{\rm c}} \quad \Rightarrow \quad I_{\rm c} = \frac{cR^2H_{\rm c}}{3\sqrt{6}} \quad . \tag{11.138}$$

The critical current density is then

$$j_{\rm c} = \frac{I_{\rm c}}{\pi R^2} = \frac{c\,H_{\rm c}}{3\sqrt{6}\,\pi\,\lambda_{\rm L}} \quad , \tag{11.139}$$

where we have restored physical units.

Thick wire : $1 \ll R \ll \kappa^{-1}$

For a thick wire, we use the asymptotic behavior of $I_n(z)$ for large argument:

$$I_\nu(z) \sim \frac{e^z}{\sqrt{2\pi z}} \sum_{k=0}^{\infty} (-1)^k \, \frac{a_k(\nu)}{z^k} \quad , \tag{11.140}$$

which is known as Hankel's expansion. The expansion coefficients are given by[12]

$$a_k(\nu) = \frac{\big(4\nu^2 - 1^2\big)\big(4\nu^2 - 3^2\big)\cdots\big(4\nu^2 - (2k-1)^2\big)}{8^k\,k!} \quad , \tag{11.141}$$

and we then obtain

$$b^2(R) = f_0^3(1 - f_0^2)R + \mathcal{O}(R^0) \quad . \tag{11.142}$$

Extremizing with respect to f_0, we obtain $f_{0,\rm max} = \sqrt{\frac{3}{5}}$ and

$$b_{\rm c}(R) = \left(\frac{4\cdot 3^3}{5^5}\right)^{1/4} R^{1/2} \quad . \tag{11.143}$$

Restoring units, the critical current of a thick wire is

$$I_{\rm c} = \frac{3^{3/4}}{5^{5/4}}\, c\, H_{\rm c}\, R^{3/2}\, \lambda_{\rm L}^{-1/2} \quad . \tag{11.144}$$

To be consistent, we must have $R \ll \kappa^{-1}$, which explains why our result here does not coincide with the bulk critical current density obtained in Eqn. 11.62.

11.5.4 Magnetic properties of type-II superconductors

Consider an incipient type-II superconductor, when the order parameter is just beginning to form. In this case we can neglect the nonlinear terms in ψ in the Ginzburg-Landau equations 11.95. The first of these equations then yields

$$\big(-i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}\big)^2\psi = \psi + \overbrace{\mathcal{O}\big(|\psi|^2\psi\big)}^{\approx 0} \quad . \tag{11.145}$$

[12] See *e.g.* the *NIST Handbook of Mathematical Functions*, §10.40.1 and §10.17.1.

We neglect the second term on the RHS. This is an eigenvalue equation, with the eigenvalue fixed at 1. In fact, this is to be regarded as an equation for $\boldsymbol{a}$, or, more precisely, for the gauge-invariant content of $\boldsymbol{a}$, which is $\boldsymbol{b} = \boldsymbol{\partial} \times \boldsymbol{a}$. The second of the GL equations says $\boldsymbol{\partial} \times (\boldsymbol{b} - \boldsymbol{h}) = \mathcal{O}\big(|\psi|^2\big)$, from which we conclude $\boldsymbol{b} = \boldsymbol{h} + \boldsymbol{\partial}\zeta$, but inspection of the free energy itself tells us $\boldsymbol{\partial}\zeta = 0$.

We assume $\boldsymbol{b} = b\hat{\boldsymbol{z}}$ and choose a gauge for $\boldsymbol{a}$:

$$\boldsymbol{a} = -\tfrac{1}{2}\, b\, y\, \hat{\boldsymbol{x}} + \tfrac{1}{2}\, b\, x\, \hat{\boldsymbol{y}} \quad , \tag{11.146}$$

with $b = h$. We define the operators

$$\pi_x = \frac{1}{i\kappa}\,\frac{\partial}{\partial x} - \tfrac{1}{2}\, b\, y \qquad , \qquad \pi_y = \frac{1}{i\kappa}\,\frac{\partial}{\partial y} + \tfrac{1}{2}\, b\, x \quad . \tag{11.147}$$

Note that $\big[\pi_x, \pi_y\big] = b/i\kappa$, and that

$$\big(-i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}\big)^2 = -\frac{1}{\kappa^2}\,\frac{\partial^2}{\partial z^2} + \pi_x^2 + \pi_y^2 \quad . \tag{11.148}$$

We now define the ladder operators

$$\begin{aligned} \gamma &= \sqrt{\frac{\kappa}{2b}}\,\big(\pi_x - i\pi_y\big) \\ \gamma^\dagger &= \sqrt{\frac{\kappa}{2b}}\,\big(\pi_x + i\pi_y\big) \quad , \end{aligned} \tag{11.149}$$

which satisfy $\big[\gamma, \gamma^\dagger\big] = 1$. Then

$$\hat{L} \equiv \big(-i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}\big)^2 = -\frac{1}{\kappa^2}\,\frac{\partial^2}{\partial z^2} + \frac{2b}{\kappa}\,\big(\gamma^\dagger\gamma + \tfrac{1}{2}\big) \quad . \tag{11.150}$$

The eigenvalues of the operator $\hat{L}$ are therefore

$$\varepsilon_n(k_z) = \frac{k_z^2}{\kappa^2} + \big(n + \tfrac{1}{2}\big)\cdot\frac{2b}{\kappa} \quad . \tag{11.151}$$

The lowest eigenvalue is therefore b/κ. This crosses the threshold value of 1 when $b = h = \kappa$, *i.e.* when

$$B = H = \sqrt{2}\,\kappa H_{\rm c} \equiv H_{\rm c2} \quad . \tag{11.152}$$

So, what have we shown? If $H_{\rm c2} < H_{\rm c}$, and therefore $\kappa < \frac{1}{\sqrt{2}}$ (we call $H_{\rm c} = \phi_{\rm L}/\sqrt{8}\pi\xi\lambda_{\rm L}$ the *thermodynamic critical field*), a *complete Meissner effect* occurs when H is decreased below the critical field $H_{\rm c}$. The order parameter ψ jumps d*iscontinuously*, and the transition across $H_{\rm c}$ is *first order*. If $\kappa > \frac{1}{\sqrt{2}}$, then $H_{\rm c2} > H_{\rm c}$, and for H just below $H_{\rm c2}$ the system wants $\psi \neq 0$.

However, a complete Meissner effect cannot occur for $H > H_{\rm c}$, so for $H_{\rm c} < H < H_{\rm c2}$ the system is in the so-called *mixed phase*. Recall again that $H_{\rm c} = \phi_{\rm L}/\sqrt{8}\,\pi\,\xi\lambda_{\rm L}$, hence

$$H_{\rm c2} = \sqrt{2}\,\kappa\,H_{\rm c} = \frac{\phi_{\rm L}}{2\pi\xi^2} \quad . \tag{11.153}$$

Thus, $H_{\rm c2}$ is the field at which neighboring vortex lines, each of which carry flux $\phi_{\rm L}$, are separated by a distance on the order of ξ.

Materials for which $\kappa < \frac{1}{\sqrt{2}}$ are called *type-I superconductors*. Materials for which $\kappa > \frac{1}{\sqrt{2}}$ are called *type-II superconductors*.

11.5.5 Lower critical field of a type-II superconductor

We now compute the energy of a perfectly straight vortex line, and ask at what field $H_{\rm c1}$ vortex lines first penetrate. Let's consider the regime $\rho > \xi$, where $\psi \simeq e^{i\varphi}$, *i.e.* $|\psi| \simeq 1$. Then the second of the Ginzburg-Landau equations gives

$$\boldsymbol{\partial} \times \boldsymbol{b} = -\big(\kappa^{-1}\boldsymbol{\partial}\varphi + \boldsymbol{a}\big) \quad . \tag{11.154}$$

Therefore the Gibbs free energy is

$$G_{\rm v} = \frac{H_{\rm c}^2\,\lambda_{\rm L}^3}{4\pi}\int d^3r\,\Big\{-\tfrac{1}{2} + \boldsymbol{b}^2 + (\boldsymbol{\partial}\times\boldsymbol{b})^2 - 2\boldsymbol{h}\cdot\boldsymbol{b}\Big\} \quad . \tag{11.155}$$

The first term in the brackets is the condensation energy density $-H_{\rm c}^2/8\pi$. The second term is the electromagnetic field energy density $\boldsymbol{B}^2/8\pi$. The third term is $\lambda_{\rm L}^2\,(\boldsymbol{\nabla}\times\boldsymbol{B})^2/8\pi$, and accounts for the kinetic energy density in the superflow.

The energy penalty for a vortex is proportional to its length. We have

$$\begin{aligned}\frac{G_{\rm v} - G_0}{L} &= \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{4\pi}\int d^2\rho\,\Big\{\boldsymbol{b}^2 + (\boldsymbol{\partial}\times\boldsymbol{b})^2 - 2\,\boldsymbol{h}\cdot\boldsymbol{b}\Big\} \\ &= \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{4\pi}\int d^2\rho\,\Big\{\boldsymbol{b}\cdot\big[\boldsymbol{b} + \boldsymbol{\partial}\times(\boldsymbol{\partial}\times\boldsymbol{b})\big] - 2\boldsymbol{h}\cdot\boldsymbol{b}\Big\} \quad .\end{aligned} \tag{11.156}$$

The total flux is

$$\int d^2\rho\,\boldsymbol{b}(\boldsymbol{\rho}) = -2\pi n\kappa^{-1}\,\hat{\boldsymbol{z}} \quad , \tag{11.157}$$

in units of $\sqrt{2}\,H_{\rm c}\,\lambda_{\rm L}^2$, where n is the integer winding of the angle variable φ about the origin. We also have $b(\boldsymbol{\rho}) = -n\kappa^{-1}K_0(\rho)$ and, taking the curl of Eqn. 11.154, we have

$$\boldsymbol{b} + \boldsymbol{\partial}\times(\boldsymbol{\partial}\times\boldsymbol{b}) = -2\pi n\kappa^{-1}\,\delta(\boldsymbol{\rho})\,\hat{\boldsymbol{z}} \quad . \tag{11.158}$$

As mentioned earlier above, the logarithmic divergence of $b(\rho \to 0)$ is an artifact of the London limit, where the vortices have no core structure. The core can crudely be accounted for by simply replacing $B(0)$ by $B(\xi)$, *i.e.* replacing $b(0)$ by $b(\xi/\lambda_{\rm L}) = b(\kappa^{-1})$. Then, for $\kappa \gg 1$, after invoking Eqn. 11.69,

$$\frac{G_{\rm v} - G_0}{L} = \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{4\pi}\left\{2\pi n^2\kappa^{-2}\ln\left(2\,e^{-{\rm C}}\kappa\right) + 4\pi n h\kappa^{-1}\right\} \quad . \tag{11.159}$$

For vortices with vorticity $n = -1$, this first turns negative at a field

$$h_{\rm c1} = \tfrac{1}{2}\kappa^{-1}\ln\left(2\,e^{-{\rm C}}\kappa\right) \quad . \tag{11.160}$$

With $2\,e^{-{\rm C}} \simeq 1.123$, we have, restoring units,

$$H_{\rm c1} = \frac{H_{\rm c}}{\sqrt{2}\,\kappa}\,\ln\left(2\,e^{-{\rm C}}\kappa\right) = \frac{\phi_{\rm L}}{4\pi\lambda_{\rm L}^2}\,\ln(1.123\,\kappa) \quad . \tag{11.161}$$

So we have

$$\begin{aligned} H_{\rm c1} &= \frac{\ln(1.123\,\kappa)}{\sqrt{2}\,\kappa}\,H_{\rm c} \quad (\kappa \gg 1) \\ H_{\rm c2} &= \sqrt{2}\,\kappa\,H_{\rm c} \quad . \end{aligned} \tag{11.162}$$

Note in general that if $E_{\rm v}$ is the energy of a single vortex, then the lower critical field is given by the relation $H_{\rm c1}\,\phi_{\rm L} = 4\pi E_{\rm v}$, *i.e.*

$$H_{\rm c1} = \frac{4\pi E_{\rm v}}{\phi_{\rm L}} \quad . \tag{11.163}$$

11.5.6 Abrikosov vortex lattice

Consider again the linearized GL equation $\left(-i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}\right)^2\psi \equiv \hat{L}\psi = \psi$ with $\boldsymbol{b} = \boldsymbol{\partial}\times\boldsymbol{a} = b\,\hat{\boldsymbol{z}}$, with $b = \kappa$, *i.e.* $B = H_{\rm c2}$. We chose the gauge $\boldsymbol{a} = \frac{1}{2}b\,(-y, x, 0)$. We showed that $\psi(\boldsymbol{\rho})$ with no z-dependence is an eigenfunction with unit eigenvalue. Recall also that $\gamma\,\psi(\boldsymbol{\rho}) = 0$, where

$$\begin{aligned} \gamma &= \frac{1}{\sqrt{2}}\left(\frac{1}{i\kappa}\,\frac{\partial}{\partial x} - \frac{\kappa}{2}\,y - \frac{1}{\kappa}\,\frac{\partial}{\partial y} - \frac{i\kappa}{2}\,x\right) \\ &= \frac{\sqrt{2}}{i\kappa}\left(\frac{\partial}{\partial w} + \tfrac{1}{4}\kappa^2\bar{w}\right) \quad , \end{aligned} \tag{11.164}$$

where $w = x + iy$ and $\bar{w} = x - iy$ are complex. We may define another ladder operator,

$$\beta = \frac{\sqrt{2}}{i\kappa}\left(\frac{\partial}{\partial\bar{w}} + \tfrac{1}{4}\kappa^2 w\right) \quad , \tag{11.165}$$

which commutes with γ and $\gamma^\dagger$, *i.e.* $[\gamma,\beta]=[\gamma,\beta^\dagger]=0$. Note that $\gamma\,\psi_0(\boldsymbol{\rho})=\beta\,\psi_0(\boldsymbol{\rho})=0$, where $\psi_0(\boldsymbol{\rho})=(\kappa/\sqrt{2\pi})\exp(-\kappa^2\boldsymbol{\rho}^2/2)$. Then we can build up the eigenspectrum of $\hat{L}=\big(-i\kappa^{-1}\boldsymbol{\partial}+\boldsymbol{a}\big)^2$ by writing

$$\psi_{m,n,k_z}(\boldsymbol{r})=\frac{e^{ik_z z}}{\sqrt{L_z}}\times\frac{(\beta^\dagger)^m\,(\gamma^\dagger)^n}{\sqrt{m!\,n!}}\,\psi_0(\boldsymbol{\rho})\quad . \tag{11.166}$$

Thus $\hat{L}\,\psi_{m,n,k_z}(\boldsymbol{r})=\varepsilon_{m,n}(k_z)\,\psi_{m,n,k_z}(\boldsymbol{r})$, with

$$\varepsilon_{m,n}(k_z)=\frac{k_z^2}{\kappa^2}+\big(2n+1\big)\cdot\frac{b}{\kappa}\quad , \tag{11.167}$$

where here we are taking $b=\kappa$. The index $n\in\{0,1,2,\ldots\}$ is the *Landau level index*. The index m is cyclic in the energy $\varepsilon_{m,n}(k_z)$, hence each Landau level is massively degenerate.

To find general solutions of $\gamma\,\psi=0$, note that

$$\gamma=\frac{\sqrt{2}}{i\kappa}\,e^{-\kappa^2\bar{w}w/4}\,\frac{\partial}{\partial w}\,e^{+\kappa^2\bar{w}w/4}\quad . \tag{11.168}$$

Thus, $\gamma\,\psi(x,y)=0$ is satisfied by any function of the form

$$\psi(x,y)=f(\bar{w})\,e^{-\kappa^2\bar{w}w/4}\quad . \tag{11.169}$$

where $f(\bar{w})$ is analytic in the complex coordinate $\bar{w}$. The most general such function[13] is of the form

$$f(\bar{w})=C\prod_{i=1}^{N_{\rm V}}(\bar{w}-\bar{w}_i)\quad , \tag{11.170}$$

where each $\bar{w}_i$ is a zero of $f(\bar{w})$. Any analytic function on the plane is, up to a constant, uniquely specified by the positions of its zeros. Note that

$$\big|\psi(x,y)\big|^2=|C|^2\,e^{-\kappa^2\bar{w}w/2}\prod_{i=1}^{N_{\rm V}}\big|w-w_i\big|^2\equiv|C|^2\,e^{-\Phi(\boldsymbol{\rho})}\quad , \tag{11.171}$$

where

$$\Phi(\boldsymbol{\rho})=\tfrac{1}{2}\kappa^2\boldsymbol{\rho}^2-2\sum_{i=1}^{N_{\rm V}}\ln\big|\boldsymbol{\rho}-\boldsymbol{\rho}_i\big|\quad . \tag{11.172}$$

$\Phi(\boldsymbol{\rho})$ may be interpreted as the electrostatic potential of a set of point charges located at $\boldsymbol{\rho}_i$, in the presence of a uniform neutralizing background. To see this, recall that $\nabla^2\ln\rho=2\pi\,\delta(\boldsymbol{\rho})$, so

$$\nabla^2\Phi(\boldsymbol{\rho})=2\kappa^2-4\pi\sum_{i=1}^{N_{\rm V}}\delta\big(\boldsymbol{\rho}-\boldsymbol{\rho}_i\big)\quad . \tag{11.173}$$

[13] We assume that ψ is square-integrable, which excludes poles in $f(\bar{w})$.

Therefore if we are to describe a state where the local density $|\psi|^2$ is uniform on average, we must impose $\big\langle \nabla^2 \Phi \big\rangle = 0$, which says

$$\Big\langle \sum_i \delta(\boldsymbol{\rho} - \boldsymbol{\rho}_i) \Big\rangle = \frac{\kappa^2}{2\pi} \quad . \tag{11.174}$$

The zeroes $\boldsymbol{\rho}_i$ are of course the positions of (anti)vortices, hence the uniform state has vortex density $n_{\rm v} = \kappa^2/2\pi$. Recall that in these units each vortex carries $2\pi/\kappa$ London flux quanta, which upon restoring units is

$$\frac{2\pi}{\kappa} \cdot \sqrt{2}\, H_{\rm c}\, \lambda_{\rm L}^2 = 2\pi \cdot \sqrt{2}\, H_{\rm c}\, \lambda_{\rm L}\, \xi = \phi_{\rm L} = \frac{hc}{e^*} \quad . \tag{11.175}$$

Multiplying the vortex density $n_{\rm v}$ by the vorticity $2\pi/\kappa$, we obtain the magnetic field strength,

$$b = h = \frac{\kappa^2}{2\pi} \times \frac{2\pi}{\kappa} = \kappa \quad . \tag{11.176}$$

In other words, $H = H_{\rm c2}$.

Aside: The freedom to choose the zeros $\{\bar{w}_i\}$ is associated with the aforementioned degeneracy of each Landau level.

Just below the upper critical field

Next, we consider the case where H is just below the upper critical field $H_{\rm c2}$. We write $\psi = \psi_0 + \delta\psi$, and $b = \kappa + \delta b$, with $\delta b < 0$. We apply the method of successive approximation, and solve for b using the second GL equation. This yields

$$b = h - \frac{|\psi_0|^2}{2\kappa} \qquad , \qquad \delta b = h - \kappa - \frac{|\psi_0|^2}{2\kappa} \tag{11.177}$$

where $\psi_0(\boldsymbol{\rho})$ is our initial solution for $\delta b = 0$. To see this, note that the second GL equation may be written

$$\boldsymbol{\partial} \times (\boldsymbol{h} - \boldsymbol{b}) = \tfrac{1}{2}\Big(\psi^* \, \boldsymbol{\pi}\, \psi + \psi\, \boldsymbol{\pi}^*\, \psi^* \Big) = \mathsf{Re}\, \Big(\psi^* \, \boldsymbol{\pi}\, \psi \Big) \quad , \tag{11.178}$$

where $\boldsymbol{\pi} = -i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}$. On the RHS we now replace ψ by ψ_0 and b by κ, corresponding to our lowest order solution. This means we write $\boldsymbol{\pi} = \boldsymbol{\pi}_0 + \delta\boldsymbol{a}$, with $\boldsymbol{\pi}_0 = -i\kappa^{-1}\boldsymbol{\partial} + \boldsymbol{a}_0$, $\boldsymbol{a}_0 = \frac{1}{2}\kappa\, \hat{\boldsymbol{z}} \times \boldsymbol{\rho}$, and $\boldsymbol{\partial} \times \delta\boldsymbol{a} = \delta b\, \hat{\boldsymbol{z}}$. Assuming $h - b = |\psi_0|^2/2\kappa$, we have

$$\begin{aligned} \boldsymbol{\partial} \times \left(\frac{|\psi_0|^2}{2\kappa} \hat{\boldsymbol{z}} \right) &= \frac{1}{2\kappa} \left[\frac{\partial}{\partial y} \left(\psi_0^* \, \psi_0 \right) \hat{\boldsymbol{x}} - \frac{\partial}{\partial x} \left(\psi_0^* \, \psi_0 \right) \hat{\boldsymbol{y}} \right] \\ &= \frac{1}{\kappa}\, \mathsf{Re} \left[\psi_0^* \, \partial_y \, \psi_0 \, \hat{\boldsymbol{x}} - \psi_0^* \, \partial_x \, \psi_0 \, \hat{\boldsymbol{y}} \right] \\ &= \mathsf{Re} \left[\psi_0^* \, i\pi_{0y} \, \psi_0 \, \hat{\boldsymbol{x}} - \psi_0^* \, i\pi_{0x} \, \psi_0 \, \hat{\boldsymbol{y}} \right] = \mathsf{Re} \left[\psi_0^* \, \boldsymbol{\pi}_0 \, \psi_0 \right] \quad , \end{aligned} \tag{11.179}$$

since $i\pi_{0y} = \kappa^{-1}\partial_y + ia_{0y}$ and $\mathrm{Re}\left[i\psi_0^*\,\psi_0\,a_{0y}\right] = 0$. Note also that since $\gamma\,\psi_0 = 0$ and $\gamma = \frac{1}{\sqrt{2}}\big(\pi_{0x} - i\pi_{0y}\big) = \frac{1}{\sqrt{2}}\pi_0^\dagger$, we have $\pi_{0y}\psi_0 = -i\pi_{0x}\psi_0$ and, equivalently, $\pi_{0x}\psi_0 = i\pi_{0y}\psi_0$.

Inserting this result into the first GL equation yields an inhomogeneous equation for $\delta\psi$. The original equation is

$$\Big(\boldsymbol{\pi}^2 - 1\Big)\psi = -|\psi|^2\psi \quad . \tag{11.180}$$

With $\boldsymbol{\pi} = \boldsymbol{\pi}_0 + \delta\boldsymbol{a}$, we then have

$$\Big(\boldsymbol{\pi}_0^2 - 1\Big)\delta\psi = -\delta\boldsymbol{a}\cdot\boldsymbol{\pi}_0\,\psi_0 - \boldsymbol{\pi}_0\cdot\delta\boldsymbol{a}\,\psi_0 - |\psi_0|^2\psi_0 \quad . \tag{11.181}$$

The RHS of the above equation must be orthogonal to ψ_0, since $\big(\boldsymbol{\pi}_0^2 - 1\big)\psi_0 = 0$. That is to say,

$$\int d^2r\;\psi_0^*\,\Big[\delta\boldsymbol{a}\cdot\boldsymbol{\pi}_0 + \boldsymbol{\pi}_0\cdot\delta\boldsymbol{a} + |\psi_0|^2\Big]\psi_0 = 0 \quad . \tag{11.182}$$

Note that

$$\delta\boldsymbol{a}\cdot\boldsymbol{\pi}_0 + \boldsymbol{\pi}_0\cdot\delta\boldsymbol{a} = \tfrac{1}{2}\,\delta a\,\pi_0^\dagger + \tfrac{1}{2}\,\pi_0^\dagger\,\delta a + \tfrac{1}{2}\,\delta\bar a\,\pi_0 + \tfrac{1}{2}\,\pi_0\,\delta\bar a \quad , \tag{11.183}$$

where

$$\pi_0 = \pi_{0x} + i\pi_{0y} \quad , \quad \pi_0^\dagger = \pi_{0x} - i\pi_{0y} \quad , \quad \delta a = \delta a_x + i\delta a_y \quad , \quad \delta\bar a = \delta a_x - i\delta a_y \quad . \tag{11.184}$$

We also have, from Eqn. 11.147,

$$\pi_0 = -2i\kappa^{-1}\big(\partial_{\bar w} - \tfrac{1}{4}\kappa^2 w\big) \qquad , \qquad \pi_0^\dagger = -2i\kappa^{-1}\big(\partial_w + \tfrac{1}{4}\kappa^2\bar w\big) \quad . \tag{11.185}$$

Note that

$$\begin{aligned}\pi_0^\dagger\,\delta a &= \big[\pi_0^\dagger\,,\,\delta a\big] + \delta a\,\pi_0^\dagger = -2i\kappa^{-1}\,\partial_w\,\delta a + \delta a\,\pi_0^\dagger \\ \delta\bar a\,\pi_0 &= \big[\delta\bar a\,,\,\pi_0\big] + \pi_0\,\delta\bar a = +2i\kappa^{-1}\,\partial_{\bar w}\,\delta\bar a + \pi_0\,\delta\bar a\end{aligned} \tag{11.186}$$

Therefore,

$$\int d^2r\;\psi_0^*\,\Big[\delta a\,\pi_0^\dagger + \pi_0\,\delta\bar a - i\kappa^{-1}\,\partial_w\,\delta a + i\kappa^{-1}\,\partial_{\bar w}\,\delta\bar a + |\psi_0|^2\Big]\psi_0 = 0 \quad . \tag{11.187}$$

We now use the fact that $\pi_0^\dagger\,\psi_0 = 0$ and $\psi_0^*\,\pi_0 = 0$ (integrating by parts) to kill off the first two terms inside the square brackets. The third and fourth term combine to give

$$-i\,\partial_w\,\delta a + i\,\partial_{\bar w}\,\delta\bar a = \partial_x\,\delta a_y - \partial_y\,\delta a_x = \delta b \quad . \tag{11.188}$$

Plugging in our expression for δb, we finally have our prize:

$$\int d^2r\,\left[\left(\frac{h}{\kappa} - 1\right)|\psi_0|^2 + \left(1 - \frac{1}{2\kappa^2}\right)|\psi_0|^4\right] = 0 \quad . \tag{11.189}$$

We may write this as

$$\left(1-\frac{h}{\kappa}\right)\big\langle|\psi_0|^2\big\rangle = \left(1-\frac{1}{2\kappa^2}\right)\big\langle|\psi_0|^4\big\rangle \quad , \tag{11.190}$$

where

$$\big\langle F(\boldsymbol{\rho})\big\rangle = \frac{1}{A}\int d^2\rho\, F(\boldsymbol{\rho}) \tag{11.191}$$

denotes the global spatial average of $F(\boldsymbol{\rho})$. It is customary to define the ratio

$$\beta_{\rm A} \equiv \frac{\big\langle|\psi_0|^4\big\rangle}{\big\langle|\psi_0|^2\big\rangle^2} \quad , \tag{11.192}$$

which depends on the distribution of the zeros $\{\boldsymbol{\rho}_i\}$. Note that

$$\big\langle|\psi_0|^2\big\rangle = \frac{1}{\beta_{\rm A}}\cdot\frac{\big\langle|\psi_0|^4\big\rangle}{\big\langle|\psi_0|^2\big\rangle} = \frac{2\kappa(\kappa-h)}{(2\kappa^2-1)\beta_{\rm A}} \quad . \tag{11.193}$$

Now let's compute the Gibbs free energy density. We have

$$\begin{aligned} g_{\rm s}-g_{\rm n} &= -\tfrac{1}{2}\big\langle|\psi_0|^4\big\rangle + \big\langle(b-h)^2\big\rangle \\ &= -\frac{1}{2}\left(1-\frac{1}{2\kappa^2}\right)\big\langle|\psi_0|^4\big\rangle = -\frac{1}{2}\left(1-\frac{h}{\kappa}\right)\big\langle|\psi_0|^2\big\rangle = -\frac{(\kappa-h)^2}{(2\kappa^2-1)\beta_{\rm A}} \quad . \end{aligned} \tag{11.194}$$

Since $g_{\rm n} = -h^2$, we have, restoring physical units

$$g_{\rm s} = -\frac{1}{8\pi}\left[H^2 + \frac{(H_{\rm c2}-H)^2}{(2\kappa^2-1)\beta_{\rm A}}\right] \quad . \tag{11.195}$$

The average magnetic field is then

$$\bar{B} = -4\pi\,\frac{\partial g_{\rm s}}{\partial H} = H - \frac{H_{\rm c2}-H}{(2\kappa^2-1)\beta_{\rm A}} \quad , \tag{11.196}$$

hence

$$M = \frac{B-H}{4\pi} = \frac{H-H_{\rm c2}}{4\pi\,(2\kappa^2-1)\,\beta_{\rm A}} \quad \Rightarrow \quad \chi = \frac{\partial M}{\partial H} = \frac{1}{4\pi\,(2\kappa^2-1)\,\beta_{\rm A}} \quad . \tag{11.197}$$

Clearly $g_{\rm s}$ is minimized by making $\beta_{\rm A}$ as small as possible, which is achieved by a regular lattice structure. Since $\beta_{\rm A}^{\rm square} = 1.18$ and $\beta_{\rm A}^{\rm triangular} = 1.16$, the triangular lattice just barely wins.

Just above the lower critical field

When H is just slightly above H_{c1}, vortex lines penetrate the superconductor, but their density is very low. To see this, we once again invoke the result of Eqn. 11.156, extending that result to the case of many vortices:

$$\frac{G_{\rm VL} - G_0}{L} = \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{4\pi}\int d^2\rho\,\Big\{\boldsymbol{b}\cdot\big[\boldsymbol{b} + \boldsymbol{\partial}\times(\boldsymbol{\partial}\times\boldsymbol{b})\big] - 2\boldsymbol{h}\cdot\boldsymbol{b}\Big\} \quad . \tag{11.198}$$

Here we have

$$\begin{aligned}\boldsymbol{b} - \boldsymbol{\partial}^2\boldsymbol{b} &= -\frac{2\pi}{\kappa}\sum_{i=1}^{N_{\rm V}} n_i\,\delta(\boldsymbol{\rho}-\boldsymbol{\rho}_i)\\ \boldsymbol{b} &= -\frac{1}{\kappa}\sum_{i=1}^{N_{\rm V}} n_i\,K_0\big(|\boldsymbol{\rho}-\boldsymbol{\rho}_i|\big) \quad .\end{aligned} \tag{11.199}$$

Thus, again replacing $K_0(0)$ by $K_0(\kappa^{-1})$ and invoking Eqn. 11.69 for $\kappa \gg 1$,

$$\frac{G_{\rm VL} - G_0}{L} = \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{\kappa^2}\left\{\tfrac{1}{2}\ln\big(2e^{-C}\kappa\big)\sum_{i=1}^{N_{\rm V}} n_i^2 + \sum_{i<j}^{N_{\rm V}} n_i\,n_j\,K_0\big(|\boldsymbol{\rho}_i-\boldsymbol{\rho}_j|\big) + \kappa h\sum_{i=1}^{N_{\rm V}} n_i\right\} \quad . \tag{11.200}$$

The first term on the RHS is the self-interaction, cut off at a length scale κ^{-1} (ξ in physical units). The second term is the interaction between different vortex lines. We've assumed a perfectly straight set of vortex lines – no wiggling! The third term arises from $\boldsymbol{B}\cdot\boldsymbol{H}$ in the Gibbs free energy. If we assume a finite density of vortex lines, we may calculate the magnetization. For $H - H_{c1} \ll H_{c1}$, the spacing between the vortices is huge, and since $K_0(r) \simeq (\pi/2r)^{1/2}\exp(-r)$ for large $|r|$, we may safely neglect all but nearest neighbor interaction terms. We assume $n_i = -1$ for all i. Let the vortex lines form a regular lattice of coordination number z and nearest neighbor separation a. Then

$$\frac{G_{\rm VL} - G_0}{L} = \frac{N_{\rm V} H_{\rm c}^2\,\lambda_{\rm L}^2}{\kappa^2}\left\{\tfrac{1}{2}\ln\big(2e^{-C}\kappa\big) + \tfrac{1}{2}zK_0(a) - \kappa h\right\} \quad , \tag{11.201}$$

where N is the total number of vortex lines, given by $N = A/\Omega$ for a lattice with unit cell area Ω. Assuming a triangular lattice, $\Omega = \frac{\sqrt{3}}{2}\,a^2$ and $z = 6$. Then, dividing by the cross-sectional area A, we have that the difference in free energy densities is ($\kappa \gg 1$ assumed)

$$\frac{\Delta G}{V} = \frac{H_{\rm c}^2\,\lambda_{\rm L}^2}{\sqrt{3}\,\kappa^2}\Big\{6K_0(a)\,a^{-2} - 2\kappa\,(h - h_{c1})\,a^{-2}\Big\} \quad , \tag{11.202}$$

where $h_{c1} = \ln\big(2e^{-C}\kappa\big)/2\kappa$. To find the optimal lattice spacing a, differentiate with respect to a, which yields the condition

$$K_0(a) + \tfrac{1}{2}\,K_1(a) = 4\kappa\,(h - h_{c1}) \quad . \tag{11.203}$$

The LHS is a monotonically decreasing function of a, behaving as $1/a$ as $a \to 0$ and decreasing as $(\pi/2a)^{1/2}\, e^{-a}$ as $a \to \infty$, hence there is a unique solution for $a(h)$ provided $h > h_{\rm c1}$. This analysis is valid provided $\kappa \gg 1$ and also $\kappa(h - h_{\rm c1}) \ll 1$. The latter requirement pertains because we have not done a proper accounting of the vortex core, hence we must assume a is much larger than κ^{-1}, which is the coherence length ξ in units of $\lambda_{\rm L}$.

Chapter 12

BCS Theory of Superconductivity

12.1 Binding and Dimensionality

Consider a spherically symmetric potential $U(\boldsymbol{r}) = -U_0\,\Theta(a-r)$. Are there bound states, *i.e.* states in the eigenspectrum of negative energy? What role does dimension play? It is easy to see that if $U_0 > 0$ is large enough, there are always bound states. A trial state completely localized within the well has kinetic energy $T_0 \simeq \hbar^2/ma^2$, while the potential energy is $-U_0$, so if $U_0 > \hbar^2/ma^2$, we have a variational state with energy $E = T_0 - U_0 < 0$, which is of course an upper bound on the true ground state energy.

What happens, though, if $U_0 < T_0$? We again appeal to a variational argument. Consider a Gaussian or exponentially localized wavefunction with characteristic size $\xi \equiv \lambda a$, with $\lambda > 1$. The variational energy is then

$$E \simeq \frac{\hbar^2}{m\xi^2} - U_0\left(\frac{a}{\xi}\right)^d = T_0\,\lambda^{-2} - U_0\,\lambda^{-d} \quad . \tag{12.1}$$

Extremizing with respect to λ, we obtain $-2T_0\,\lambda^{-3} + d\,U_0\,\lambda^{-(d+1)}$ and $\lambda = \left(d\,U_0/2T_0\right)^{1/(d-2)}$. Inserting this into our expression for the energy, we find

$$E = \left(\frac{2}{d}\right)^{2/(d-2)}\left(1-\frac{2}{d}\right)T_0^{d/(d-2)}\,U_0^{-2/(d-2)} \quad . \tag{12.2}$$

We see that for $d = 1$ we have $\lambda = 2T_0/U_0$ and $E = -U_0^2/4T_0 < 0$. In $d = 2$ dimensions, we have $E = (T_0 - U_0)/\lambda^2$, which says $E \geqslant 0$ unless $U_0 > T_0$. For weak attractive $U(\boldsymbol{r})$, the minimum energy solution is $E \to 0^+$, with $\lambda \to \infty$. It turns out that $d = 2$ is a marginal dimension, and we shall show that we always get localized states with a ballistic dispersion and an attractive potential well. For $d > 2$ we have $E > 0$ which suggests that we cannot have bound states unless $U_0 > T_0$, in which case $\lambda \leqslant 1$ and we must appeal to the analysis in the previous paragraph.

We can firm up this analysis a bit by considering the Schrödinger equation,

$$-\frac{\hbar^2}{2m}\,\nabla^2\psi(\boldsymbol{x}) + V(\boldsymbol{x})\,\psi(\boldsymbol{x}) = E\,\psi(\boldsymbol{x}) \quad . \tag{12.3}$$

Fourier transforming, we have

$$\varepsilon(\boldsymbol{k})\,\hat\psi(\boldsymbol{k}) + \int \frac{d^dk'}{(2\pi)^d}\,\hat V(\boldsymbol{k}-\boldsymbol{k}')\,\hat\psi(\boldsymbol{k}') = E\,\hat\psi(\boldsymbol{k}) \quad , \tag{12.4}$$

where $\varepsilon(\boldsymbol{k}) = \hbar^2\boldsymbol{k}^2/2m$. We may now write $\hat V(\boldsymbol{k}-\boldsymbol{k}') = \sum_n \lambda_n\,\alpha_n(\boldsymbol{k})\,\alpha_n^*(\boldsymbol{k}')$, which is a decomposition of the Hermitian matrix $\hat V_{\boldsymbol{k},\boldsymbol{k}'} \equiv \hat V(\boldsymbol{k}-\boldsymbol{k}')$ into its (real) eigenvalues λ_n and eigenvectors $\alpha_n(\boldsymbol{k})$. Let's approximate $V_{\boldsymbol{k},\boldsymbol{k}'}$ by its leading eigenvalue, which we call λ, and the corresponding eigenvector $\alpha(\boldsymbol{k})$. That is, we write $\hat V_{\boldsymbol{k},\boldsymbol{k}'} \simeq \lambda\,\alpha(\boldsymbol{k})\,\alpha^*(\boldsymbol{k}')$. We then have

$$\hat\psi(\boldsymbol{k}) = \frac{\lambda\,\alpha(\boldsymbol{k})}{E-\varepsilon(\boldsymbol{k})}\int\frac{d^dk'}{(2\pi)^d}\,\alpha^*(\boldsymbol{k}')\,\hat\psi(\boldsymbol{k}') \quad . \tag{12.5}$$

Multiply the above equation by $\alpha^*(\boldsymbol{k})$ and integrate over $\boldsymbol{k}$, resulting in

$$\frac{1}{\lambda} = \int\frac{d^dk}{(2\pi)^d}\,\frac{\big|\alpha(\boldsymbol{k})\big|^2}{E-\varepsilon(\boldsymbol{k})} = \frac{1}{\lambda} = \int\limits_0^\infty d\varepsilon\,\frac{g(\varepsilon)}{E-\varepsilon}\,\big|\alpha(\varepsilon)\big|^2 \quad , \tag{12.6}$$

where $g(\varepsilon)$ is the density of states $g(\varepsilon) = \mathrm{Tr}\,\delta\big(\varepsilon-\varepsilon(\boldsymbol{k})\big)$. Here, we assume that $\alpha(\boldsymbol{k}) = \alpha(k)$ is isotropic. It is generally the case that if $V_{\boldsymbol{k},\boldsymbol{k}'}$ is isotropic, *i.e.* if it is invariant under a simultaneous $\mathsf{O}(3)$ rotation $\boldsymbol{k}\to R\boldsymbol{k}$ and $\boldsymbol{k}'\to R\boldsymbol{k}'$, then so will be its lowest eigenvector. Furthermore, since $\varepsilon = \hbar^2k^2/2m$ is a function of the scalar $k = |\boldsymbol{k}|$, this means $\alpha(k)$ can be considered a function of ε. We then have

$$\frac{1}{|\lambda|} = \int\limits_0^\infty d\varepsilon\,\frac{g(\varepsilon)}{|E|+\varepsilon}\,\big|\alpha(\varepsilon)\big|^2 \quad , \tag{12.7}$$

where we have we assumed an attractive potential ($\lambda<0$), and, as we are looking for a bound state, $E<0$.

If $\alpha(0)$ and $g(0)$ are finite, then in the limit $|E|\to 0$ we have

$$\frac{1}{|\lambda|} = g(0)\,|\alpha(0)|^2\,\ln\big(1/|E|\big) + \text{finite}\,. \tag{12.8}$$

This equation may be solved for arbitrarily small $|\lambda|$ because the RHS of Eqn. 12.7 diverges as $|E|\to 0$. If, on the other hand, $g(\varepsilon)\sim\varepsilon^p$ where $p>0$, then the RHS is finite even when $E=0$. In this case, bound states can only exist for $|\lambda|>\lambda_{\rm c}$, where

$$\lambda_{\rm c} = 1\Big/\int\limits_0^\infty d\varepsilon\,\frac{g(\varepsilon)}{\varepsilon}\,\big|\alpha(\varepsilon)\big|^2 \quad . \tag{12.9}$$

Typically the integral has a finite upper limit, given by the bandwidth B. For the ballistic dispersion, one has $g(\varepsilon) \propto \varepsilon^{(d-2)/2}$, so $d = 2$ is the marginal dimension. In dimensions $d \leqslant 2$, bound states form for arbitrarily weak attractive potentials.

12.2 Cooper's Problem

In 1956, Leon Cooper considered the problem of two electrons interacting in the presence of a quiescent Fermi sea. The background electrons comprising the Fermi sea enter the problem only through their *Pauli blocking*. Since spin and total momentum are conserved, Cooper first considered a zero momentum singlet,

$$|\,\Psi\,\rangle = \frac{1}{\sqrt{2}} \sum_{\boldsymbol{k}} A_{\boldsymbol{k}} \left(c^\dagger_{\boldsymbol{k}\uparrow} c^\dagger_{-\boldsymbol{k}\downarrow} - c^\dagger_{\boldsymbol{k}\downarrow} c^\dagger_{-\boldsymbol{k}\uparrow} \right) |\,\mathrm{F}\,\rangle \quad , \tag{12.10}$$

where $|\,\mathrm{F}\,\rangle$ is the filled Fermi sea, $|\,\mathrm{F}\,\rangle = \prod_{|\boldsymbol{p}|<k_{\mathrm{F}}} c^\dagger_{\boldsymbol{p}\uparrow} c^\dagger_{\boldsymbol{p}\downarrow} \,|\,0\,\rangle$. Only states with $k > k_{\mathrm{F}}$ contribute to the RHS of Eqn. 12.10, due to Pauli blocking. The real space wavefunction is

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{1}{\sqrt{2}} \sum_{\boldsymbol{k}} A_{\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot(\boldsymbol{r}_1 - \boldsymbol{r}_2)} \left(|\uparrow_1 \downarrow_2\rangle - |\downarrow_1 \uparrow_2\rangle \right) \quad , \tag{12.11}$$

with $A_{\boldsymbol{k}} = A_{-\boldsymbol{k}}$ to enforce symmetry of the orbital part. It should be emphasized that this is a two-particle wavefunction, and not an $(N+2)$-particle wavefunction, with N the number of electrons in the Fermi sea. Again, the Fermi sea in this analysis has no dynamics of its own. Its presence is reflected only in the restriction $k > k_{\mathrm{F}}$ for the states which participate in the Cooper pair.

The many-body Hamiltonian is written

$$\hat{H} = \sum_{\boldsymbol{k}\sigma} \varepsilon_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\sigma} c_{\boldsymbol{k}\sigma} + \tfrac{1}{2} \sum_{\boldsymbol{k}_1\sigma_1} \sum_{\boldsymbol{k}_2\sigma_2} \sum_{\boldsymbol{k}_3\sigma_3} \sum_{\boldsymbol{k}_4\sigma_4} \langle\, \boldsymbol{k}_1\sigma_1, \boldsymbol{k}_2\sigma_2 \,|\, v \,|\, \boldsymbol{k}_3\sigma_3, \boldsymbol{k}_4\sigma_4 \,\rangle\, c^\dagger_{\boldsymbol{k}_1\sigma_1} c^\dagger_{\boldsymbol{k}_2\sigma_2} c_{\boldsymbol{k}_4\sigma_4} c_{\boldsymbol{k}_3\sigma_3} \,. \tag{12.12}$$

We treat $|\,\Psi\,\rangle$ as a variational state, which means we set

$$\frac{\delta}{\delta A^*_{\boldsymbol{k}}} \frac{\langle\,\Psi\,|\,\hat{H}\,|\,\Psi\,\rangle}{\langle\,\Psi\,|\,\Psi\,\rangle} = 0 \quad , \tag{12.13}$$

resulting in

$$(E - E_0)\, A_{\boldsymbol{k}} = 2\varepsilon_{\boldsymbol{k}}\, A_{\boldsymbol{k}} + \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, A_{\boldsymbol{k}'} \quad , \tag{12.14}$$

where

$$V_{\boldsymbol{k},\boldsymbol{k}'} = \langle\, \boldsymbol{k}\uparrow, -\boldsymbol{k}\downarrow \,|\, v \,|\, \boldsymbol{k}'\uparrow, -\boldsymbol{k}'\downarrow \,\rangle = \frac{1}{V} \int d^3r\; v(\boldsymbol{r})\, e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{r}} \quad . \tag{12.15}$$

Here $E_0 = \langle\, \mathrm{F} \,|\, \hat{H} \,|\, \mathrm{F} \,\rangle$ is the energy of the Fermi sea.

We write $\varepsilon_{\boldsymbol{k}} = \varepsilon_{\rm F} + \xi_{\boldsymbol{k}}$, and we define $E \equiv E_0 + 2\varepsilon_{\rm F} + W$. Then

$$W A_{\boldsymbol{k}} = 2\xi_{\boldsymbol{k}}\, A_{\boldsymbol{k}} + \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, A_{\boldsymbol{k}'} \quad . \tag{12.16}$$

If $V_{\boldsymbol{k},\boldsymbol{k}'}$ is rotationally invariant, meaning it is left unchanged by $\boldsymbol{k} \to R\boldsymbol{k}$ and $\boldsymbol{k}' \to R\boldsymbol{k}'$ where $R \in \mathsf{O}(3)$, then we may write

$$V_{\boldsymbol{k},\boldsymbol{k}'} = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} V_\ell(k,k')\, Y_{\ell,m}(\hat{\boldsymbol{k}})\, Y^*_{\ell,m}(\hat{\boldsymbol{k}}') \quad . \tag{12.17}$$

We assume that $V_l(k,k')$ is *separable*, meaning we can write

$$V_\ell(k,k') = \frac{1}{V}\,\lambda_\ell\,\alpha_\ell(k)\,\alpha^*_\ell(k') \quad . \tag{12.18}$$

This simplifies matters and affords us an exact solution, for now we take $A_{\boldsymbol{k}} = A_k\, Y_{\ell,m}(\hat{\boldsymbol{k}})$ to obtain a solution in the ℓ angular momentum channel:

$$W_\ell\, A_k = 2\xi_k\, A_k + \lambda_\ell\,\alpha_\ell(k) \cdot \frac{1}{V} \sum_{\boldsymbol{k}'} \alpha^*_\ell(k')\, A_{k'} \quad , \tag{12.19}$$

which may be recast as

$$A_k = \frac{\lambda_\ell\,\alpha_\ell(k)}{W_\ell - 2\xi_k} \cdot \frac{1}{V} \sum_{\boldsymbol{k}'} \alpha^*_\ell(k')\, A_{k'} \quad . \tag{12.20}$$

Now multiply by α^*_k and sum over k to obtain

$$\frac{1}{\lambda_\ell} = \frac{1}{V} \sum_{\boldsymbol{k}} \frac{\big|\alpha_\ell(k)\big|^2}{W_\ell - 2\xi_k} \equiv \Phi(W_\ell) \quad . \tag{12.21}$$

We solve this for W_ℓ.

We may find a graphical solution. Recall that the sum is restricted to $\boldsymbol{k}$ such that $k > k_{\rm F}$, and that $\xi_k \geqslant 0$. The denominator on the RHS of Eqn. 12.21 changes sign as a function of W_ℓ every time $\frac{1}{2}W_\ell$ passes through one of the ξ_k values[1]. A sketch of the graphical solution is provided in Fig. 12.1. One sees that if $\lambda_\ell < 0$, *i.e.* if the potential is attractive, then a bound state exists. This is true for arbitrarily weak $|\lambda_\ell|$, a situation not usually encountered in three-dimensional problems, where there is usually a critical strength of the attractive potential in order to form a bound state[2]. This is a density of states effect – by restricting our attention to electrons near the

[1] We imagine quantizing in a finite volume, so the allowed k values are discrete.

[2] For example, the He_2 molecule is unbound, despite the attractive $-1/r^6$ van der Waals attractive tail in the interatomic potential.

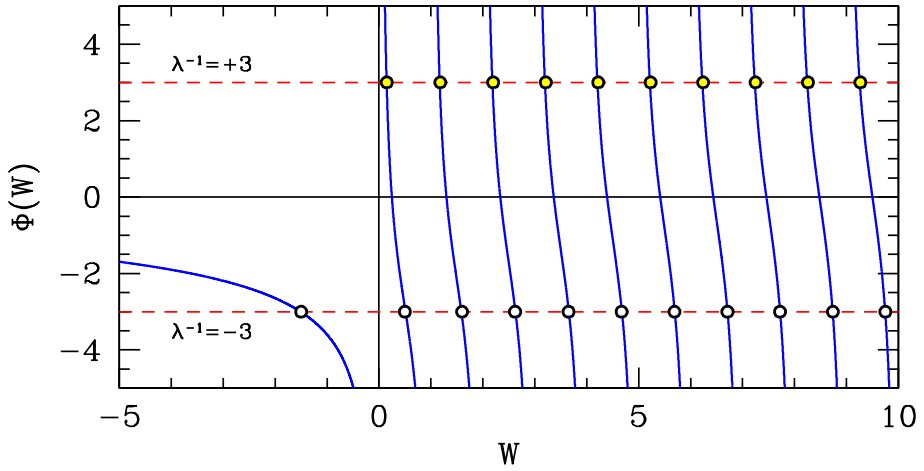


Figure 12.1: Graphical solution to the Cooper problem. A bound state exists for arbitrarily weak $\lambda < 0$.

Fermi level, where the DOS is roughly constant at $g(\varepsilon_{\rm F}) = m^* k_{\rm F}/\pi^2\hbar^2$, rather than near $\boldsymbol{k} = 0$, where $g(\varepsilon)$ vanishes as $\sqrt{\varepsilon}$, the pairing problem is effectively rendered two-dimensional. We can make further progress by assuming a particular form for $\alpha_\ell(k)$:

$$\alpha_\ell(k) = \begin{cases} 1 & \text{if } 0 < \xi_k < B_\ell \\ 0 & \text{otherwise} \end{cases} \quad , \tag{12.22}$$

where B_ℓ is an effective bandwidth for the ℓ channel. Then

$$1 = \tfrac{1}{2}\,|\lambda_\ell| \int\limits_0^{B_\ell} d\xi\, \frac{g(\varepsilon_{\rm F}+\xi)}{|W_\ell| + 2\xi} \quad . \tag{12.23}$$

The factor of $\frac{1}{2}$ is because it is the DOS per spin here, and not the total DOS. We assume $g(\varepsilon)$ does not vary significantly in the vicinity of $\varepsilon = \varepsilon_{\rm F}$, and pull $g(\varepsilon_{\rm F})$ out from the integrand. Integrating and solving for $|W_\ell|$,

$$|W_\ell| = \frac{2B_\ell}{\exp\big[4/|\lambda_\ell|\, g(\varepsilon_{\rm F})\big] - 1} \quad . \tag{12.24}$$

In the *weak coupling* limit, where $|\lambda_\ell|\, g(\varepsilon_{\rm F}) \ll 1$, we have

$$|W_\ell| \simeq 2B_\ell\, \exp\left(-\frac{4}{|\lambda_\ell|\, g(\varepsilon_{\rm F})}\right) \quad . \tag{12.25}$$

As we shall see when we study BCS theory, the factor in the exponent is twice too large. The coefficient $2B_\ell$ will be shown to be the Debye energy of the phonons; we will see that it is only over a narrow range of energies about the Fermi surface that the effective electron-electron interaction is attractive. For strong coupling,

$$|W_\ell| = \tfrac{1}{2}\,|\lambda_\ell|\, B_\ell\, g(\varepsilon_{\rm F}) \quad . \tag{12.26}$$

Finite momentum Cooper pair

We can construct a finite momentum Cooper pair as follows:

$$|\,\Psi_{\boldsymbol{q}}\,\rangle = \frac{1}{\sqrt{2}}\sum_{\boldsymbol{k}} A_{\boldsymbol{k}}\left(c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow} - c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow}c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\right)|\,\mathrm{F}\,\rangle \quad . \tag{12.27}$$

This wavefunction is a momentum eigenstate, with total momentum $\boldsymbol{P} = \hbar\boldsymbol{q}$. The eigenvalue equation is then

$$W A_{\boldsymbol{k}} = \left(\xi_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} + \xi_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}}\right)A_{\boldsymbol{k}} + \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,A_{\boldsymbol{k}'} \quad . \tag{12.28}$$

Assuming $\xi_{\boldsymbol{k}} = \xi_{-\boldsymbol{k}}$,

$$\xi_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} + \xi_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} = 2\,\xi_{\boldsymbol{k}} + \tfrac{1}{4}\,q^\alpha q^\beta\,\frac{\partial^2\xi_{\boldsymbol{k}}}{\partial k^\alpha\,\partial k^\beta} + \ldots \quad . \tag{12.29}$$

The binding energy is thus reduced by an amount proportional to q^2 ; the $\boldsymbol{q} = 0$ Cooper pair has the greatest binding energy[3].

Mean square radius of the Cooper pair

We have

$$\left\langle \boldsymbol{r}^2\right\rangle = \frac{\int d^3r\,\left|\Psi(\boldsymbol{r})\right|^2\boldsymbol{r}^2}{\int d^3r\,\left|\Psi(\boldsymbol{r})\right|^2} = \frac{\int d^3k\,\left|\boldsymbol{\nabla}_{\boldsymbol{k}}A_{\boldsymbol{k}}\right|^2}{\int d^3k\,\left|A_{\boldsymbol{k}}\right|^2} \simeq \frac{g(\varepsilon_{\rm F})\,\xi'(k_{\rm F})^2\int\limits_0^\infty d\xi\,\left|\frac{\partial A}{\partial\xi}\right|^2}{g(\varepsilon_{\rm F})\int\limits_0^\infty d\xi\,|A|^2} \tag{12.30}$$

with $A(\xi) = -C/\big(|W|+2\xi\big)$ and thus $A'(\xi) = 2C/\big(|W|+2\xi\big)^2$, where C is a constant independent of ξ. Ignoring the upper cutoff on ξ at B_ℓ, we have

$$\left\langle \boldsymbol{r}^2\right\rangle = 4\,\xi'(k_{\rm F})^2\cdot\frac{\int\limits_{|W|}^\infty du\,u^{-4}}{\int\limits_{|W|}^\infty du\,u^{-2}} = \tfrac{4}{3}\,(\hbar v_{\rm F})^2\,|W|^{-2} \quad , \tag{12.31}$$

where we have used $\xi'(k_{\rm F}) = \hbar v_{\rm F}$. Thus, $R_{\rm RMS} = 2\hbar v_{\rm F}/\sqrt{3}\,|W|$. In the weak coupling limit, where $|W|$ is exponentially small in $1/|\lambda|$, the Cooper pair radius is huge. Indeed it is so large that many other Cooper pairs have their centers of mass within the radius of any given pair. This feature is what makes the BCS mean field theory of superconductivity so successful. Recall in our discussion of the Ginzburg criterion in §11.4.5, we found that mean field theory was qualitatively correct down to the Ginzburg reduced temperature $t_{\rm G} = (\mathsf{a}/R_*)^{2d/(4-d)}$, *i.e.* $t_{\rm G} = (\mathsf{a}/R_*)^6$ for $d = 3$. In this expression, R_* should be the mean Cooper pair size, and a a microscopic length (*i.e.* lattice constant). Typically $R_*/\mathsf{a} \sim 10^2 - 10^3$, so $t_{\rm G}$ is very tiny indeed.

[3]We assume the matrix $\partial_\alpha\partial_\beta\,\xi_{\boldsymbol{k}}$ is positive definite.

12.3 Effective Attraction Due to Phonons

The solution to Cooper's problem provided the first glimpses into the pairing nature of the superconducting state. But why should $V_{\boldsymbol{k},\boldsymbol{k}'}$ be attractive? One possible mechanism is an *induced* attraction due to phonons.

12.3.1 Electron-phonon Hamiltonian

In §5.12 we derived the electron-phonon Hamiltonian,

$$\hat{H}_{\rm el-ph} = \frac{1}{\sqrt{V}} \sum_{\substack{\boldsymbol{k},\boldsymbol{k}'\sigma \\ \boldsymbol{q},\lambda,\boldsymbol{G}}} g_\lambda(\boldsymbol{k},\boldsymbol{k}')\,(a^\dagger_{\boldsymbol{q}\lambda} + a_{-\boldsymbol{q}\lambda})\,c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}'\sigma}\,\delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}+\boldsymbol{G}} \quad , \tag{12.32}$$

where $c^\dagger_{\boldsymbol{k}\sigma}$ creates an electron in state $|\,\boldsymbol{k}\,\sigma\,\rangle$ and $a^\dagger_{\boldsymbol{q}\lambda}$ creates a phonon in state $|\,\boldsymbol{q}\,\lambda\,\rangle$, where λ is the phonon polarization state. $\boldsymbol{G}$ is a reciprocal lattice vector, and

$$g_\lambda(\boldsymbol{k},\boldsymbol{k}') = -i\left(\frac{\hbar}{2\Omega\,\omega_\lambda(\boldsymbol{q})}\right)^{1/2} \frac{4\pi Z e^2}{(\boldsymbol{q}+\boldsymbol{G})^2 + \lambda_{\rm TF}^{-2}}\,(\boldsymbol{q}+\boldsymbol{G})\cdot\hat{\mathbf{e}}^*_\lambda(\boldsymbol{q})\;. \tag{12.33}$$

is the electron-phonon coupling constant, with $\hat{\mathbf{e}}_\lambda(\boldsymbol{q})$ the phonon polarization vector, Ω the Wigner-Seitz unit cell volume, and $\omega_\lambda(\boldsymbol{q})$ the phonon frequency dispersion of the λ branch.

Recall that in an isotropic 'jellium' solid, the phonon polarization at wavevector $\boldsymbol{q}$ either is parallel to $\boldsymbol{q}$ (longitudinal waves), or perpendicular to $\boldsymbol{q}$ (transverse waves). We then have that only longitudinal waves couple to the electrons. This is because transverse waves do not result in any local accumulation of charge density, and the Coulomb interaction couples electrons to density fluctuations. Restricting our attention to the longitudinal phonon, we found for small $\boldsymbol{q}$ the electron-longitudinal phonon coupling $g_{\rm L}(\boldsymbol{k},\boldsymbol{k}+\boldsymbol{q}) \equiv g_{\boldsymbol{q}}$ satisfies

$$|g_{\boldsymbol{q}}|^2 = \lambda_{\rm el-ph}\cdot\frac{\hbar c_{\rm L} q}{g(\varepsilon_{\rm F})}\;, \tag{12.34}$$

where $g(\varepsilon_{\rm F})$ is the electronic density of states, $c_{\rm L}$ is the longitudinal phonon speed, and where the dimensionless *electron-phonon coupling constant* is

$$\lambda_{\rm el-ph} = \frac{Z^2}{2Mc_{\rm L}^2\,\Omega\,g(\varepsilon_{\rm F})} = \frac{2Z}{3}\,\frac{m^*}{M}\left(\frac{\varepsilon_{\rm F}}{k_{\rm B}\Theta_{\rm s}}\right)^2\;, \tag{12.35}$$

with $\Theta_{\rm s} \equiv \hbar c_{\rm L} k_{\rm F}/k_{\rm B}$.

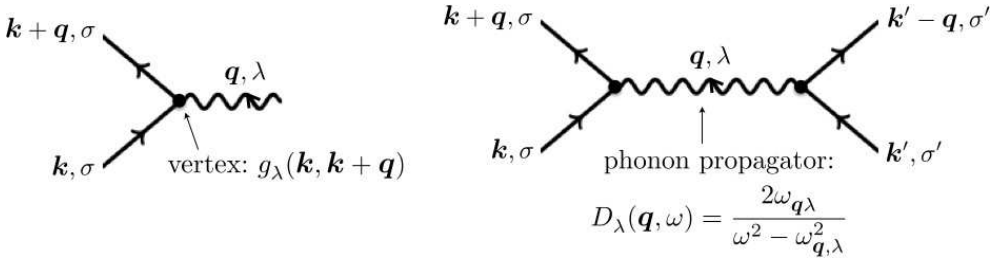


Figure 12.2: Feynman diagrams for electron-phonon processes.

12.3.2 Effective interaction between electrons

Consider now the problem of two particle scattering $|\,\boldsymbol{k}\,\sigma\,,\,-\boldsymbol{k}\;-\sigma\,\rangle\;\rightarrow\;|\,\boldsymbol{k}'\,\sigma\,,\,-\boldsymbol{k}'\;-\sigma\,\rangle$. We assume no phonons are present in the initial state, *i.e.* we work at $T=0$. The initial state energy is $E_{\rm i}=2\xi_{\boldsymbol{k}}$ and the final state energy is $E_{\rm f}=2\xi_{\boldsymbol{k}'}$. There are two intermediate states:[4]

$$\begin{aligned}|\,{\rm I}_1\,\rangle &= |\,\boldsymbol{k}'\,\sigma\,,\,-\boldsymbol{k}\;-\sigma\,\rangle\otimes|\,-\boldsymbol{q}\,\lambda\,\rangle\\ |\,{\rm I}_2\,\rangle &= |\,\boldsymbol{k}\,\sigma\,,\,-\boldsymbol{k}'\;-\sigma\,\rangle\otimes|\,+\boldsymbol{q}\,\lambda\,\rangle\quad,\end{aligned}\tag{12.36}$$

with $\boldsymbol{k}'=\boldsymbol{k}+\boldsymbol{q}$ in each case. The energies of these intermediate states are

$$E_1=\xi_{-\boldsymbol{k}}+\xi_{\boldsymbol{k}'}+\hbar\omega_{-\boldsymbol{q}\,\lambda}\qquad,\qquad E_2=\xi_{\boldsymbol{k}}+\xi_{-\boldsymbol{k}'}+\hbar\omega_{\boldsymbol{q}\,\lambda}\quad.\tag{12.37}$$

The second order matrix element is then

$$\begin{aligned}\langle\,\boldsymbol{k}'\,\sigma\,,\,-\boldsymbol{k}'\;-\sigma\,|\,\hat H_{\rm indirect}\,|\,\boldsymbol{k}\,\sigma\,,\,-\boldsymbol{k}\;-\sigma\,\rangle &= \sum_n\langle\,\boldsymbol{k}'\,\sigma\,,\,-\boldsymbol{k}'\;-\sigma\,|\,\hat H_{\rm el-ph}\,|\,n\,\rangle\langle\,n\,|\,\hat H_{\rm el-ph}\,|\,\boldsymbol{k}\,\sigma\,,\,-\boldsymbol{k}\;-\sigma\,\rangle\\ &\qquad\times\left(\frac{1}{E_{\rm f}-E_n}+\frac{1}{E_{\rm i}-E_n}\right)\\ &= \big|g_{\boldsymbol{k}'-\boldsymbol{k}}\big|^2\left(\frac{1}{\xi_{\boldsymbol{k}'}-\xi_{\boldsymbol{k}}-\hbar\omega_{\boldsymbol{q}}}+\frac{1}{\xi_{\boldsymbol{k}}-\xi_{\boldsymbol{k}'}-\hbar\omega_{\boldsymbol{q}}}\right)\quad.\end{aligned}\tag{12.38}$$

Here we have assumed $\xi_{\boldsymbol{k}}=\xi_{-\boldsymbol{k}}$ and $\omega_{\boldsymbol{q}}=\omega_{-\boldsymbol{q}}$, and we have chosen λ to correspond to the longitudinal acoustic phonon branch. We add this to the Coulomb interaction $\hat v\big(|\boldsymbol{k}-\boldsymbol{k}'|\big)$ to get the net effective interaction between electrons,

$$\langle\,\boldsymbol{k}'\,\sigma\,,\,-\boldsymbol{k}'\;-\sigma\,|\,\hat H_{\rm eff}\,|\,\boldsymbol{k}\,\sigma\,,\,-\boldsymbol{k}\;-\sigma\,\rangle=\hat v\big(|\boldsymbol{k}-\boldsymbol{k}'|\big)+\big|g_{\boldsymbol{q}}\big|^2\times\frac{2\hbar\omega_{\boldsymbol{q}}}{(\xi_{\boldsymbol{k}}-\xi_{\boldsymbol{k}'})^2-(\hbar\omega_{\boldsymbol{q}})^2}\quad,\tag{12.39}$$

where $\boldsymbol{k}'=\boldsymbol{k}+\boldsymbol{q}$. We see that the effective interaction can be attractive, but only so long as $|\xi_{\boldsymbol{k}}-\xi_{\boldsymbol{k}'}|<\hbar\omega_{\boldsymbol{q}}$.

[4]The annihilation operator in the Hamiltonian $\hat H_{\rm el-ph}$ can act on either of the two electrons.

Another way to evoke this effective attraction is via the jellium model studied in §10.5.6. There we found the effective interaction between unit charges was given by

$$\hat{V}_{\rm eff}(\boldsymbol{q},\omega) = \frac{4\pi e^2}{\boldsymbol{q}^2\,\epsilon(\boldsymbol{q},\omega)} \tag{12.40}$$

where

$$\frac{1}{\epsilon(\boldsymbol{q},\omega)} \simeq \frac{q^2}{q^2+q_{\rm TF}^2}\left\{1+\frac{\omega_{\boldsymbol{q}}^2}{\omega^2-\omega_{\boldsymbol{q}}^2}\right\} \quad , \tag{12.41}$$

where the first term in the curly brackets is due to Thomas-Fermi screening (§10.5.2) and the second from ionic screening (§10.5.6). Recall that the Thomas-Fermi wavevector is given by $q_{\rm TF} = \sqrt{4\pi e^2 g(\varepsilon_{\rm F})}$, where $g(\varepsilon_{\rm F})$ is the electronic density of states at the Fermi level, and that $\omega_{\boldsymbol{q}} = \Omega_{\rm p,i}\,q\big/\sqrt{q^2+q_{\rm TF}^2}$, where $\Omega_{\rm p,i} = \sqrt{4\pi n_{\rm i}^0 Z_{\rm i} e^2/M_{\rm i}}$ is the ionic plasma frequency.

12.4 Reduced BCS Hamiltonian

The operator which creates a Cooper pair with total momentum $\boldsymbol{q}$ is $b^\dagger_{\boldsymbol{k},\boldsymbol{q}} + b^\dagger_{-\boldsymbol{k},\boldsymbol{q}}$, where

$$b^\dagger_{\boldsymbol{k},\boldsymbol{q}} = c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\,c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow} \tag{12.42}$$

is a composite operator which creates the state $|\,\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow,\,-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow\rangle$. We learned from the solution to the Cooper problem that the $\boldsymbol{q}=0$ pairs have the greatest binding energy. This motivates consideration of the so-called *reduced BCS Hamiltonian*,

$$\hat{H}_{\rm red} = \sum_{\boldsymbol{k},\sigma}\varepsilon_{\boldsymbol{k}}\,c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k},\boldsymbol{k}'}V_{\boldsymbol{k},\boldsymbol{k}'}\,b^\dagger_{\boldsymbol{k},0}\,b_{\boldsymbol{k}',0} \quad . \tag{12.43}$$

The most general form for a momentum-conserving interaction is[5]

$$\hat{V} = \frac{1}{2V}\sum_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}}\sum_{\sigma,\sigma'}\hat{u}_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})\,c^\dagger_{\boldsymbol{k}+\boldsymbol{q}\,\sigma}\,c^\dagger_{\boldsymbol{p}-\boldsymbol{q}\,\sigma'}\,c_{\boldsymbol{p}\,\sigma'}\,c_{\boldsymbol{k}\,\sigma} \quad . \tag{12.44}$$

Taking $\boldsymbol{p}=-\boldsymbol{k}$, $\sigma'=-\sigma$, and defining $\boldsymbol{k}'\equiv\boldsymbol{k}+\boldsymbol{q}$, we have

$$\hat{V} \to \frac{1}{2V}\sum_{\boldsymbol{k},\boldsymbol{k}',\sigma}\hat{v}(\boldsymbol{k},\boldsymbol{k}')\,c^\dagger_{\boldsymbol{k}'\sigma}\,c^\dagger_{-\boldsymbol{k}'\,-\sigma}\,c_{-\boldsymbol{k}\,-\sigma}\,c_{\boldsymbol{k}\sigma} \quad , \tag{12.45}$$

where $\hat{v}(\boldsymbol{k},\boldsymbol{k}') = \hat{u}_{\uparrow\downarrow}(\boldsymbol{k},-\boldsymbol{k},\boldsymbol{k}'-\boldsymbol{k})$, which is equivalent to $\hat{H}_{\rm red}$.

[5]See the discussion in Appendix I, §12.13.



Figure 12.3: John Bardeen, Leon Cooper, and J. Robert Schrieffer.

If $V_{\boldsymbol{k},\boldsymbol{k}'}$ is attractive, then the ground state will have no pair $(\boldsymbol{k}\uparrow, -\boldsymbol{k}\downarrow)$ occupied by a single electron; the pair states are either empty or doubly occupied. In that case, the reduced BCS Hamiltonian may be written as[6]

$$H^0_{\rm red} = \sum_{\boldsymbol{k}} 2\varepsilon_{\boldsymbol{k}}\, b^\dagger_{\boldsymbol{k},0}\, b_{\boldsymbol{k},0} + \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, b^\dagger_{\boldsymbol{k},0}\, b_{\boldsymbol{k}',0} \quad . \tag{12.46}$$

This has the innocent appearance of a noninteracting bosonic Hamiltonian – an exchange of Cooper pairs restores the many-body wavefunction without a sign change because the Cooper pair is a composite object consisting of an even number of fermions[7]. However, this is not quite correct, because the operators $b_{\boldsymbol{k},0}$ and $b_{\boldsymbol{k}',0}$ do not satisfy canonical bosonic commutation relations. Rather,

$$\begin{aligned} \big[b_{\boldsymbol{k},0}\,,\, b_{\boldsymbol{k}',0}\big] &= \big[b^\dagger_{\boldsymbol{k},0}\,,\, b^\dagger_{\boldsymbol{k}',0}\big] = 0 \\ \big[b_{\boldsymbol{k},0}\,,\, b^\dagger_{\boldsymbol{k}',0}\big] &= \big(1 - c^\dagger_{\boldsymbol{k}\uparrow} c_{\boldsymbol{k}\uparrow} - c^\dagger_{-\boldsymbol{k}\downarrow} c_{-\boldsymbol{k}\downarrow}\big)\,\delta_{\boldsymbol{k}\boldsymbol{k}'} \quad . \end{aligned} \tag{12.47}$$

Because of this, $\hat{H}^0_{\rm red}$ cannot naïvely be diagonalized. The extra terms inside the round brackets on the RHS arise due to the Pauli blocking effects. Indeed, one has $(b^\dagger_{\boldsymbol{k},0})^2 = 0$, so $b^\dagger_{\boldsymbol{k},0}$ is no ordinary boson operator.

While the composite operators $b_{\boldsymbol{k},\boldsymbol{q}}$ do not obey bosonic commutation relations, they are still boson-like in the sense that they can have a nonzero expectation value, which composite fermion operators cannot. This suggest we try a mean field Hartree-Fock approach. Accordingly, we write

$$b_{\boldsymbol{k},0} = \langle b_{\boldsymbol{k},0}\rangle + \overbrace{\big(b_{\boldsymbol{k},0} - \langle b_{\boldsymbol{k},0}\rangle\big)}^{\delta b_{\boldsymbol{k},0}} \quad , \tag{12.48}$$

[6]Spin rotation invariance and a singlet Cooper pair requires that $V_{\boldsymbol{k},\boldsymbol{k}'} = V_{\boldsymbol{k},-\boldsymbol{k}'} = V_{-\boldsymbol{k},\boldsymbol{k}'}$.

[7]Recall that the atom ^{4}He, which consists of six fermions (two protons, two neutrons, and two electrons), is a boson, while ^{3}He, which has only one neutron and thus five fermions, is itself a fermion.

and we neglect terms in $\hat{H}_{\rm red}$ proportional to $\delta b^\dagger_{\boldsymbol{k},0}\,\delta b_{\boldsymbol{k}',0}$. We have

$$\hat{H}_{\rm red} = \sum_{\boldsymbol{k},\sigma} \varepsilon_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'} \Big(\overbrace{-\langle b^\dagger_{\boldsymbol{k},0}\rangle\,\langle b_{\boldsymbol{k}',0}\rangle}^{\text{energy shift}} + \overbrace{\langle b_{\boldsymbol{k}',0}\rangle\, b^\dagger_{\boldsymbol{k},0} + \langle b^\dagger_{\boldsymbol{k},0}\rangle\, b_{\boldsymbol{k}',0}}^{\text{keep this}} + \overbrace{\delta b^\dagger_{\boldsymbol{k},0}\,\delta b_{\boldsymbol{k}',0}}^{\text{drop this!}} \Big) \quad . \tag{12.49}$$

Dropping the last term, which is quadratic in fluctuations, we obtain

$$\hat{H}^{\rm MF}_{\rm red} = \sum_{\boldsymbol{k},\sigma} \varepsilon_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k}} \left(\Delta_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\uparrow}\, c^\dagger_{-\boldsymbol{k}\downarrow} + \Delta^*_{\boldsymbol{k}}\, c_{-\boldsymbol{k}\downarrow}\, c_{\boldsymbol{k}\uparrow} \right) - \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \langle c^\dagger_{\boldsymbol{k}\uparrow}\, c^\dagger_{-\boldsymbol{k}\downarrow}\rangle\, \langle c_{-\boldsymbol{k}'\downarrow}\, c_{\boldsymbol{k}'\uparrow}\rangle \quad , \tag{12.50}$$

where

$$\Delta_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \langle c_{-\boldsymbol{k}'\downarrow}\, c_{\boldsymbol{k}'\uparrow}\rangle \qquad , \qquad \Delta^*_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V^*_{\boldsymbol{k},\boldsymbol{k}'}\, \langle c^\dagger_{\boldsymbol{k}'\uparrow}\, c^\dagger_{-\boldsymbol{k}'\downarrow}\rangle \quad . \tag{12.51}$$

The first thing to notice about $\hat{H}^{\rm MF}_{\rm red}$ is that it does not preserve particle number, *i.e.* it does not commute with $\hat{N} = \sum_{\boldsymbol{k},\sigma} c^\dagger_{\boldsymbol{k}\sigma} c_{\boldsymbol{k}\sigma}$. Accordingly, we are practically forced to work in the grand canonical ensemble, and we define the grand canonical Hamiltonian $\hat{K} \equiv \hat{H} - \mu\hat{N}$.

12.5 Solution of the Mean Field Hamiltonian

We now subtract $\mu\hat{N}$ from Eqn. 12.50, and define $\hat{K}_{\rm BCS} \equiv \hat{H}^{\rm MF}_{\rm red} - \mu\hat{N}$. Thus,

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k}} \begin{pmatrix} c^\dagger_{\boldsymbol{k}\uparrow} & c_{-\boldsymbol{k}\downarrow} \end{pmatrix} \overbrace{\begin{pmatrix} \xi_{\boldsymbol{k}} & \Delta_{\boldsymbol{k}} \\ \Delta^*_{\boldsymbol{k}} & -\xi_{\boldsymbol{k}} \end{pmatrix}}^{K_{\boldsymbol{k}}} \begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^\dagger_{-\boldsymbol{k}\downarrow} \end{pmatrix} + K_0 \quad , \tag{12.52}$$

with $\xi_{\boldsymbol{k}} = \varepsilon_{\boldsymbol{k}} - \mu$, and where

$$K_0 = \sum_{\boldsymbol{k}} \xi_{\boldsymbol{k}} - \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \langle c^\dagger_{\boldsymbol{k}\uparrow} c^\dagger_{-\boldsymbol{k}\downarrow}\rangle\, \langle c_{-\boldsymbol{k}'\downarrow}\, c_{\boldsymbol{k}'\uparrow}\rangle \tag{12.53}$$

is a constant. This problem may be brought to diagonal form via a unitary transformation,

$$\begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^\dagger_{-\boldsymbol{k}\downarrow} \end{pmatrix} = \overbrace{\begin{pmatrix} \cos\vartheta_{\boldsymbol{k}} & -\sin\vartheta_{\boldsymbol{k}}\, e^{i\phi_{\boldsymbol{k}}} \\ \sin\vartheta_{\boldsymbol{k}}\, e^{-i\phi_{\boldsymbol{k}}} & \cos\vartheta_{\boldsymbol{k}} \end{pmatrix}}^{U_{\boldsymbol{k}}} \begin{pmatrix} \gamma_{\boldsymbol{k}\uparrow} \\ \gamma^\dagger_{-\boldsymbol{k}\downarrow} \end{pmatrix} \quad . \tag{12.54}$$

In order for the $\gamma_{\boldsymbol{k}\sigma}$ operators to satisfy fermionic anticommutation relations, the matrix $U_{\boldsymbol{k}}$ must be unitary[8]. We then have

$$\begin{aligned} c_{\boldsymbol{k}\sigma} &= \cos\vartheta_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\sin\vartheta_{\boldsymbol{k}}\,e^{i\phi_{\boldsymbol{k}}}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma} \\ \gamma_{\boldsymbol{k}\sigma} &= \cos\vartheta_{\boldsymbol{k}}\,c_{\boldsymbol{k}\sigma} + \sigma\sin\vartheta_{\boldsymbol{k}}\,e^{i\phi_{\boldsymbol{k}}}\,c^\dagger_{-\boldsymbol{k}\,-\sigma} \quad . \end{aligned} \tag{12.55}$$

EXERCISE: Verify that $\left\{\gamma_{\boldsymbol{k}\sigma}\,,\,\gamma^\dagger_{\boldsymbol{k}'\sigma'}\right\} = \delta_{\boldsymbol{k}\boldsymbol{k}'}\,\delta_{\sigma\sigma'}$.

We now must compute the transformed Hamiltonian. Dropping the $\boldsymbol{k}$ subscript for notational convenience, we have

$$\widetilde{K} = U^\dagger\,K\,U = \begin{pmatrix} \cos\vartheta & \sin\vartheta\,e^{i\phi} \\ -\sin\vartheta\,e^{-i\phi} & \cos\vartheta \end{pmatrix}\begin{pmatrix} \xi & \Delta \\ \Delta^* & -\xi \end{pmatrix}\begin{pmatrix} \cos\vartheta & -\sin\vartheta\,e^{i\phi} \\ \sin\vartheta\,e^{-i\phi} & \cos\vartheta \end{pmatrix} \tag{12.56}$$

$$= \begin{pmatrix} (\cos^2\vartheta - \sin^2\vartheta)\,\xi + \sin\vartheta\cos\vartheta\,(\Delta\,e^{-i\phi} + \Delta^* e^{i\phi}) & \Delta\cos^2\vartheta - \Delta^* e^{2i\phi}\sin^2\vartheta - 2\xi\sin\vartheta\cos\vartheta\,e^{i\phi} \\ \Delta^*\cos^2\vartheta - \Delta e^{-2i\phi}\sin^2\vartheta - 2\xi\sin\vartheta\cos\vartheta\,e^{-i\phi} & (\sin^2\vartheta - \cos^2\vartheta)\,\xi - \sin\vartheta\cos\vartheta\,(\Delta\,e^{-i\phi} + \Delta^* e^{i\phi}) \end{pmatrix} .$$

We now use our freedom to choose ϑ and ϕ to render $\widetilde{K}$ diagonal. That is, we demand that $\phi = \mathsf{arg}(\Delta)$ and

$$2\xi\sin\vartheta\cos\vartheta = |\Delta|\,(\cos^2\vartheta - \sin^2\vartheta) \quad . \tag{12.57}$$

This says $\tan(2\vartheta) = |\Delta|/\xi$, which means

$$\cos(2\vartheta) = \frac{\xi}{E} \qquad , \qquad \sin(2\vartheta) = \frac{|\Delta|}{E} \qquad , \qquad E = \sqrt{\xi^2 + |\Delta|^2} \quad . \tag{12.58}$$

The upper left element of $\widetilde{K}$ then becomes

$$(\cos^2\vartheta - \sin^2\vartheta)\,\xi + \sin\vartheta\cos\vartheta\,(\Delta\,e^{-i\phi} + \Delta^* e^{i\phi}) = \frac{\xi^2}{E} + \frac{|\Delta|^2}{E} = E \quad , \tag{12.59}$$

and thus $\widetilde{K} = \begin{pmatrix} E & 0 \\ 0 & -E \end{pmatrix}$. This unitary transformation, which mixes particle and hole states, is called a *Bogoliubov transformation*, because it was first discovered by Valatin.

Restoring the $\boldsymbol{k}$ subscript, we have $\phi_{\boldsymbol{k}} = \mathsf{arg}(\Delta_{\boldsymbol{k}})$, and $\tan(2\vartheta_{\boldsymbol{k}}) = |\Delta_{\boldsymbol{k}}|/\xi_{\boldsymbol{k}}$, which means

$$\cos(2\vartheta_{\boldsymbol{k}}) = \frac{\xi_{\boldsymbol{k}}}{E_{\boldsymbol{k}}} \qquad , \qquad \sin(2\vartheta_{\boldsymbol{k}}) = \frac{|\Delta_{\boldsymbol{k}}|}{E_{\boldsymbol{k}}} \qquad , \qquad E_{\boldsymbol{k}} = \sqrt{\xi_{\boldsymbol{k}}^2 + |\Delta_{\boldsymbol{k}}|^2} \quad . \tag{12.60}$$

Assuming that $\Delta_{\boldsymbol{k}}$ is not strongly momentum-dependent, we see that the dispersion $E_{\boldsymbol{k}}$ of the excitations has a nonzero minimum at $\xi_{\boldsymbol{k}} = 0$, *i.e.* at $k = k_{\mathsf{F}}$. This minimum value of $E_{\boldsymbol{k}}$ is called the *superconducting energy gap*.

[8]The most general 2×2 unitary matrix is of the above form, but with each row multiplied by an independent phase. These phases may be absorbed into the definitions of the fermion operators themselves. After absorbing these harmless phases, we have written the most general unitary transformation.

We may further write

$$\cos\vartheta_{\boldsymbol{k}} = \sqrt{\frac{E_{\boldsymbol{k}} + \xi_{\boldsymbol{k}}}{2E_{\boldsymbol{k}}}} \quad , \quad \sin\vartheta_{\boldsymbol{k}} = \sqrt{\frac{E_{\boldsymbol{k}} - \xi_{\boldsymbol{k}}}{2E_{\boldsymbol{k}}}} \quad . \tag{12.61}$$

The grand canonical BCS Hamiltonian then becomes

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k},\sigma} E_{\boldsymbol{k}}\,\gamma^\dagger_{\boldsymbol{k}\sigma}\,\gamma_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k}}(\xi_{\boldsymbol{k}} - E_{\boldsymbol{k}}) - \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\langle c^\dagger_{\boldsymbol{k}\uparrow} c^\dagger_{-\boldsymbol{k}\downarrow}\rangle\,\langle c_{-\boldsymbol{k}'\downarrow}\,c_{\boldsymbol{k}'\uparrow}\rangle \quad . \tag{12.62}$$

Finally, what of the ground state wavefunction itself? We must have $\gamma_{\boldsymbol{k}\sigma}|\,{\rm G}\,\rangle = 0$. This leads to

$$|\,{\rm G}\,\rangle = \prod_{\boldsymbol{k}} \left(\,\cos\vartheta_{\boldsymbol{k}} - \sin\vartheta_{\boldsymbol{k}}\,e^{i\phi_{\boldsymbol{k}}}\,c^\dagger_{\boldsymbol{k}\uparrow}\,c^\dagger_{-\boldsymbol{k}\downarrow}\right)|\,0\,\rangle \quad . \tag{12.63}$$

Note that $\langle\,{\rm G}\,|\,{\rm G}\,\rangle = 1$. J. R. Schrieffer conceived of this wavefunction during a subway ride in New York City sometime during the winter of 1957. At the time he was a graduate student at the University of Illinois.

Sanity check

It is good to make contact with something familiar, such as the case $\Delta_{\boldsymbol{k}} = 0$. Note that $\xi_{\boldsymbol{k}} < 0$ for $k < k_{\rm F}$ and $\xi_{\boldsymbol{k}} > 0$ for $k > k_{\rm F}$. We now have

$$\cos\vartheta_{\boldsymbol{k}} = \Theta(k - k_{\rm F}) \quad , \quad \sin\vartheta_{\boldsymbol{k}} = \Theta(k_{\rm F} - k) \quad . \tag{12.64}$$

Note that the wavefunction $|\,{\rm G}\,\rangle$ in Eqn. 12.63 correctly describes a filled Fermi sphere out to $k = k_{\rm F}$. Furthermore, the constant on the RHS of Eqn. 12.62 is $2\sum_{k<k_{\rm F}}\xi_{\boldsymbol{k}}$, which is the Landau free energy of the filled Fermi sphere. What of the excitations? We are free to take $\phi_{\boldsymbol{k}} = 0$. Then

$$\begin{aligned} k < k_{\rm F} \;&:\; \gamma^\dagger_{\boldsymbol{k}\sigma} = \sigma\,c_{-\boldsymbol{k}\,-\sigma} \\ k > k_{\rm F} \;&:\; \gamma^\dagger_{\boldsymbol{k}\sigma} = c^\dagger_{\boldsymbol{k}\sigma} \quad . \end{aligned} \tag{12.65}$$

Thus, the elementary excitations are holes below $k_{\rm F}$ and electrons above $k_{\rm F}$. All we have done, then, is to effect a particle-hole transformation on those states lying within the Fermi sea.

12.6 Self-Consistency

We now demand that the following two conditions hold:

$$N = \sum_{\boldsymbol{k},\sigma}\langle c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}\sigma}\rangle \quad , \quad \Delta_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\langle c_{-\boldsymbol{k}'\downarrow}\,c_{\boldsymbol{k}'\uparrow}\rangle \quad , \tag{12.66}$$

the second of which is from Eqn. 12.51. Thus, we need

$$\begin{aligned}\langle c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}\sigma}\rangle &= \big\langle (\cos\vartheta_{\boldsymbol{k}}\,\gamma^\dagger_{\boldsymbol{k}\sigma} - \sigma\sin\vartheta_{\boldsymbol{k}}\, e^{-i\phi_{\boldsymbol{k}}}\,\gamma_{-\boldsymbol{k}\,-\sigma})(\cos\vartheta_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\sin\vartheta_{\boldsymbol{k}}\, e^{i\phi_{\boldsymbol{k}}}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma})\big\rangle \\ &= \cos^2\vartheta_{\boldsymbol{k}}\, f_{\boldsymbol{k}} + \sin^2\vartheta_{\boldsymbol{k}}\,(1-f_{\boldsymbol{k}}) = \frac{1}{2} - \frac{\xi_{\boldsymbol{k}}}{2E_{\boldsymbol{k}}}\tanh\big(\tfrac{1}{2}\beta E_{\boldsymbol{k}}\big) \quad ,\end{aligned} \tag{12.67}$$

where

$$f_{\boldsymbol{k}} = \langle \gamma^\dagger_{\boldsymbol{k}\sigma}\,\gamma_{\boldsymbol{k}\sigma}\rangle = \frac{1}{e^{\beta E_{\boldsymbol{k}}}+1} = \tfrac{1}{2} - \tfrac{1}{2}\tanh\big(\tfrac{1}{2}\beta E_{\boldsymbol{k}}\big) \tag{12.68}$$

is the Fermi function, with $\beta = 1/k_{\rm B}T$. We also have

$$\begin{aligned}\langle c_{-\boldsymbol{k}\,-\sigma}\, c_{\boldsymbol{k}\sigma}\rangle &= \big\langle (\cos\vartheta_{\boldsymbol{k}}\,\gamma_{-\boldsymbol{k}\,-\sigma} + \sigma\sin\vartheta_{\boldsymbol{k}}\, e^{i\phi_{\boldsymbol{k}}}\,\gamma^\dagger_{\boldsymbol{k}\sigma})(\cos\vartheta_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\sin\vartheta_{\boldsymbol{k}}\, e^{i\phi_{\boldsymbol{k}}}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma})\big\rangle \\ &= \sigma\sin\vartheta_{\boldsymbol{k}}\cos\vartheta_{\boldsymbol{k}}\, e^{i\phi_{\boldsymbol{k}}}\,\big(2f_{\boldsymbol{k}}-1\big) = -\frac{\sigma\Delta_{\boldsymbol{k}}}{2E_{\boldsymbol{k}}}\,\tanh\big(\tfrac{1}{2}\beta E_{\boldsymbol{k}}\big) \quad .\end{aligned} \tag{12.69}$$

Let's evaluate at $T = 0$:

$$N = \sum_{\boldsymbol{k}}\left(1 - \frac{\xi_{\boldsymbol{k}}}{E_{\boldsymbol{k}}}\right) \qquad , \qquad \Delta_{\boldsymbol{k}} = -\sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\frac{\Delta_{\boldsymbol{k}'}}{2E_{\boldsymbol{k}'}} \quad . \tag{12.70}$$

The second of these is known as the BCS gap equation. Note that $\Delta_{\boldsymbol{k}} = 0$ is always a solution of the gap equation. To proceed further, we need a model for $V_{\boldsymbol{k},\boldsymbol{k}'}$. We shall assume

$$V_{\boldsymbol{k},\boldsymbol{k}'} = \begin{cases} -v/V & \text{if } |\xi_{\boldsymbol{k}}| < \hbar\omega_{\rm D} \text{ and } |\xi_{\boldsymbol{k}'}| < \hbar\omega_{\rm D} \\ 0 & \text{otherwise} \quad . \end{cases} \tag{12.71}$$

Here $v > 0$, so the interaction is attractive, but only when $\xi_{\boldsymbol{k}}$ and $\xi_{\boldsymbol{k}'}$ are within an energy $\hbar\omega_{\rm D}$ of zero. For phonon-mediated superconductivity, $\omega_{\rm D}$ is the Debye frequency, which is the phonon bandwidth.

12.6.1 Solution at zero temperature

We first solve the second of Eqns. 12.70, by assuming

$$\Delta_{\boldsymbol{k}} = \begin{cases} \Delta\, e^{i\phi} & \text{if } |\xi_{\boldsymbol{k}}| < \hbar\omega_{\rm D} \\ 0 & \text{otherwise} \quad , \end{cases} \tag{12.72}$$

with Δ real. We then have[9]

$$\Delta = +v\int\frac{d^3k}{(2\pi)^3}\,\frac{\Delta}{2E_{\boldsymbol{k}}}\,\Theta\big(\hbar\omega_{\rm D} - |\xi_{\boldsymbol{k}}|\big) = \tfrac{1}{2}v\,g(\varepsilon_{\rm F})\int_0^{\hbar\omega_{\rm D}} d\xi\,\frac{\Delta}{\sqrt{\xi^2+\Delta^2}} \quad . \tag{12.73}$$

[9] We assume the density of states $g(\varepsilon)$ is slowly varying in the vicinity of the chemical potential and approximate it at $g(\varepsilon_{\rm F})$. In fact, we should more properly call it $g(\mu)$, but as a practical matter $\mu \simeq \varepsilon_{\rm F}$ at temperatures low enough to be in the superconducting phase. Note that $g(\varepsilon_{\rm F})$ is the total DOS for both spin species. In the literature, one often encounters the expression $N(0)$, which is the DOS per spin at the Fermi level, *i.e.* $N(0) = \frac{1}{2}\,g(\varepsilon_{\rm F})$.

Cancelling out the common factors of Δ on each side, we obtain

$$1 = \tfrac{1}{2} v\, g(\varepsilon_{\rm F}) \int_0^{\hbar\omega_{\rm D}/\Delta} ds\, (1+s^2)^{-1/2} = \tfrac{1}{2} v\, g(\varepsilon_{\rm F})\, \sinh^{-1}\big(\hbar\omega_{\rm D}/\Delta\big) \quad . \tag{12.74}$$

Thus, writing $\Delta_0 \equiv \Delta(0)$ for the zero temperature gap,

$$\Delta_0 = \frac{\hbar\omega_{\rm D}}{\sinh\big(2/g(\varepsilon_{\rm F})\, v\big)} \simeq 2\hbar\omega_{\rm D} \exp\left(-\frac{2}{g(\varepsilon_{\rm F})\, v}\right) \quad , \tag{12.75}$$

where $g(\varepsilon_{\rm F})$ is the total electronic DOS (for both spin species) at the Fermi level. Notice that, as promised, the argument of the exponent is one half as large as what we found in our solution of the Cooper problem, in Eqn. 12.25.

12.6.2 Condensation energy

We now evaluate the zero temperature expectation of $\hat{K}_{\rm BCS}$ from Eqn. 12.62. To get the correct answer, it is essential that we retain the term corresponding to the constant energy shift in the mean field Hamiltonian, *i.e.* the last term on the RHS of Eqn. 12.62. Invoking the gap equation $\Delta_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'} \langle c_{-\boldsymbol{k}'\downarrow}\, c_{\boldsymbol{k}'\uparrow} \rangle$, we have

$$\langle\, {\rm G}\,|\, \hat{K}_{\rm BCS}\,|\, {\rm G}\,\rangle = \sum_{\boldsymbol{k}} \left(\xi_{\boldsymbol{k}} - E_{\boldsymbol{k}} + \frac{|\Delta_{\boldsymbol{k}}|^2}{2E_{\boldsymbol{k}}} \right) \quad . \tag{12.76}$$

From this we subtract the ground state energy of the metallic phase, *i.e.* when $\Delta_{\boldsymbol{k}} = 0$, which is $2\sum_{\boldsymbol{k}} \xi_{\boldsymbol{k}}\, \Theta(k_{\rm F} - k)$. The difference is the condensation energy. Adopting the model interaction potential in Eqn. 12.71, we have

$$\begin{aligned} E_{\rm s} - E_{\rm n} &= \sum_{\boldsymbol{k}} \left(\xi_{\boldsymbol{k}} - E_{\boldsymbol{k}} + \frac{|\Delta_{\boldsymbol{k}}|^2}{2E_{\boldsymbol{k}}} - 2\xi_{\boldsymbol{k}}\, \Theta(k_{\rm F} - k) \right) \\ &= 2\sum_{\boldsymbol{k}} \big(\xi_{\boldsymbol{k}} - E_{\boldsymbol{k}}\big)\, \Theta(\xi_{\boldsymbol{k}})\, \Theta(\hbar\omega_{\rm D} - \xi_{\boldsymbol{k}}) + \sum_{\boldsymbol{k}} \frac{\Delta_0^2}{2E_{\boldsymbol{k}}}\, \Theta\big(\hbar\omega_{\rm D} - |\xi_{\boldsymbol{k}}|\big) \quad , \end{aligned} \tag{12.77}$$

where we have linearized about $k = k_{\rm F}$. We then have

$$\begin{aligned} E_{\rm s} - E_{\rm n} &= V g(\varepsilon_{\rm F})\, \Delta_0^2 \int_0^{\hbar\omega_{\rm D}/\Delta_0} ds \left(s - \sqrt{s^2+1} + \frac{1}{2\sqrt{s^2+1}} \right) \\ &= \tfrac{1}{2}\, V g(\varepsilon_{\rm F})\, \Delta_0^2 \left(x^2 - x\sqrt{1+x^2} \right) \approx -\tfrac{1}{4}\, V g(\varepsilon_{\rm F})\, \Delta_0^2 \quad , \end{aligned} \tag{12.78}$$

where $x \equiv \hbar\omega_{\rm D}/\Delta_0$. The condensation energy density is therefore $-\frac{1}{4}\,g(\varepsilon_{\rm F})\,\Delta_0^2$, which may be equated with $-H_{\rm c}^2/8\pi$, where $H_{\rm c}$ is the thermodynamic critical field. Thus, we find

$$H_{\rm c}(0) = \sqrt{2\pi g(\varepsilon_{\rm F})}\,\Delta_0 \quad , \tag{12.79}$$

which relates the thermodynamic critical field to the superconducting gap, at $T = 0$.

12.7 Coherence factors and quasiparticle energies

When $\Delta_{\boldsymbol{k}} = 0$, we have $E_{\boldsymbol{k}} = |\xi_{\boldsymbol{k}}|$. When $\hbar\omega_{\rm D} \ll \varepsilon_{\rm F}$, there is a very narrow window surrounding $k = k_{\rm F}$ where $E_{\boldsymbol{k}}$ departs from $|\xi_{\boldsymbol{k}}|$, as shown in the bottom panel of Fig. 12.4. Note the *energy gap* in the quasiparticle dispersion, where the minimum excitation energy is given by[10]

$$\min_{\boldsymbol{k}} E_{\boldsymbol{k}} = E_{k_{\rm F}} = \Delta_0 \quad . \tag{12.80}$$

In the top panel of Fig. 12.4 we plot the coherence factors $\sin^2\vartheta_{\boldsymbol{k}}$ and $\cos^2\vartheta_{\boldsymbol{k}}$. Note that $\sin^2\vartheta_{\boldsymbol{k}}$ approaches unity for $k < k_{\rm F}$ and $\cos^2\vartheta_{\boldsymbol{k}}$ approaches unity for $k > k_{\rm F}$, aside for the narrow window of width $\delta k \simeq \Delta_0/\hbar v_{\rm F}$. Recall that

$$\gamma^\dagger_{\boldsymbol{k}\sigma} = \cos\vartheta_{\boldsymbol{k}}\,c^\dagger_{\boldsymbol{k}\sigma} + \sigma\sin\vartheta_{\boldsymbol{k}}\,e^{-i\phi_{\boldsymbol{k}}}\,c_{-\boldsymbol{k}\,-\sigma} \quad . \tag{12.81}$$

Thus we see that the quasiparticle creation operator $\gamma^\dagger_{\boldsymbol{k}\sigma}$ creates an electron in the state $|\,\boldsymbol{k}\,\sigma\,\rangle$ when $\cos^2\vartheta_{\boldsymbol{k}} \simeq 1$, and a hole in the state $|\,-\boldsymbol{k}\;-\sigma\,\rangle$ when $\sin^2\vartheta_{\boldsymbol{k}} \simeq 1$. In the aforementioned narrow window $|k - k_{\rm F}| \lesssim \Delta_0/\hbar v_{\rm F}$, the quasiparticle creates a linear combination of electron and hole states. Typically $\Delta_0 \sim 10^{-4}\,\varepsilon_{\rm F}$, since metallic Fermi energies are on the order of tens of thousands of Kelvins, while Δ_0 is on the order of Kelvins or tens of Kelvins. Thus, $\delta k \lesssim 10^{-3} k_{\rm F}$. The difference between the superconducting state and the metallic state all takes place within an onion skin at the Fermi surface!

Note that for the model interaction $V_{\boldsymbol{k},\boldsymbol{k}'}$ of Eqn. 12.71, the solution $\Delta_{\boldsymbol{k}}$ in Eqn. 12.72 is actually *discontinuous* when $\xi_{\boldsymbol{k}} = \pm\hbar\omega_{\rm D}$, *i.e.* when $k = k^*_\pm \equiv k_{\rm F} \pm \omega_{\rm D}/v_{\rm F}$. Therefore, the energy dispersion $E_{\boldsymbol{k}}$ is also discontinuous along these surfaces. However, the magnitude of the discontinuity is

$$\delta E = \sqrt{(\hbar\omega_{\rm D})^2 + \Delta_0^2} - \hbar\omega_{\rm D} \approx \frac{\Delta_0^2}{2\hbar\omega_{\rm D}} \quad . \tag{12.82}$$

Therefore $\delta E/E_{k^*_\pm} \approx \Delta_0^2/2(\hbar\omega_{\rm D})^2 \propto \exp\big(-4/g(\varepsilon_{\rm F})\,v\big)$, which is very tiny in weak coupling, where $g(\varepsilon_{\rm F})\,v \ll 1$. Note that the ground state is largely unaffected for electronic states in the vicinity of this (unphysical) energy discontinuity. The coherence factors are distinguished from those of a Fermi liquid only in regions where $\langle c^\dagger_{\boldsymbol{k}\uparrow} c^\dagger_{-\boldsymbol{k}\downarrow}\rangle$ is appreciable, which requires $\xi_{\boldsymbol{k}}$ to be on the order of $\Delta_{\boldsymbol{k}}$. This only happens when $|k - k_{\rm F}| \lesssim \Delta_0/\hbar v_{\rm F}$, as discussed in the previous paragraph. In a more physical model, the interaction $V_{\boldsymbol{k},\boldsymbol{k}'}$ and the solution $\Delta_{\boldsymbol{k}}$ would not be discontinuous functions of $\boldsymbol{k}$.

[10] Here we assume, without loss of generality, that Δ is real.

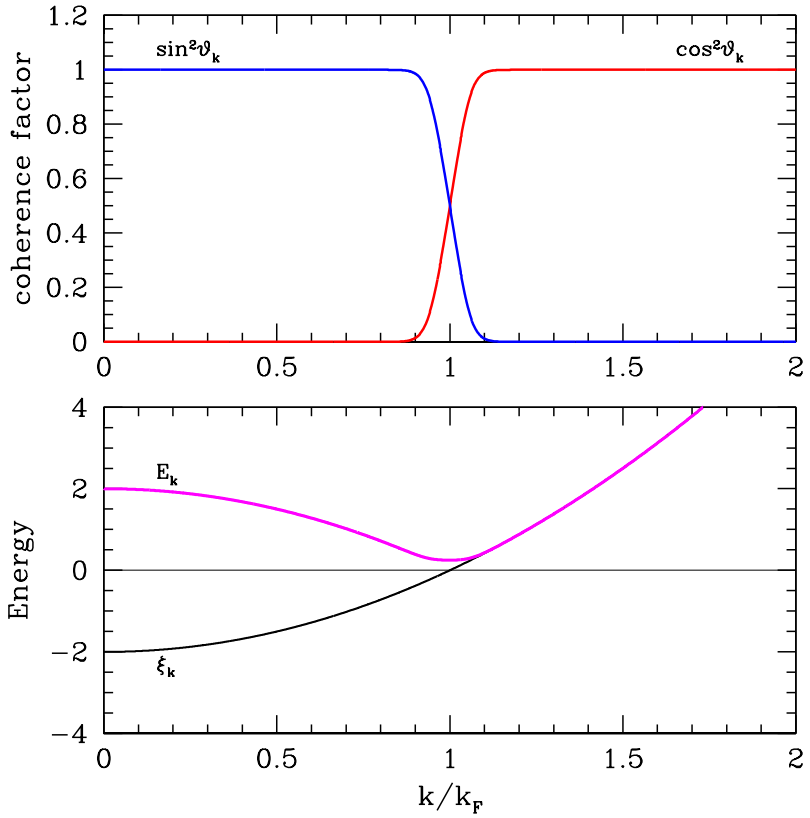


Figure 12.4: Top panel: BCS coherence factors $\sin^2\vartheta_{\boldsymbol{k}}$ (blue) and $\cos^2\vartheta_{\boldsymbol{k}}$ (red). Bottom panel: the functions $\xi_{\boldsymbol{k}}$ (black) and $E_{\boldsymbol{k}}$ (magenta). The minimum value of the magenta curve is the superconducting gap Δ_0.

12.8 Number and Phase

The BCS ground state wavefunction $|\,\mathrm{G}\,\rangle$ was given in Eqn. 12.63. Consider the state

$$|\,\mathrm{G}(\alpha)\,\rangle = \prod_{\boldsymbol{k}} \left(\cos\vartheta_{\boldsymbol{k}} - e^{i\alpha}\, e^{i\phi_{\boldsymbol{k}}} \sin\vartheta_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\uparrow}\, c^\dagger_{-\boldsymbol{k}\downarrow}\right)|\,0\,\rangle \quad . \tag{12.83}$$

This is the ground state when the gap function $\Delta_{\boldsymbol{k}}$ is multiplied by the uniform phase factor $e^{i\alpha}$. We shall here abbreviate $|\,\alpha\,\rangle \equiv |\,\mathrm{G}(\alpha)\,\rangle$.

Now consider the action of the number operator on $|\,\alpha\,\rangle$:

$$\begin{aligned}\hat{N}\,|\,\alpha\,\rangle &= \sum_{\boldsymbol{k}} \left(c^\dagger_{\boldsymbol{k}\uparrow} c_{\boldsymbol{k}\uparrow} + c^\dagger_{-\boldsymbol{k}\downarrow} c_{-\boldsymbol{k}\downarrow}\right)|\,\alpha\,\rangle \qquad (12.84)\\ &= -2\sum_{\boldsymbol{k}} e^{i\alpha}\, e^{i\phi_{\boldsymbol{k}}}\, \sin\vartheta_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\uparrow}\, c^\dagger_{-\boldsymbol{k}\downarrow} \prod_{\boldsymbol{k}'\neq\boldsymbol{k}} \left(\cos\vartheta_{\boldsymbol{k}'} - e^{i\alpha}\, e^{i\phi_{\boldsymbol{k}'}} \sin\vartheta_{\boldsymbol{k}'}\, c^\dagger_{\boldsymbol{k}'\uparrow}\, c^\dagger_{-\boldsymbol{k}'\downarrow}\right)|\,0\,\rangle = \frac{2}{i}\frac{\partial}{\partial\alpha}\,|\,\alpha\,\rangle \quad .\end{aligned}$$

If we define the number of Cooper pairs as $\hat{M} \equiv \frac{1}{2}\hat{N}$, then we may identify $\hat{M} = \frac{1}{i}\frac{\partial}{\partial\alpha}$. Furthermore, we may project $|\,\mathrm{G}\,\rangle$ onto a state of definite particle number by defining

$$|\,M\,\rangle = \int_{-\pi}^{\pi} \frac{d\alpha}{2\pi}\, e^{-iM\alpha}\,|\,\alpha\,\rangle \quad . \tag{12.85}$$

The state $|\,M\,\rangle$ has $N = 2M$ particles, *i.e.* M Cooper pairs. One can easily compute the number fluctuations in the state $|\,\mathrm{G}(\alpha)\,\rangle$:

$$\frac{\langle\,\alpha\,|\,\hat{N}^2\,|\,\alpha\,\rangle - \langle\,\alpha\,|\,\hat{N}\,|\,\alpha\,\rangle^2}{\langle\,\alpha\,|\,\hat{N}\,|\,\alpha\,\rangle} = \frac{2\int d^3k\,\sin^2\vartheta_{\boldsymbol{k}}\,\cos^2\vartheta_{\boldsymbol{k}}}{\int d^3k\,\sin^2\vartheta_{\boldsymbol{k}}} = \frac{4\Delta_0}{\hbar\omega_{\mathrm{D}}}\tan^{-1}\!\left(\frac{\hbar\omega_{\mathrm{D}}}{\Delta_0}\right) \simeq \frac{2\pi\Delta_0}{\hbar\omega_{\mathrm{D}}} \quad . \tag{12.86}$$

Thus, $(\Delta N)_{\mathrm{RMS}} \propto \langle N\rangle^{1/2}$. Note that $(\Delta N)_{\mathrm{RMS}} = 0$ in the normal state, where $\sin\vartheta_{\boldsymbol{k}}\cos\vartheta_{\boldsymbol{k}} = 0$.

12.9 Finite Temperature

The gap equation at finite temperature takes the form

$$\Delta_{\boldsymbol{k}} = -\sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\frac{\Delta_{\boldsymbol{k}'}}{2E_{\boldsymbol{k}'}}\,\tanh\!\left(\frac{E_{\boldsymbol{k}'}}{2k_{\mathrm{B}}T}\right) \quad . \tag{12.87}$$

It is easy to see that we have no solutions other than the trivial one $\Delta_{\boldsymbol{k}} = 0$ in the $T \to \infty$ limit, for the gap equation then becomes $\sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\Delta_{\boldsymbol{k}'} = -4k_{\mathrm{B}}T\,\Delta_{\boldsymbol{k}}$, and if the eigenspectrum of $V_{\boldsymbol{k},\boldsymbol{k}'}$ is bounded, there is no solution for $k_{\mathrm{B}}T$ greater than the largest eigenvalue of $-V_{\boldsymbol{k},\boldsymbol{k}'}$.

To find the critical temperature where the gap collapses, again we assume the forms in Eqns. 12.71 and 12.72, in which case we have

$$1 = \tfrac{1}{2}\,g(\varepsilon_{\mathsf{F}})\,v\int_0^{\hbar\omega_{\mathrm{D}}} \frac{d\xi}{\sqrt{\xi^2+\Delta^2}}\,\tanh\!\left(\frac{\sqrt{\xi^2+\Delta^2}}{2k_{\mathrm{B}}T}\right) \quad . \tag{12.88}$$

It is clear that $\Delta(T)$ is a decreasing function of temperature, which vanishes at $T = T_{\mathrm{c}}$, where T_{c} is determined by the equation

$$\int_0^{\Lambda/2} ds\; s^{-1}\tanh(s) = \frac{2}{g(\varepsilon_{\mathsf{F}})\,v} \quad , \tag{12.89}$$

where $\Lambda = \hbar\omega_{\mathrm{D}}/k_{\mathrm{B}}T_{\mathrm{c}}$. One finds, for large Λ,

$$\begin{aligned} I(\Lambda) = \int_0^{\Lambda/2} ds\; s^{-1}\tanh(s) &= \ln\left(\tfrac{1}{2}\Lambda\right)\tanh\left(\tfrac{1}{2}\Lambda\right) - \int_0^{\Lambda/2} ds\,\frac{\ln s}{\cosh^2 s} \\ &= \ln\Lambda + \ln\left(2\,e^{\mathrm{C}}/\pi\right) + \mathcal{O}\left(e^{-\Lambda/2}\right) \quad , \end{aligned} \tag{12.90}$$

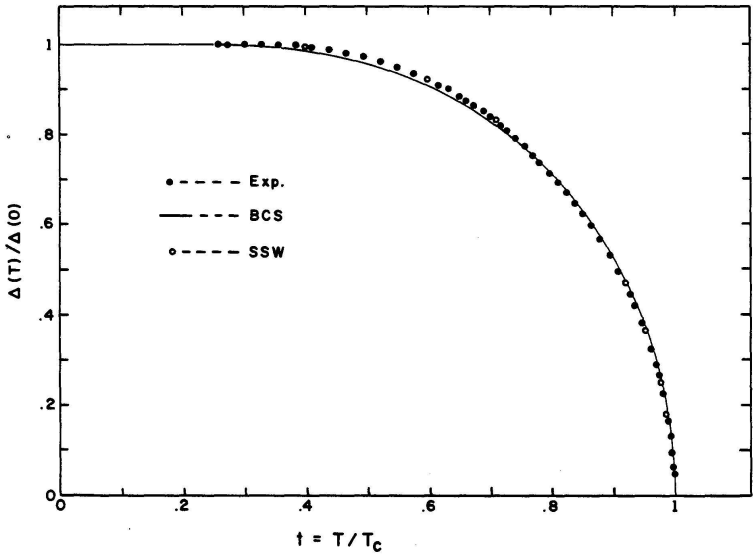


Figure 12.5: Temperature dependence of the energy gap in Pb as determined by tunneling *versus* prediction of BCS theory. From R. F. Gasparovic, B. N. Taylor, and R. E. Eck, *Sol. State Comm.* **4**, 59 (1966). Deviations from the BCS theory are accounted for by numerical calculations at strong coupling by Swihart, Scalapino, and Wada (1965).

where $\mathrm{C} = 0.57721566\ldots$ is the Euler-Mascheroni constant. One has $2\,e^{\mathrm{C}}/\pi = 1.134$, so

$$k_{\rm B}T_{\rm c} = 1.134\,\hbar\omega_{\rm D}\,e^{-2/g(\varepsilon_{\rm F})\,v} \quad . \tag{12.91}$$

Comparing with Eqn. 12.75, we obtain the famous result

$$2\Delta(0) = 2\pi e^{-\mathrm{C}}\,k_{\rm B}T_{\rm c} \simeq 3.52\,k_{\rm B}T_{\rm c} \quad . \tag{12.92}$$

As we shall derive presently, just below the critical temperature, one has

$$\Delta(T) = 1.734\,\Delta(0)\left(1-\frac{T}{T_{\rm c}}\right)^{1/2} \simeq 3.06\,k_{\rm B}T_{\rm c}\left(1-\frac{T}{T_{\rm c}}\right)^{1/2} \quad . \tag{12.93}$$

12.9.1 Isotope effect

The prefactor in Eqn. 12.91 is proportional to the Debye energy $\hbar\omega_{\rm D}$. Thus,

$$\ln T_{\rm c} = \ln\omega_{\rm D} - \frac{2}{g(\varepsilon_{\rm F})\,v} + \text{const.} \quad . \tag{12.94}$$

If we imagine varying only the mass of the ions, via isotopic substitution, then $g(\varepsilon_{\rm F})$ and v do not change, and we have

$$\delta\ln T_{\rm c} = \delta\ln\omega_{\rm D} = -\tfrac{1}{2}\,\delta\ln M \quad , \tag{12.95}$$

where M is the ion mass. Thus, isotopically increasing the ion mass leads to a concomitant reduction in $T_{\rm c}$ according to BCS theory. This is fairly well confirmed in experiments on low $T_{\rm c}$ materials.

12.9.2 Landau free energy of a superconductor

Quantum statistical mechanics of noninteracting fermions applied to $\hat{K}_{\rm BCS}$ in Eqn. 12.62 yields the Landau free energy

$$\Omega_{\rm s} = -2k_{\rm B}T \sum_{\boldsymbol{k}} \ln\bigl(1 + e^{-E_{\boldsymbol{k}}/k_{\rm B}T}\bigr) + \sum_{\boldsymbol{k}} \left\{ \xi_{\boldsymbol{k}} - E_{\boldsymbol{k}} + \frac{|\Delta_{\boldsymbol{k}}|^2}{2E_{\boldsymbol{k}}} \, \tanh\left(\frac{E_{\boldsymbol{k}}}{2k_{\rm B}T}\right) \right\} \quad . \tag{12.96}$$

The corresponding result for the normal state ($\Delta_{\boldsymbol{k}} = 0$) is

$$\Omega_{\rm n} = -2k_{\rm B}T \sum_{\boldsymbol{k}} \ln\bigl(1 + e^{-|\xi_{\boldsymbol{k}}|/k_{\rm B}T}\bigr) + \sum_{\boldsymbol{k}} \Bigl(\xi_{\boldsymbol{k}} - |\xi_{\boldsymbol{k}}|\Bigr) \quad . \tag{12.97}$$

Thus, the difference is

$$\Omega_{\rm s} - \Omega_{\rm n} = -2k_{\rm B}T \sum_{\boldsymbol{k}} \ln\left(\frac{1 + e^{-E_{\boldsymbol{k}}/k_{\rm B}T}}{1 + e^{-|\xi_{\boldsymbol{k}}|/k_{\rm B}T}}\right) + \sum_{\boldsymbol{k}} \left\{ |\xi_{\boldsymbol{k}}| - E_{\boldsymbol{k}} + \frac{|\Delta_{\boldsymbol{k}}|^2}{2E_{\boldsymbol{k}}} \, \tanh\left(\frac{E_{\boldsymbol{k}}}{2k_{\rm B}T}\right) \right\} \quad . \tag{12.98}$$

We now invoke the model interaction in Eqn. 12.71. Recall that the solution to the gap equation is of the form $\Delta_{\boldsymbol{k}}(T) = \Delta(T)\,\Theta\bigl(\hbar\omega_{\rm D} - |\xi_{\boldsymbol{k}}|\bigr)$. We then have

$$\begin{aligned} \frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = \frac{\Delta^2}{v} - \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, \Delta^2 \left\{ \frac{\hbar\omega_{\rm D}}{\Delta} \sqrt{1 + \left(\frac{\hbar\omega_{\rm D}}{\Delta}\right)^2} - \left(\frac{\hbar\omega_{\rm D}}{\Delta}\right)^2 + \sinh^{-1}\left(\frac{\hbar\omega_{\rm D}}{\Delta}\right) \right\} \\ - 2\, g(\varepsilon_{\rm F})\, k_{\rm B}T\, \Delta \int_0^\infty ds \, \ln\left(1 + e^{-\sqrt{1+s^2}\,\Delta/k_{\rm B}T}\right) + \tfrac{1}{6}\,\pi^2\, g(\varepsilon_{\rm F})\, (k_{\rm B}T)^2 + \ldots \quad , \end{aligned} \tag{12.99}$$

where the terms in the ellipsis are of $\mathcal{O}(T^4)$, arising from the Sommerfeld expansion of the low temperature normal state free energy. We now expand this result in the vicinity of $T = 0$ and $T = T_{\rm c}$. In the weak coupling limit, throughout this entire region we have $\Delta \ll \hbar\omega_{\rm D}$, so we proceed to expand in the small ratio, writing

$$\begin{aligned} \frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = -\tfrac{1}{4}\, g(\varepsilon_{\rm F})\, \Delta^2 \left\{ 1 + 2\ln\left(\frac{\Delta_0}{\Delta}\right) - \left(\frac{\Delta}{2\hbar\omega_{\rm D}}\right)^2 + \mathcal{O}\bigl(\Delta^4\bigr) \right\} \\ - 2\, g(\varepsilon_{\rm F})\, k_{\rm B}T\Delta \int_0^\infty ds \, \ln\left(1 + e^{-\sqrt{1+s^2}\,\Delta/k_{\rm B}T}\right) + \tfrac{1}{6}\,\pi^2\, g(\varepsilon_{\rm F})\, (k_{\rm B}T)^2 + \ldots \quad , \end{aligned} \tag{12.100}$$

where $\Delta_0 = \Delta(0) = \pi e^{-{\rm C}}\, k_{\rm B}T_{\rm c}$. The asymptotic analysis of this expression in the limits $T \to 0^+$ and $T \to T_{\rm c}^-$ is discussed in the appendix §12.14.

$T \to 0^+$

In the limit $T \to 0$, we find

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = -\tfrac{1}{4}\, g(\varepsilon_{\rm F})\, \Delta^2 \left\{ 1 + 2\ln\left(\frac{\Delta_0}{\Delta}\right) + \mathcal{O}\big(\Delta^2\big) \right\} \tag{12.101}$$

$$- g(\varepsilon_{\rm F}) \sqrt{2\pi (k_{\rm B}T)^3 \Delta}\, e^{-\Delta/k_{\rm B}T} + \tfrac{1}{6}\,\pi^2\, g(\varepsilon_{\rm F})\, (k_{\rm B}T)^2 + \ldots \quad .$$

Differentiating the above expression with respect to Δ, we obtain a self-consistent equation for the gap $\Delta(T)$ at low temperatures:

$$\ln\left(\frac{\Delta}{\Delta_0}\right) = -\sqrt{\frac{2\pi k_{\rm B}T}{\Delta}}\, e^{-\Delta/k_{\rm B}T} \left(1 - \frac{k_{\rm B}T}{2\Delta} + \ldots\right) \tag{12.102}$$

Thus,

$$\Delta(T) = \Delta_0 - \sqrt{2\pi \Delta_0 k_{\rm B}T}\, e^{-\Delta_0/k_{\rm B}T} + \ldots \quad . \tag{12.103}$$

Substituting this expression into Eqn. 12.101, we find

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = -\tfrac{1}{4}\, g(\varepsilon_{\rm F})\, \Delta_0^2 - g(\varepsilon_{\rm F}) \sqrt{2\pi \Delta_0\, (k_{\rm B}T)^3}\, e^{-\Delta_0/k_{\rm B}T} + \tfrac{1}{6}\,\pi^2\, g(\varepsilon_{\rm F})\, (k_{\rm B}T)^2 + \ldots \quad . \tag{12.104}$$

Equating this with the condensation energy density, $-H_{\rm c}^2(T)/8\pi$, and invoking our previous result, $\Delta_0 = \pi e^{-\rm C}\, k_{\rm B}T_{\rm c}$, we find

$$H_{\rm c}(T) = H_{\rm c}(0) \left\{ 1 - \overbrace{\tfrac{1}{3}\, e^{2\rm C}}^{\approx 1.057} \left(\frac{T}{T_{\rm c}}\right)^2 + \ldots \right\} \quad , \tag{12.105}$$

where $H_{\rm c}(0) = \sqrt{2\pi\, g(\varepsilon_{\rm F})}\, \Delta_0$.

$T \to T_{\rm c}^-$

In this limit, one finds

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, \ln\left(\frac{T}{T_{\rm c}}\right) \Delta^2 + \frac{7\,\zeta(3)}{32\pi^2}\, \frac{g(\varepsilon_{\rm F})}{(k_{\rm B}T)^2}\, \Delta^4 + \mathcal{O}\big(\Delta^6\big) \quad . \tag{12.106}$$

This is of the standard Landau form,

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = \tilde{a}(T)\, \Delta^2 + \tfrac{1}{2}\, \tilde{b}(T)\, \Delta^4 \quad , \tag{12.107}$$

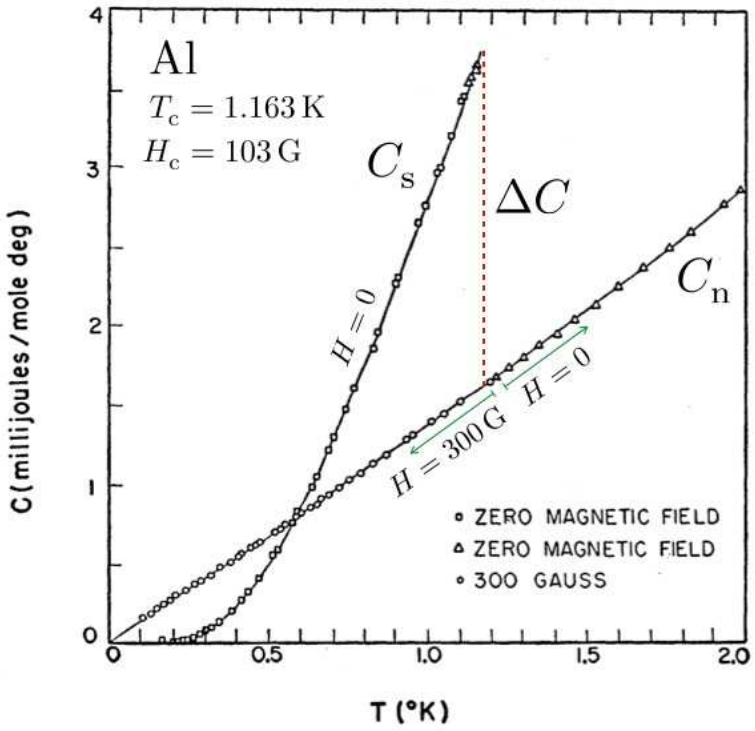


Figure 12.6: Heat capacity in aluminum at low temperatures, from N. K. Phillips, *Phys. Rev. 114*, **3** (1959). The zero field superconducting transition occurs at $T_c = 1.163\,\mathrm{K}$. Comparison with normal state C below T_c is made possible by imposing a magnetic field $H > H_c$. This destroys the superconducting state, but has little effect on the metal. A jump ΔC is observed at T_c, quantitatively in agreement BCS theory.

with coefficients

$$\tilde{a}(T) = \tfrac{1}{2}\, g(\varepsilon_{\rm F})\left(\frac{T}{T_c} - 1\right) \quad , \qquad \tilde{b} = \frac{7\,\zeta(3)}{16\pi^2}\,\frac{g(\varepsilon_{\rm F})}{(k_{\rm B}T_c)^2} \quad , \tag{12.108}$$

working here to lowest nontrivial order in $T - T_c$. The head capacity jump, according to Eqn. 1.44, is

$$c_{\rm s}(T_c^-) - c_{\rm n}(T_c^+) = \frac{T_c\big[\tilde{a}'(T_c)\big]^2}{\tilde{b}(T_c)} = \frac{4\pi^2}{7\,\zeta(3)}\, g(\varepsilon_{\rm F})\, k_{\rm B}^2 T_c \quad . \tag{12.109}$$

The normal state heat capacity at $T = T_c$ is $c_{\rm n} = \frac{1}{3}\pi^2 g(\varepsilon_{\rm F})\, k_{\rm B}^2 T_c$, hence

$$\frac{c_{\rm s}(T_c^-) - c_{\rm n}(T_c^+)}{c_{\rm n}(T_c^+)} = \frac{12}{7\,\zeta(3)} = 1.43 \quad . \tag{12.110}$$

This universal ratio is closely reproduced in many experiments; see, for example, Fig. 12.6.

The order parameter is given by

$$\Delta^2(T) = -\frac{\tilde{a}(T)}{\tilde{b}(T)} = \frac{8\pi^2(k_{\rm B}T_{\rm c})^2}{7\,\zeta(3)}\left(1-\frac{T}{T_{\rm c}}\right) = \frac{8\,e^{2{\rm C}}}{7\,\zeta(3)}\left(1-\frac{T}{T_{\rm c}}\right)\Delta^2(0) \quad , \tag{12.111}$$

where we have used $\Delta(0) = \pi\,e^{-{\rm C}}\,k_{\rm B}T_{\rm c}$. Thus,

$$\frac{\Delta(T)}{\Delta(0)} = \overbrace{\left(\frac{8\,e^{2{\rm C}}}{7\,\zeta(3)}\right)^{1/2}}^{\approx 1.734}\left(1-\frac{T}{T_{\rm c}}\right)^{1/2} \quad . \tag{12.112}$$

The thermodynamic critical field just below $T_{\rm c}$ is obtained by equating the energies $-\tilde{a}^2/2\tilde{b}$ and $-H_{\rm c}^2/8\pi$. Therefore

$$\frac{H_{\rm c}(T)}{H_{\rm c}(0)} = \left(\frac{8\,e^{2{\rm C}}}{7\,\zeta(3)}\right)^{1/2}\left(1-\frac{T}{T_{\rm c}}\right) \simeq 1.734\left(1-\frac{T}{T_{\rm c}}\right) \quad . \tag{12.113}$$

12.10 Paramagnetic Susceptibility

Suppose we add a weak magnetic field, the effect of which is described by the perturbation Hamiltonian

$$\hat{H}_1 = -\mu_{\rm B}H\sum_{\boldsymbol{k},\sigma}\sigma\,c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}\sigma} = -\mu_{\rm B}H\sum_{\boldsymbol{k},\sigma}\sigma\,\gamma^\dagger_{\boldsymbol{k}\sigma}\,\gamma_{\boldsymbol{k}\sigma} \quad . \tag{12.114}$$

The shift in the Landau free energy due to the field is then $\Delta\Omega_{\rm s}(T,V,\mu,H) = \Omega_{\rm s}(T,V,\mu,H) - \Omega_{\rm s}(T,V,\mu,0)$. We have

$$\begin{aligned}\Delta\Omega_{\rm s}(T,V,\mu,H) &= -k_{\rm B}T\sum_{\boldsymbol{k},\sigma}\ln\left(\frac{1+e^{-\beta(E_{\boldsymbol{k}}+\sigma\mu_{\rm B}H)}}{1+e^{-\beta E_{\boldsymbol{k}}}}\right)\\ &= -\beta\,(\mu_{\rm B}H)^2\sum_{\boldsymbol{k}}\frac{e^{\beta E_{\boldsymbol{k}}}}{\left(e^{\beta E_{\boldsymbol{k}}}+1\right)^2} + \mathcal{O}(H^4) \quad .\end{aligned} \tag{12.115}$$

The magnetic susceptibility is then

$$\chi_{\rm s} = -\frac{1}{V}\,\frac{\partial^2\Delta\Omega_{\rm s}}{\partial H^2} = g(\varepsilon_{\rm F})\,\mu_{\rm B}^2\,\mathcal{Y}(T) \quad , \tag{12.116}$$

where

$$\mathcal{Y}(T) = 2\int\limits_0^\infty d\xi\,\left(-\frac{\partial f}{\partial E}\right) = \tfrac{1}{2}\beta\int\limits_0^\infty d\xi\,\mathrm{sech}^2\left(\tfrac{1}{2}\beta\sqrt{\xi^2+\Delta^2}\right) \tag{12.117}$$

is the *Yoshida function*. Note that $\mathcal{Y}(T_{\rm c}) = \int_0^\infty du\,\mathrm{sech}^2 u = 1$, and $\mathcal{Y}(T\to 0) \simeq (2\pi\beta\Delta)^{1/2}\exp(-\beta\Delta)$, which is exponentially suppressed. Since $\chi_{\rm n} = g(\varepsilon_{\rm F})\,\mu_{\rm B}^2$ is the normal state Pauli susceptibility, we have that the ratio of superconducting to normal state susceptibilities is $\chi_{\rm s}(T)/\chi_{\rm n}(T) = \mathcal{Y}(T)$. This vanishes exponentially as $T\to 0$ because it takes a finite energy Δ to create a Bogoliubov quasiparticle out of the spin singlet BCS ground state.

In metals, the nuclear spins experience a shift in their resonance energy in the presence of an external magnetic field, due to their coupling to conduction electrons via the hyperfine interaction. This is called the *Knight shift*, after Walter Knight, who first discovered this phenomenon at Berkeley in 1949. The magnetic field polarizes the metallic conduction electrons, which in turn impose an extra effective field, through the hyperfine coupling, on the nuclei. In superconductors, the electrons remain unpolarized in a weak magnetic field owing to the superconducting gap. Thus there is no Knight shift.

As we have seen from the Ginzburg-Landau theory, when the field is sufficiently strong, superconductivity is destroyed (type I), or there is a mixed phase at intermediate fields where magnetic flux penetrates the superconductor in the form of vortex lines. Our analysis here is valid only for weak fields.

12.11 Finite Momentum Condensate

The BCS reduced Hamiltonian of Eqn. 12.43 involved interactions between $\boldsymbol{q} = 0$ Cooper pairs only. The reduced BCS Hamiltonian was given by

$$\hat{H}_{\rm red} = \sum_{\boldsymbol{k},\sigma} \varepsilon_{\boldsymbol{k}}\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{p}} V_{\boldsymbol{k},\boldsymbol{k}'}\, b^\dagger_{\boldsymbol{k},\boldsymbol{p}}\, b_{\boldsymbol{k}',\boldsymbol{p}} \quad . \tag{12.118}$$

where $b_{\boldsymbol{k},\boldsymbol{p}} = c_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{p}\downarrow}\, c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{p}\uparrow}$, and with zero momentum pairing, $\langle\, b_{\boldsymbol{k},\boldsymbol{p}}\,\rangle = \langle\, b_{\boldsymbol{k},\boldsymbol{0}}\,\rangle\,\delta_{\boldsymbol{p},\boldsymbol{0}}$. The gap $\Delta_{\boldsymbol{k}}$ is then given by $\Delta_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\langle\, b_{\boldsymbol{k}',\boldsymbol{0}}\,\rangle$. What happens, though, if we take

$$\langle\, b_{\boldsymbol{k},\boldsymbol{p}}\,\rangle = \big\langle c_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow}\, c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\big\rangle\,\delta_{\boldsymbol{p},\boldsymbol{q}} \qquad , \qquad \Delta_{\boldsymbol{k},\boldsymbol{q}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\big\langle c_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow}\, c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\big\rangle \quad , \tag{12.119}$$

corresponding to a finite momentum condensate? We then obtain

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k}} \begin{pmatrix} c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow} & c_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow} \end{pmatrix} \begin{pmatrix} \omega_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}} & \Delta_{\boldsymbol{k},\boldsymbol{q}} \\ \Delta^*_{\boldsymbol{k},\boldsymbol{q}} & -\omega_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}} \end{pmatrix} \begin{pmatrix} c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow} \\ c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow} \end{pmatrix} + \sum_{\boldsymbol{k}} \Big(\xi_{\boldsymbol{k}} - \Delta_{\boldsymbol{k},\boldsymbol{q}}\,\langle\, b^\dagger_{\boldsymbol{k},\boldsymbol{q}}\,\rangle \Big) \quad , \tag{12.120}$$

where

$$\omega_{\boldsymbol{k},\boldsymbol{q}} = \tfrac{1}{2}\Big(\xi_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} + \xi_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}}\Big) \qquad\qquad \xi_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} = \omega_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}} \tag{12.121}$$

$$\nu_{\boldsymbol{k},\boldsymbol{q}} = \tfrac{1}{2}\Big(\xi_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} - \xi_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}}\Big) \qquad\qquad \xi_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}} = \omega_{\boldsymbol{k},\boldsymbol{q}} - \nu_{\boldsymbol{k},\boldsymbol{q}} \quad . \tag{12.122}$$

Note that $\omega_{\boldsymbol{k},\boldsymbol{q}}$ is even under reversal of either $\boldsymbol{k}$ or $\boldsymbol{q}$, while $\nu_{\boldsymbol{k},\boldsymbol{q}}$ is odd under reversal of either $\boldsymbol{k}$ or $\boldsymbol{q}$. That is,

$$\omega_{\boldsymbol{k},\boldsymbol{q}} = \omega_{-\boldsymbol{k},\boldsymbol{q}} = \omega_{\boldsymbol{k},-\boldsymbol{q}} = \omega_{-\boldsymbol{k},-\boldsymbol{q}} \qquad , \qquad \nu_{\boldsymbol{k},\boldsymbol{q}} = -\nu_{-\boldsymbol{k},\boldsymbol{q}} = -\nu_{\boldsymbol{k},-\boldsymbol{q}} = \nu_{-\boldsymbol{k},-\boldsymbol{q}} \quad . \tag{12.123}$$

We now make a Bogoliubov transformation,

$$\begin{aligned} c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow} &= \cos\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,\gamma_{\boldsymbol{k},\boldsymbol{q},\uparrow} - \sin\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,e^{i\phi_{\boldsymbol{k},\boldsymbol{q}}}\,\gamma^\dagger_{-\boldsymbol{k},\boldsymbol{q},\downarrow} \\ c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow} &= \cos\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,\gamma^\dagger_{-\boldsymbol{k},\boldsymbol{q},\downarrow} + \sin\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,e^{i\phi_{\boldsymbol{k},\boldsymbol{q}}}\,\gamma_{\boldsymbol{k},\boldsymbol{q},\uparrow} \end{aligned} \tag{12.124}$$

with

$$\cos\vartheta_{\boldsymbol{k},\boldsymbol{q}} = \sqrt{\frac{E_{\boldsymbol{k},\boldsymbol{q}} + \omega_{\boldsymbol{k},\boldsymbol{q}}}{2E_{\boldsymbol{k},\boldsymbol{q}}}} \qquad\qquad \phi_{\boldsymbol{k},\boldsymbol{q}} = \mathsf{arg}(\Delta_{\boldsymbol{k},\boldsymbol{q}}) \tag{12.125}$$

$$\sin\vartheta_{\boldsymbol{k},\boldsymbol{q}} = \sqrt{\frac{E_{\boldsymbol{k},\boldsymbol{q}} - \omega_{\boldsymbol{k},\boldsymbol{q}}}{2E_{\boldsymbol{k},\boldsymbol{q}}}} \qquad\qquad E_{\boldsymbol{k},\boldsymbol{q}} = \sqrt{\omega^2_{\boldsymbol{k},\boldsymbol{q}} + |\Delta_{\boldsymbol{k},\boldsymbol{q}}|^2} \quad . \tag{12.126}$$

We then obtain

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k},\sigma}(E_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}})\,\gamma^\dagger_{\boldsymbol{k},\boldsymbol{q},\sigma}\gamma_{\boldsymbol{k},\boldsymbol{q},\sigma} + \sum_{\boldsymbol{k}}\Big(\xi_{\boldsymbol{k}} - E_{\boldsymbol{k},\boldsymbol{q}} + \Delta_{\boldsymbol{k},\boldsymbol{q}}\,\langle b^\dagger_{\boldsymbol{k},\boldsymbol{q}}\rangle\Big)\,. \tag{12.127}$$

Next, we compute the quantum statistical averages

$$\begin{aligned} \big\langle c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\,c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\big\rangle &= \cos^2\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,f(E_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}}) + \sin^2\vartheta_{\boldsymbol{k},\boldsymbol{q}}\Big[1 - f(E_{\boldsymbol{k},\boldsymbol{q}} - \nu_{\boldsymbol{k},\boldsymbol{q}})\Big] \\ &= \frac{1}{2}\left(1 + \frac{\omega_{\boldsymbol{k},\boldsymbol{q}}}{E_{\boldsymbol{k},\boldsymbol{q}}}\right) f(E_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}}) + \frac{1}{2}\left(1 - \frac{\omega_{\boldsymbol{k},\boldsymbol{q}}}{E_{\boldsymbol{k},\boldsymbol{q}}}\right)\Big[1 - f(E_{\boldsymbol{k},\boldsymbol{q}} - \nu_{\boldsymbol{k},\boldsymbol{q}})\Big] \end{aligned} \tag{12.128}$$

and

$$\begin{aligned} \big\langle c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\,c^\dagger_{-\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\downarrow}\big\rangle &= -\sin\vartheta_{\boldsymbol{k},\boldsymbol{q}}\cos\vartheta_{\boldsymbol{k},\boldsymbol{q}}\,e^{-i\phi_{\boldsymbol{k},\boldsymbol{q}}}\Big[1 - f(E_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}}) - f(E_{\boldsymbol{k},\boldsymbol{q}} - \nu_{\boldsymbol{k},\boldsymbol{q}})\Big] \\ &= -\frac{\Delta^*_{\boldsymbol{k},\boldsymbol{q}}}{2E_{\boldsymbol{k},\boldsymbol{q}}}\Big[1 - f(E_{\boldsymbol{k},\boldsymbol{q}} + \nu_{\boldsymbol{k},\boldsymbol{q}}) - f(E_{\boldsymbol{k},\boldsymbol{q}} - \nu_{\boldsymbol{k},\boldsymbol{q}})\Big] \quad . \end{aligned} \tag{12.129}$$

12.11.1 Gap equation for finite momentum condensate

We may now solve the $T = 0$ gap equation,

$$1 = -\sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\,\frac{1}{2E_{\boldsymbol{k}',\boldsymbol{q}}} = \tfrac{1}{2}\,g(\varepsilon_{\rm F})\,v\int\limits_0^{\hbar\omega_{\rm D}}\frac{d\xi}{\sqrt{(\xi+\eta_{\boldsymbol{q}})^2+|\Delta_{0,\boldsymbol{q}}|^2}}\ . \tag{12.130}$$

Here we have assumed the interaction $V_{\boldsymbol{k},\boldsymbol{k}'}$ of Eqn. 12.71, and we take

$$\Delta_{\boldsymbol{k},\boldsymbol{q}} = \Delta_{0,\boldsymbol{q}}\,\Theta\big(\hbar\omega_{\rm D} - |\xi_{\boldsymbol{k}}|\big)\quad . \tag{12.131}$$

We have also written $\omega_{\boldsymbol{k},\boldsymbol{q}} = \xi_{\boldsymbol{k}} + \eta_{\boldsymbol{q}}$. This form is valid for quadratic $\xi_{\boldsymbol{k}} = \frac{\hbar^2\boldsymbol{k}^2}{2m^*} - \mu$, in which case $\eta_{\boldsymbol{q}} = \hbar^2\boldsymbol{q}^2/8m^*$. We take $\Delta_{0,\boldsymbol{q}} \in \mathbb{R}$. We may now compute the critical wavevector $q_{\rm c}$ at which the $T = 0$ gap collapses:

$$1 = \tfrac{1}{2}\,g(\varepsilon_{\rm F})\,v\,\ln\left(\frac{\hbar\omega_{\rm D}+\eta_{q_{\rm c}}}{\eta_{q_{\rm c}}}\right)\qquad\Rightarrow\qquad \eta_{q_{\rm c}} \simeq \hbar\omega_{\rm D}\,e^{-2/g(\varepsilon_{\rm F})\,v} = \tfrac{1}{2}\,\Delta_0\quad , \tag{12.132}$$

whence $q_{\rm c} = 2\sqrt{m^*\Delta_0}\,/\hbar$. Here we have assumed weak coupling, *i.e.* $g(\varepsilon_{\rm F})\,v \ll 1$

Next, we compute the gap $\Delta_{0,\boldsymbol{q}}$. We have

$$\sinh^{-1}\left(\frac{\hbar\omega_{\rm D}+\eta_{\boldsymbol{q}}}{\Delta_{0,\boldsymbol{q}}}\right) = \frac{2}{g(\varepsilon_{\rm F})\,v} + \sinh^{-1}\left(\frac{\eta_{\boldsymbol{q}}}{\Delta_{0,\boldsymbol{q}}}\right)\quad . \tag{12.133}$$

Assuming $\eta_{\boldsymbol{q}} \ll \Delta_{0,\boldsymbol{q}}$, we obtain

$$\Delta_{0,\boldsymbol{q}} = \Delta_0 - \eta_{\boldsymbol{q}} = \Delta_0 - \frac{\hbar^2\boldsymbol{q}^2}{8m^*}\quad . \tag{12.134}$$

12.11.2 Supercurrent

We assume a quadratic dispersion $\varepsilon_{\boldsymbol{k}} = \hbar^2\boldsymbol{k}^2/2m^*$, so $\boldsymbol{v}_{\boldsymbol{k}} = \hbar\boldsymbol{k}/m^*$. The current density is then given by

$$\boldsymbol{j} = \frac{2e\hbar}{m^*V}\sum_{\boldsymbol{k}}\big(\boldsymbol{k}+\tfrac{1}{2}\boldsymbol{q}\big)\big\langle\, c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\,c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\big\rangle = \frac{ne\hbar}{2m^*}\,\boldsymbol{q} + \frac{2e\hbar}{m^*V}\sum_{\boldsymbol{k}}\boldsymbol{k}\,\big\langle\, c^\dagger_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\,c_{\boldsymbol{k}+\frac{1}{2}\boldsymbol{q}\uparrow}\big\rangle\quad , \tag{12.135}$$

where $n = N/V$ is the total electron number density. Appealing to Eqn. 12.128, we have

$$\boldsymbol{j} = \frac{e\hbar}{m^*V}\sum_{\boldsymbol{k}}\boldsymbol{k}\left\{\Big[1+f(E_{\boldsymbol{k},\boldsymbol{q}}+\nu_{\boldsymbol{k},\boldsymbol{q}})-f(E_{\boldsymbol{k},\boldsymbol{q}}-\nu_{\boldsymbol{k},\boldsymbol{q}})\Big]\right. \tag{12.136}$$
$$\left. + \frac{\omega_{\boldsymbol{k},\boldsymbol{q}}}{E_{\boldsymbol{k},\boldsymbol{q}}}\Big[f(E_{\boldsymbol{k},\boldsymbol{q}}+\nu_{\boldsymbol{k},\boldsymbol{q}})+f(E_{\boldsymbol{k},\boldsymbol{q}}-\nu_{\boldsymbol{k},\boldsymbol{q}})-1\Big]\right\} + \frac{ne\hbar}{2m^*}\,\boldsymbol{q}\quad .$$

We now write $f(E_{\boldsymbol{k},\boldsymbol{q}} \pm \nu_{\boldsymbol{k},\boldsymbol{q}}) = f(E_{\boldsymbol{k},\boldsymbol{q}}) \pm f'(E_{\boldsymbol{k},\boldsymbol{q}})\,\nu_{\boldsymbol{k},\boldsymbol{q}} + \ldots$, obtaining

$$\boldsymbol{j} = \frac{e\hbar}{m^* V}\sum_{\boldsymbol{k}} \boldsymbol{k}\left[1 + 2\,\nu_{\boldsymbol{k},\boldsymbol{q}}\,f'(E_{\boldsymbol{k},\boldsymbol{q}})\right] + \frac{ne\hbar}{2m^*}\,\boldsymbol{q} \quad . \tag{12.137}$$

For the ballistic dispersion, $\nu_{\boldsymbol{k},\boldsymbol{q}} = \hbar^2\boldsymbol{k}\cdot\boldsymbol{q}/2m^*$, so

$$\begin{aligned}\boldsymbol{j} - \frac{ne\hbar}{2m^*}\,\boldsymbol{q} &= \frac{e\hbar}{m^* V}\,\frac{\hbar^2}{m^*}\sum_{\boldsymbol{k}}(\boldsymbol{q}\cdot\boldsymbol{k})\,\boldsymbol{k}\,f'(E_{\boldsymbol{k},\boldsymbol{q}}) \\ &= \frac{e\hbar^3}{3m^{*2}V}\boldsymbol{q}\sum_{\boldsymbol{k}}\boldsymbol{k}^2\,f'(E_{\boldsymbol{k},\boldsymbol{q}}) \simeq \frac{ne\hbar}{m^*}\,\boldsymbol{q}\int_0^\infty d\xi\,\frac{\partial f}{\partial E} \quad ,\end{aligned} \tag{12.138}$$

where we have set $\boldsymbol{k}^2 = k_{\rm F}^2$ inside the sum, since it is only appreciable in the vicinity of $k = k_{\rm F}$, and we have invoked $g(\varepsilon_{\rm F}) = m^* k_{\rm F}/\pi^2\hbar^2$ and $n = k_{\rm F}^3/3\pi^2$. Thus,

$$\boldsymbol{j} = \frac{ne\hbar}{2m^*}\left(1 + 2\int_0^\infty d\xi\,\frac{\partial f}{\partial E}\right)\boldsymbol{q} \equiv \frac{n_{\rm s}(T)\,e\hbar\boldsymbol{q}}{2m^*} \quad . \tag{12.139}$$

This defines the superfluid density,

$$n_{\rm s}(T) = n\left(1 + 2\int_0^\infty d\xi\,\frac{\partial f}{\partial E}\right) \quad . \tag{12.140}$$

Note that the second term in round brackets on the RHS is always negative. Thus, at $T = 0$, we have $n_{\rm s} = n$, but at $T = T_{\rm c}$, where the gap vanishes, we find $n_{\rm s}(T_{\rm c}) = 0$, since $E = |\xi|$ and $f(0) = \frac{1}{2}$. We may write $n_{\rm s}(T) = n - n_{\rm n}(T)$, where $n_{\rm n}(T) = n\,\mathcal{Y}(T)$ is the normal fluid density.

Ginzburg-Landau theory

We may now expand the free energy near $T = T_{\rm c}$ at finite condensate $\boldsymbol{q}$. We will only quote the result. One finds

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = \tilde{a}(T)\,|\Delta|^2 + \tfrac{1}{2}\,\tilde{b}(T)\,|\Delta|^4 + \frac{n\,\tilde{b}(T)}{g(\varepsilon_{\rm F})}\,\frac{\hbar^2\boldsymbol{q}^2}{2m^*}\,|\Delta|^2 \quad , \tag{12.141}$$

where the Landau coefficients $\tilde{a}(T)$ and $\tilde{b}(T)$ are given in Eqn. 12.108. Identifying the last term as $\tilde{K}\,|\boldsymbol{\nabla}\Delta|^2$, where $\tilde{K}$ is the stiffness, we have

$$\tilde{K} = \frac{\hbar^2}{2m^*}\,\frac{n\,\tilde{b}(T)}{g(\varepsilon_{\rm F})} \quad . \tag{12.142}$$

12.12 Effect of Repulsive Interactions

Let's modify our model in Eqns. 12.71 and 12.72 and write

$$V_{\boldsymbol{k},\boldsymbol{k}'} = \begin{cases} (v_{\rm C} - v_{\rm p})/V & \text{if } |\xi_{\boldsymbol{k}}| < \hbar\omega_{\rm D} \text{ and } |\xi_{\boldsymbol{k}'}| < \hbar\omega_{\rm D} \\ v_{\rm C}/V & \text{otherwise} \end{cases} \tag{12.143}$$

and

$$\Delta_{\boldsymbol{k}} = \begin{cases} \Delta_0 & \text{if } |\xi_{\boldsymbol{k}}| < \hbar\omega_{\rm D} \\ \Delta_1 & \text{otherwise} \end{cases} . \tag{12.144}$$

Here $-v_{\rm p} < 0$ is the attractive interaction mediated by phonons, while $v_{\rm C} > 0$ is the Coulomb repulsion. We presume $v_{\rm p} > v_{\rm C}$ so that there is a net attraction at low energies, although below we will show this assumption is overly pessimistic. We take $\Delta_{0,1}$ both to be real.

At $T = 0$, the gap equation then gives

$$\begin{aligned} \Delta_0 &= \tfrac{1}{2}\, g(\varepsilon_{\rm F})\,(v_{\rm p} - v_{\rm C}) \int\limits_0^{\hbar\omega_{\rm D}} d\xi\, \frac{\Delta_0}{\sqrt{\xi^2 + \Delta_0^2}} - \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C} \int\limits_{\hbar\omega_{\rm D}}^{B} d\xi\, \frac{\Delta_1}{\sqrt{\xi^2 + \Delta_1^2}} \\ \Delta_1 &= -\tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C} \int\limits_0^{\hbar\omega_{\rm D}} d\xi\, \frac{\Delta_0}{\sqrt{\xi^2 + \Delta_0^2}} - \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C} \int\limits_{\hbar\omega_{\rm D}}^{B} d\xi\, \frac{\Delta_1}{\sqrt{\xi^2 + \Delta_1^2}} \quad , \end{aligned} \tag{12.145}$$

where $\hbar\omega_{\rm D}$ is once again the Debye energy, and B is the full electronic bandwidth. Performing the integrals, and assuming $\Delta_{0,1} \ll \hbar\omega_{\rm D} \ll B$, we obtain

$$\begin{aligned} \Delta_0 &= \tfrac{1}{2}\, g(\varepsilon_{\rm F})\,(v_{\rm p} - v_{\rm C})\, \Delta_0\, \ln\!\left(\frac{2\hbar\omega_{\rm D}}{\Delta_0}\right) - \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C}\, \Delta_1 \ln\!\left(\frac{B}{\hbar\omega_{\rm D}}\right) \\ \Delta_1 &= -\tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C}\, \Delta_0\, \ln\!\left(\frac{2\hbar\omega_{\rm D}}{\Delta_0}\right) - \tfrac{1}{2}\, g(\varepsilon_{\rm F})\, v_{\rm C}\, \Delta_1 \ln\!\left(\frac{B}{\hbar\omega_{\rm D}}\right) \quad . \end{aligned} \tag{12.146}$$

The second of these equations gives

$$\Delta_1 = -\frac{\frac{1}{2} g(\varepsilon_{\rm F})\, v_{\rm C} \ln(2\hbar\omega_{\rm D}/\Delta_0)}{1 + \frac{1}{2} g(\varepsilon_{\rm F})\, v_{\rm C}\, \ln(B/\hbar\omega_{\rm D})}\, \Delta_0 \quad . \tag{12.147}$$

Inserting this into the first equation then results in

$$\frac{2}{g(\varepsilon_{\rm F})\, v_{\rm p}} = \ln\!\left(\frac{2\hbar\omega_{\rm D}}{\Delta_0}\right) \cdot \left\{ 1 - \frac{v_{\rm C}}{v_{\rm p}} \cdot \frac{1}{1 + \frac{1}{2}\, g(\varepsilon_{\rm F})\, \ln(B/\hbar\omega_{\rm D})} \right\} \quad . \tag{12.148}$$

This has a solution only if the attractive potential $v_{\rm p}$ is greater than the repulsive factor $v_{\rm C}\Big/\big[1+\frac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm C}\ln(B/\hbar\omega_{\rm D})\big]$. Note that it is a renormalized and reduced value of the bare repulsion $v_{\rm C}$ which enters here. Thus, it is possible to have

$$v_{\rm C} > v_{\rm p} > \frac{v_{\rm C}}{1+\frac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm C}\ln(B/\hbar\omega_{\rm D})} \quad , \tag{12.149}$$

so that $v_{\rm C} > v_{\rm p}$ and the potential is *always* repulsive, yet still the system is superconducting!

Working at finite temperature, we must include factors of $\tanh\Big(\frac{1}{2}\beta\sqrt{\xi^2+\Delta_{0,1}^2}\Big)$ inside the appropriate integrands in Eqn. 12.145, with $\beta = 1/k_{\rm B}T$. The equation for $T_{\rm c}$ is then obtained by examining the limit $\Delta_{0,1}\to 0$, with the ratio $r\equiv\Delta_1/\Delta_0$ finite. We then have

$$\begin{aligned}\frac{2}{g(\varepsilon_{\rm F})} &= (v_{\rm p}-v_{\rm C})\int\limits_0^{\widetilde{\Omega}} ds\; s^{-1}\,\tanh(s) - r\,v_{\rm C}\int\limits_{\widetilde{\Omega}}^{\widetilde{B}} ds\; s^{-1}\,\tanh(s) \\ \frac{2}{g(\varepsilon_{\rm F})} &= -r^{-1}\,v_{\rm C}\int\limits_0^{\widetilde{\Omega}} ds\; s^{-1}\,\tanh(s) - v_{\rm C}\int\limits_{\widetilde{\Omega}}^{\widetilde{B}} ds\; s^{-1}\,\tanh(s) \quad ,\end{aligned} \tag{12.150}$$

where $\widetilde{\Omega}\equiv\hbar\omega_{\rm D}/2k_{\rm B}T_{\rm c}$ and $\widetilde{B}\equiv B/2k_{\rm B}T_{\rm c}$. We now use

$$\int\limits_0^{\Lambda} ds\; s^{-1}\,\tanh(s) = \ln\Lambda + \ln\big(\overbrace{4e^{\rm C}/\pi}^{\approx 2.268}\big) + \mathcal{O}\big(e^{-\Lambda}\big) \tag{12.151}$$

to obtain

$$\frac{2}{g(\varepsilon_{\rm F})\,v_{\rm p}} = \ln\left(\frac{1.134\,\hbar\omega_{\rm D}}{k_{\rm B}T_{\rm c}}\right)\cdot\left\{1-\frac{v_{\rm C}}{v_{\rm p}}\cdot\frac{1}{1+\frac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm C}\,\ln(B/\hbar\omega_{\rm D})}\right\} \quad . \tag{12.152}$$

Comparing with Eqn. 12.148, we see that once again we have $2\Delta_0(T=0) = 3.52\,k_{\rm B}T_{\rm c}$. Note, however, that

$$k_{\rm B}T_{\rm c} = 1.134\,\hbar\omega_{\rm D}\,\exp\left(-\frac{2}{g(\varepsilon_{\rm F})\,v_{\rm eff}}\right) \quad , \tag{12.153}$$

where

$$v_{\rm eff} = v_{\rm p} - \frac{v_{\rm C}}{1+\frac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm C}\,\ln(B/\hbar\omega_{\rm D})} \quad . \tag{12.154}$$

It is customary to define

$$\lambda\equiv\tfrac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm p} \qquad , \qquad \mu\equiv\tfrac{1}{2}\,g(\varepsilon_{\rm F})\,v_{\rm C} \qquad , \qquad \mu^*\equiv\frac{\mu}{1+\mu\ln(B/\hbar\omega_{\rm D})} \quad , \tag{12.155}$$

so that

$$k_{\rm B}T_{\rm c} = 1.134\,\hbar\omega_{\rm D}\,e^{-1/(\lambda-\mu^*)} \quad , \quad \Delta_0 = 2\hbar\omega_{\rm D}\,e^{-1/(\lambda-\mu^*)} \quad , \quad \Delta_1 = -\frac{\mu^*\Delta_0}{\lambda-\mu^*} \quad . \tag{12.156}$$

Since μ^* depends on $\omega_{\rm D}$, the isotope effect is modified:

$$\delta \ln T_{\rm c} = \delta \ln \omega_{\rm D} \cdot \left\{ 1 - \frac{\mu^2}{1+\mu\ln(B/\hbar\omega_{\rm D})} \right\} \quad . \tag{12.157}$$

12.13 Appendix I : General Variational Formulation

We consider a more general grand canonical Hamiltonian of the form

$$\hat{K} = \sum_{\boldsymbol{k}\sigma} (\varepsilon_{\boldsymbol{k}} - \mu)\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}\sigma} + \frac{1}{2V} \sum_{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}} \sum_{\sigma,\sigma'} \hat{u}_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})\, c^\dagger_{\boldsymbol{k}+\boldsymbol{q}\,\sigma}\, c^\dagger_{\boldsymbol{p}-\boldsymbol{q}\,\sigma'}\, c_{\boldsymbol{p}\,\sigma'}\, c_{\boldsymbol{k}\,\sigma} \quad . \tag{12.158}$$

In order that the Hamiltonian be Hermitian, we may require, without loss of generality,

$$\hat{u}^*_{\sigma\sigma'}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}) = \hat{u}_{\sigma\sigma'}(\boldsymbol{k}+\boldsymbol{q}\,,\,\boldsymbol{p}-\boldsymbol{q}\,,\,-\boldsymbol{q}) \quad . \tag{12.159}$$

In addition, spin rotation invariance says that $\hat{u}_{\uparrow\uparrow}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}) = \hat{u}_{\downarrow\downarrow}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})$ and $\hat{u}_{\uparrow\downarrow}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}) = \hat{u}_{\downarrow\uparrow}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})$. We now take the thermal expectation of $\hat{K}$ using a density matrix derived from the BCS Hamiltonian,

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k}} \begin{pmatrix} c^\dagger_{\boldsymbol{k}\uparrow} & c_{-\boldsymbol{k}\downarrow} \end{pmatrix} \begin{pmatrix} \xi_{\boldsymbol{k}} & \Delta_{\boldsymbol{k}} \\ \Delta^*_{\boldsymbol{k}} & -\xi_{\boldsymbol{k}} \end{pmatrix} \begin{pmatrix} c_{\boldsymbol{k}\uparrow} \\ c^\dagger_{-\boldsymbol{k}\downarrow} \end{pmatrix} + K_0 \quad . \tag{12.160}$$

The energy shift K_0 will not be important in our subsequent analysis. From the BCS Hamiltonian,

$$\langle c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma'} \rangle = n_{\boldsymbol{k}}\, \delta_{\boldsymbol{k},\boldsymbol{k}'}\, \delta_{\sigma\sigma'} \qquad , \qquad \langle c^\dagger_{\boldsymbol{k}\sigma}\, c^\dagger_{\boldsymbol{k}'\sigma'} \rangle = \Psi^*_{\boldsymbol{k}}\, \delta_{\boldsymbol{k}',-\boldsymbol{k}}\, \varepsilon_{\sigma\sigma'} \quad , \tag{12.161}$$

where $\varepsilon_{\sigma\sigma'} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. We don't yet need the detailed forms of $n_{\boldsymbol{k}}$ and $\Psi_{\boldsymbol{k}}$ either. Using Wick's theorem, we find

$$\langle \hat{K} \rangle = \sum_{\boldsymbol{k}} 2(\varepsilon_{\boldsymbol{k}} - \mu)\, n_{\boldsymbol{k}} + \sum_{\boldsymbol{k},\boldsymbol{k}'} W_{\boldsymbol{k},\boldsymbol{k}'}\, n_{\boldsymbol{k}}\, n_{\boldsymbol{k}'} - \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \Psi^*_{\boldsymbol{k}}\, \Psi_{\boldsymbol{k}'} \quad , \tag{12.162}$$

where

$$\begin{aligned} W_{\boldsymbol{k},\boldsymbol{k}'} &= \frac{1}{V} \Big\{ \hat{u}_{\uparrow\uparrow}(\boldsymbol{k},\boldsymbol{k}',0) + \hat{u}_{\uparrow\downarrow}(\boldsymbol{k},\boldsymbol{k}',0) - \hat{u}_{\uparrow\uparrow}(\boldsymbol{k},\boldsymbol{k}',\boldsymbol{k}'-\boldsymbol{k}) \Big\} \\ V_{\boldsymbol{k},\boldsymbol{k}'} &= -\frac{1}{V}\, \hat{u}_{\uparrow\downarrow}(\boldsymbol{k}',-\boldsymbol{k}',\boldsymbol{k}-\boldsymbol{k}') \quad . \end{aligned} \tag{12.163}$$

We may assume $W_{\boldsymbol{k},\boldsymbol{k}'}$ is real and symmetric, and $V_{\boldsymbol{k},\boldsymbol{k}'}$ is Hermitian.

Now let's vary $\langle \hat{K} \rangle$ by changing the distribution. We have

$$\delta\langle \hat{K} \rangle = 2 \sum_{\boldsymbol{k}} \left(\varepsilon_{\boldsymbol{k}} - \mu + \sum_{\boldsymbol{k}'} W_{\boldsymbol{k},\boldsymbol{k}'}\, n_{\boldsymbol{k}'} \right) \delta n_{\boldsymbol{k}} + \sum_{\boldsymbol{k},\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'} \Big(\Psi^*_{\boldsymbol{k}}\, \delta\Psi_{\boldsymbol{k}'} + \delta\Psi^*_{\boldsymbol{k}}\, \Psi_{\boldsymbol{k}'} \Big) \quad . \tag{12.164}$$

On the other hand,

$$\delta\langle \hat{K}_{\rm BCS} \rangle = 2 \sum_{\boldsymbol{k}} \Big(\xi_{\boldsymbol{k}}\, \delta n_{\boldsymbol{k}} + \Delta_{\boldsymbol{k}}\, \delta\Psi^*_{\boldsymbol{k}} + \Delta^*_{\boldsymbol{k}}\, \delta\Psi_{\boldsymbol{k}} \Big) \quad . \tag{12.165}$$

Setting these variations to be equal, we obtain

$$\begin{aligned} \xi_{\boldsymbol{k}} &= \varepsilon_{\boldsymbol{k}} - \mu + \sum_{\boldsymbol{k}'} W_{\boldsymbol{k},\boldsymbol{k}'}\, n_{\boldsymbol{k}'} \\ &= \varepsilon_{\boldsymbol{k}} - \mu + \sum_{\boldsymbol{k}'} W_{\boldsymbol{k},\boldsymbol{k}'} \left[\frac{1}{2} - \frac{\xi_{\boldsymbol{k}'}}{2E_{\boldsymbol{k}'}} \tanh\big(\tfrac{1}{2}\beta E_{\boldsymbol{k}'}\big) \right] \end{aligned} \tag{12.166}$$

and

$$\Delta_{\boldsymbol{k}} = \sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \Psi_{\boldsymbol{k}'} = -\sum_{\boldsymbol{k}'} V_{\boldsymbol{k},\boldsymbol{k}'}\, \frac{\Delta_{\boldsymbol{k}'}}{2E_{\boldsymbol{k}'}}\, \tanh\big(\tfrac{1}{2}\beta E_{\boldsymbol{k}'}\big) \quad . \tag{12.167}$$

These are to be regarded as self-consistent equations for $\xi_{\boldsymbol{k}}$ and $\Delta_{\boldsymbol{k}}$.

12.14 Appendix II : Superconducting Free Energy

We start with the Landau free energy difference from Eqn. 12.100,

$$\frac{\Omega_{\rm s} - \Omega_{\rm n}}{V} = -\tfrac{1}{4}\, g(\varepsilon_{\rm F})\, \Delta^2 \left\{ 1 + 2\ln\left(\frac{\Delta_0}{\Delta}\right) - \left(\frac{\Delta}{2\hbar\omega_{\rm D}}\right)^2 + \mathcal{O}\big(\Delta^4\big) \right\} \tag{12.168}$$

$$- 2\, g(\varepsilon_{\rm F})\, \Delta^2\, I(\delta) + \tfrac{1}{6}\pi^2\, g(\varepsilon_{\rm F})\, (k_{\rm B}T)^2 \quad ,$$

where

$$I(\delta) = \frac{1}{\delta} \int_0^\infty ds\ \ln\left(1 + e^{-\delta\sqrt{1+s^2}}\right) . \tag{12.169}$$

We now proceed to examine the integral $I(\delta)$ in the limits $\delta \to \infty$ (*i.e.* $T \to 0^+$) and $\delta \to 0^+$ (*i.e.* $T \to T_{\rm c}^-$, where $\Delta \to 0$).

When $\delta \to \infty$, we may safely expand the logarithm in a Taylor series, and

$$I(\delta) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n\delta}\, K_1(n\delta) \quad , \tag{12.170}$$

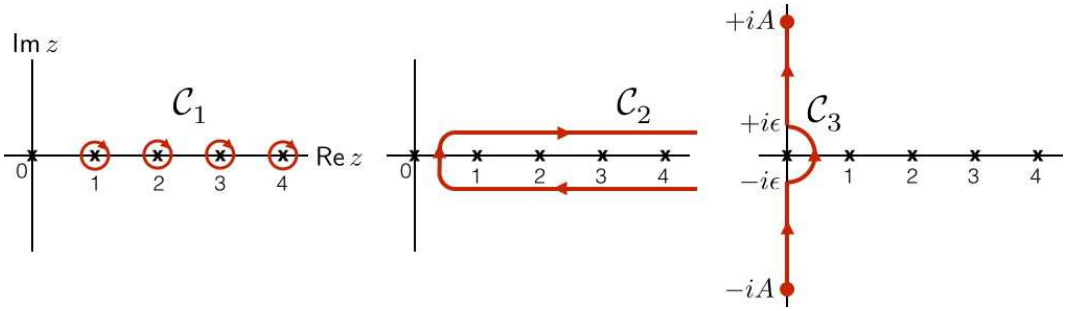


Figure 12.7: Contours for complex integration for calculating $I(\delta)$ as described in the text.

where $K_1(\delta)$ is the modified Bessel function, also called the MacDonald function. Asymptotically, we have[11]

$$K_1(z) = \left(\frac{\pi}{2z}\right)^{1/2} e^{-z} \cdot \Big\{1 + \mathcal{O}\big(z^{-1}\big)\Big\} \quad . \tag{12.171}$$

We may then retain only the $n = 1$ term to leading nontrivial order. This immediately yields the expression in Eqn. 12.101.

The limit $\delta \to 0$ is much more subtle. We begin by integrating once by parts, to obtain

$$I(\delta) = \int_1^\infty dt\, \frac{\sqrt{t^2-1}}{e^{\delta t}+1} \quad . \tag{12.172}$$

We now appeal to the tender mercies of `Mathematica`. Alas, this avenue is to no avail, for the program gags when asked to expand $I(\delta)$ for small δ. We need something better than `Mathematica`. We need Professor Michael Fogler.

Fogler says[12]: start by writing Eqn. 12.170 in the form

$$I(\delta) = \sum_{n=1}^\infty \frac{(-1)^{n-1}}{n\delta}\, K_1(n\delta) = +\int_{\mathcal{C}_1} \frac{dz}{2\pi i}\, \frac{\pi}{\sin \pi z}\, \frac{K_1(\delta z)}{\delta z} \quad . \tag{12.173}$$

The initial contour $\mathcal{C}_1$ consists of a disjoint set of small loops circling the points $z = \pi n$, where $n \in \mathbb{Z}_+$. Note that the sense of integration is clockwise rather than counterclockwise. This accords with an overall minus sign in the RHS above, because the residues contain a factor of $\cos(\pi n) = (-1)^n$ rather than the desired $(-1)^{n-1}$. Following Fig. 12.7, the contour may now be deformed into $\mathcal{C}_2$, and then into $\mathcal{C}_3$. Contour $\mathcal{C}_3$ lies along the imaginary z axis, aside from a small semicircle of radius $\epsilon \to 0$ avoiding the origin, and terminates at $z = \pm iA$. We will later take $A \to \infty$, but for the moment we consider $1 \ll A \ll \delta^{-1}$. So long as $A \gg 1$, the denominator $\sin \pi z = i \sinh \pi u$, with $z = iu$, will be exponentially large at $u = \pm A$, so we are safe in making

[11] See, *e.g.*, the *NIST Handbook of Mathematical Functions*, §10.25.
[12] M. Fogler, private communications.

this initial truncation. We demand $A \ll \delta^{-1}$, however, which means $|\delta z| \ll 1$ everywhere along $\mathcal{C}_3$. *This allows us to expand $K_1(\delta z)$ for small values of the argument.* One has

$$\frac{K_1(w)}{w} = \frac{1}{w^2} + \tfrac{1}{2}\ln w\left(1 + \tfrac{1}{8}w^2 + \tfrac{1}{192}w^4 + \ldots\right) + \left(\mathrm{C} - \ln 2 - \tfrac{1}{2}\right) + \tfrac{1}{16}\left(C - \ln 2 - \tfrac{5}{4}\right)w^2 + \tfrac{1}{384}\left(C - \ln 2 - \tfrac{5}{3}\right)w^4 + \ldots \quad , \tag{12.174}$$

where $\mathrm{C} \simeq 0.577216$ is the Euler-Mascheroni constant. The integral is then given by

$$I(\delta) = \int\limits_{\epsilon}^{A} \frac{du}{2\pi i}\,\frac{\pi}{\sinh \pi u}\left[\frac{K_1(i\delta u)}{i\delta u} - \frac{K_1(-i\delta u)}{-i\delta u}\right] + \int\limits_{-\pi/2}^{\pi/2} \frac{d\theta}{2\pi}\,\frac{\pi\epsilon\, e^{i\theta}}{\sin\left(\pi\epsilon\, e^{i\theta}\right)}\,\frac{K_1\left(\delta\epsilon\, e^{i\theta}\right)}{\delta\epsilon\, e^{i\theta}} \quad . \tag{12.175}$$

Using the above expression for $K_1(w)/w$, we have

$$\frac{K_1(i\delta u)}{i\delta u} - \frac{K_1(-i\delta u)}{-i\delta u} = \frac{i\pi}{2}\left(1 - \tfrac{1}{8}\delta^2 u^2 + \tfrac{1}{192}\delta^4 u^4 + \ldots\right) \quad . \tag{12.176}$$

At this point, we may take $A \to \infty$. The integral along the two straight parts of the $\mathcal{C}_3$ contour is then

$$\begin{aligned} I_1(\delta) &= \tfrac{1}{4}\pi \int\limits_{\epsilon}^{\infty} \frac{du}{\sinh \pi u}\left(1 - \tfrac{1}{8}\delta^2 u^2 + \tfrac{1}{192}\delta^4 u^4 + \ldots\right) \\ &= -\tfrac{1}{4}\ln\tanh\left(\tfrac{1}{2}\pi\epsilon\right) - \frac{7\,\zeta(3)}{64\,\pi^2}\,\delta^2 + \frac{31\,\zeta(5)}{512\,\pi^4}\,\delta^4 + \mathcal{O}\left(\delta^6\right) \quad . \end{aligned} \tag{12.177}$$

The integral around the semicircle is

$$\begin{aligned} I_2(\delta) &= \int\limits_{-\pi/2}^{\pi/2} \frac{d\theta}{2\pi}\,\frac{1}{1 - \frac{1}{6}\pi^2\epsilon^2\, e^{2i\theta}}\left\{\frac{1}{\delta^2\epsilon^2\, e^{2i\theta}} + \tfrac{1}{2}\ln\left(\delta\epsilon\, e^{i\theta}\right) + \tfrac{1}{2}(\mathrm{C} - \ln 2 - \tfrac{1}{2}) + \ldots\right\} \\ &= \int\limits_{-\pi/2}^{\pi/2} \frac{d\theta}{2\pi}\left(1 + \tfrac{1}{6}\pi^2\epsilon^2\, e^{2i\theta} + \ldots\right)\left\{\frac{e^{-2i\theta}}{\delta^2\epsilon^2} + \tfrac{1}{2}\ln(\delta\epsilon) + \tfrac{i}{2}\theta + \tfrac{1}{2}(\mathrm{C} - \ln 2 - \tfrac{1}{2}) + \ldots\right\} \\ &= \frac{\pi^2}{12\,\delta^2} + \tfrac{1}{4}\ln\delta + \tfrac{1}{4}\ln\epsilon + \tfrac{1}{4}(\mathrm{C} - \ln 2 - \tfrac{1}{2}) + \mathcal{O}\left(\epsilon^2\right) \quad . \end{aligned} \tag{12.178}$$

We now add the results to obtain $I(\delta) = I_1(\delta) + I_2(\delta)$. Note that there are divergent pieces, each proportional to $\ln\epsilon$, which cancel as a result of this addition. The final result is

$$I(\delta) = \frac{\pi^2}{12\,\delta^2} + \tfrac{1}{4}\ln\left(\frac{2\delta}{\pi}\right) + \tfrac{1}{4}(\mathrm{C} - \ln 2 - \tfrac{1}{2}) - \frac{7\,\zeta(3)}{64\,\pi^2}\,\delta^2 + \frac{31\,\zeta(5)}{512\,\pi^4}\,\delta^4 + \mathcal{O}\left(\delta^6\right) \quad . \tag{12.179}$$

Inserting this result in Eqn. 12.168 above, we thereby recover Eqn. 12.106.

Chapter 13

Applications of BCS Theory

13.1 Quantum XY Model for Granular Superconductors

Consider a set of superconducting grains, each of which is large enough to be modeled by BCS theory, but small enough that the self-capacitance (*i.e.* Coulomb interaction) cannot be neglected. The Coulomb energy of the $j^{\rm th}$ grain is written as

$$\hat{U}_j = \frac{2e^2}{C_j}\big(\hat{M}_j - \bar{M}_j\big)^2 \quad , \tag{13.1}$$

where $\hat{M}_j$ is the operator which counts the number of Cooper pairs on grain j, and $\bar{M}_j$ is the mean number of pairs in equilibrium, which is given by half the total ionic charge on the grain. The capacitance C_j is a geometrical quantity which is proportional to the radius of the grain, assuming the grain is roughly spherical. For very large grains, the Coulomb interaction is negligible. It should be stressed that here we are accounting for only the long wavelength part of the Coulomb interaction, which is proportional to $4\pi\big|\delta\hat{\rho}(q_{\rm min})\big|^2/q_{\rm min}^2$, where $q_{\rm min} \sim 1/R_j$ is the inverse grain size. The remaining part of the Coulomb interaction is included in the BCS part of the Hamiltonian for each grain.

We assume that $\hat{K}_{{\rm BCS},j}$ describes a simple s-wave superconductor with gap $\Delta_j = |\Delta_j|\, e^{i\phi_j}$. We saw in chapter 3 how ϕ_j is conjugate to the Cooper pair number operator $\hat{M}_j$, with

$$\hat{M}_j = \frac{1}{i}\,\frac{\partial}{\partial\phi_j} \quad . \tag{13.2}$$

The operator which adds one Cooper pair to grain j is therefore $e^{i\phi_j}$, because

$$\hat{M}_j\, e^{i\phi_j} = e^{i\phi_j}\,(\hat{M}_j + 1) \quad . \tag{13.3}$$

Thus, accounting for the hopping of Cooper pairs between neighboring grains, the effective Hamiltonian for a granular superconductor should be given by

$$\hat{H}_{\rm gr} = -\tfrac{1}{2}\sum_{i,j} J_{ij}\big(e^{i\phi_i}\,e^{-i\phi_j} + e^{-i\phi_i}\,e^{i\phi_j}\big) + \sum_i \frac{2e^2}{C_j}\big(\hat{M}_j - \bar{M}_j\big)^2 \quad , \tag{13.4}$$

where J_{ij} is the hopping matrix element for the Cooper pairs, here assumed to be real.

Before we calculate J_{ij}, note that we can eliminate the constants $\bar{M}_i$ from the Hamiltonian via the unitary transformation $\hat{H}_{\rm gr} \to \hat{H}'_{\rm gr} = V^\dagger \hat{H}_{\rm gr} V$, where $V = \prod_j e^{i\,[\bar{M}_j]\,\phi_j}$, where $[\bar{M}_j]$ is defined as the integer nearest to $\bar{M}_j$. The difference, $\delta\bar{M}_j = \bar{M}_j - [\bar{M}_j]$, cannot be removed. This transformation commutes with the hopping part of $\hat{H}_{\rm gr}$, so, after dropping the prime on $\hat{H}'_{\rm gr}$, we are left with

$$\hat{H}_{\rm gr} = \sum_j \frac{2e^2}{C_j}\left(\frac{1}{i}\,\frac{\partial}{\partial\phi_j} - \delta\bar{M}_j\right)^2 - \sum_{i,j} J_{ij}\cos(\phi_i - \phi_j) \quad . \tag{13.5}$$

In the presence of an external magnetic field,

$$\hat{H}_{\rm gr} = \sum_j \frac{2e^2}{C_j}\left(\frac{1}{i}\,\frac{\partial}{\partial\phi_j} - \delta\bar{M}_j\right)^2 - \sum_{i,j} J_{ij}\cos(\phi_i - \phi_j - \mathcal{A}_{ij}) \quad , \tag{13.6}$$

where

$$\mathcal{A}_{ij} = \frac{2e}{\hbar c}\int_{\boldsymbol{R}_i}^{\boldsymbol{R}_j} d\boldsymbol{l}\cdot\boldsymbol{A} \tag{13.7}$$

is a lattice vector potential, with $\boldsymbol{R}_i$ the position of grain i.

13.1.1 No disorder

In a perfect lattice of identical grains, with $J_{ij} = J$ for nearest neighbors, $\delta\bar{M}_j = 0$ and $2e^2/C_j = U$ for all j, we have

$$\hat{H}_{\rm gr} = -U\sum_i \frac{\partial^2}{\partial\phi_i^2} - 2J\sum_{\langle ij\rangle}\cos(\phi_i - \phi_j) \quad , \tag{13.8}$$

where $\langle ij\rangle$ indicates a nearest neighbor pair. This model, known as the *quantum rotor model*, features competing interactions. The potential energy, proportional to U, favors each grain being in a state $\psi(\phi_i) = 1$, corresponding to $M = 0$, which minimizes the Coulomb interaction. However, it does a poor job with the hopping, since $\big\langle\cos(\phi_i - \phi_j)\big\rangle = 0$ in this state. The kinetic (hopping) energy, proportional to J, favors that all grains be coherent with $\phi_i = \alpha$ for all i, where α is a constant. This state has significant local charge fluctuations which cost Coulomb

energy – an infinite amount, in fact! Some sort of compromise must be reached. One important issue is whether the ground state exhibits a finite order parameter $\langle e^{i\phi_i}\rangle$.

The model has been simulated numerically using a cluster Monte Carlo algorithm[1], and is known to exhibit a quantum phase transition between superfluid and insulating states at a critical value of J/U. The superfluid state is that in which $\langle e^{i\phi_i}\rangle \neq 0$.

13.1.2 Self-consistent harmonic approximation

The self-consistent harmonic approximation (SCHA) is a variational approach in which we approximate the ground state wavefunction as a Gaussian function of the many phase variables $\{\phi_i\}$. Specifically, we write

$$\Psi[\phi] = \mathcal{C}\ \exp\bigl(-\tfrac{1}{4}\,A_{ij}\,\phi_i\,\phi_j\bigr) \quad , \tag{13.9}$$

where $\mathcal{C}$ is a normalization constant. The matrix elements A_{ij} is assumed to be a function of the separation $\boldsymbol{R}_i - \boldsymbol{R}_j$, where $\boldsymbol{R}_i$ is the position of lattice site i. We define the *generating function*

$$Z[\mathcal{J}] = \int D\phi\,\bigl|\Psi[\phi]\bigr|^2\,e^{-\mathcal{J}_i\,\phi_i} = Z[0]\ \exp\bigl(\tfrac{1}{2}\,\mathcal{J}_i\,A^{-1}_{ij}\,\mathcal{J}_j\bigr) \quad . \tag{13.10}$$

Here $\mathcal{J}_i$ is a *source field* with respect to which we differentiate in order to compute correlation functions, as we shall see. Here $D\phi = \prod_i d\phi_i$, and all the phase variables are integrated over the $\phi_i \in (-\infty, +\infty)$. Right away we see something is fishy, since in the original model there is a periodicity under $\phi_i \to \phi_i + 2\pi$ at each site. The individual basis functions are $\psi_n(\phi) = e^{in\phi}$, corresponding to $M = n$ Cooper pairs. Taking linear combinations of these basis states preserves the 2π periodicity, but this is not present in our variational wavefunction. Nevertheless, we can extract some useful physics using the SCHA.

The first order of business is to compute the correlator

$$\langle\,\Psi\,|\,\phi_i\,\phi_j\,|\,\Psi\,\rangle = \frac{1}{Z[0]}\,\frac{\partial^2 Z[\mathcal{J}]}{\partial\mathcal{J}_i\,\partial\mathcal{J}_j}\bigg|_{\mathcal{J}=0} = A^{-1}_{ij} \quad . \tag{13.11}$$

This means that

$$\langle\,\Psi\,|\,e^{i(\phi_i-\phi_j)}\,|\,\Psi\,\rangle = e^{-\langle(\phi_i-\phi_j)^2\rangle/2} = e^{-(A^{-1}_{ii}-A^{-1}_{ij})} \quad . \tag{13.12}$$

Here we have used that $\langle e^Q\rangle = e^{\langle Q^2\rangle/2}$ where Q is a sum of Gaussian-distributed variables. Next, we need

$$\begin{aligned}\langle\,\Psi\,|\,\frac{\partial^2}{\partial\phi_i^2}\,|\,\Psi\,\rangle &= -\langle\,\Psi\,|\,\frac{\partial}{\partial\phi_i}\,\tfrac{1}{2}\,A_{ik}\,\phi_k\,|\,\Psi\,\rangle \\ &= -\tfrac{1}{2}\,A_{ii} + \tfrac{1}{4}\,A_{ik}\,A_{li}\,\langle\,\Psi\,|\,\phi_k\,\phi_l\,|\,\Psi\,\rangle = -\tfrac{1}{4}\,A_{ii} \quad .\end{aligned} \tag{13.13}$$

[1] See F. Alet and E. Sørensen, *Phys. Rev. E* **67**, 015701(R) (2003) and references therein.

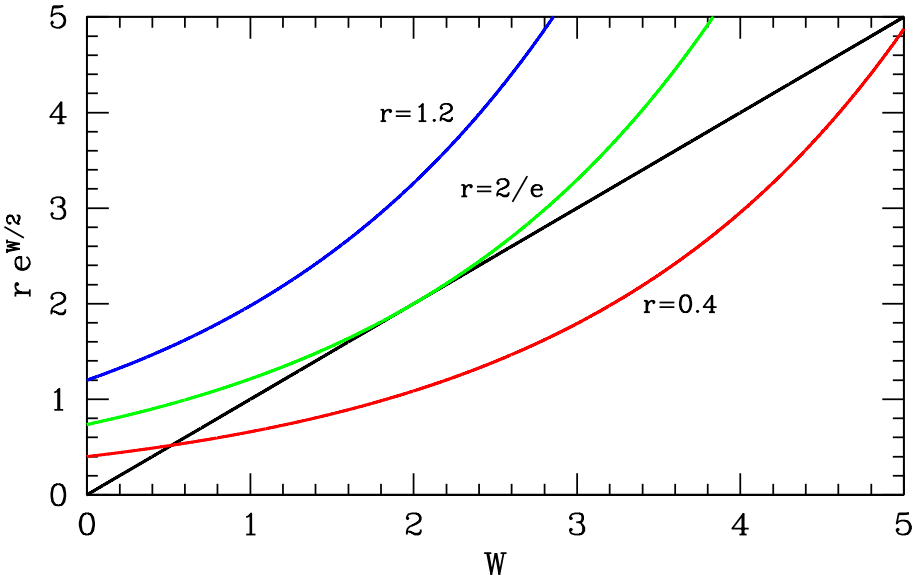


Figure 13.1: Graphical solution to the SCHA equation $W = r\,\exp\left(\frac{1}{2}W\right)$ for three representative values of r. The critical value is $r_{\rm c} = 2/e = 0.73576$.

Thus, the variational energy per site is

$$\begin{aligned}\frac{1}{N}\,\langle\,\Psi\,|\,\hat{H}_{\rm gr}\,|\,\Psi\,\rangle &= \tfrac{1}{4}\,UA_{ii} - zJ\,e^{-(A_{ii}^{-1}-A_{ij}^{-1})} \\ &= \tfrac{1}{4}U\int\frac{d^dk}{(2\pi)^d}\,\hat{A}(\boldsymbol{k}) - zJ\,\exp\left\{-\int\frac{d^dk}{(2\pi)^d}\,\frac{1-\gamma_{\boldsymbol{k}}}{\hat{A}(\boldsymbol{k})}\right\}\quad ,\end{aligned} \tag{13.14}$$

where z is the lattice coordination number ($N_{\rm links} = \frac{1}{2}zN$),

$$\gamma_{\boldsymbol{k}} = \frac{1}{z}\sum_{\boldsymbol{\delta}} e^{i\boldsymbol{k}\cdot\boldsymbol{\delta}} \tag{13.15}$$

is a sum over the z nearest neighbor vectors $\boldsymbol{\delta}$, and $\hat{A}(\boldsymbol{k})$ is the Fourier transform of A_{ij},

$$A_{ij} = \int\frac{d^dk}{(2\pi)^d}\,\hat{A}(\boldsymbol{k})\,e^{i(\boldsymbol{R}_i-\boldsymbol{R}_j)}\quad . \tag{13.16}$$

Note that $\hat{A}^*(\boldsymbol{k}) = \hat{A}(-\boldsymbol{k})$ since $\hat{A}(\boldsymbol{k})$ is the (discrete) Fourier transform of a real quantity.

We are now in a position to vary the energy in Eqn. 13.14 with respect to the variational parameters $\{\hat{A}(\boldsymbol{k})\}$. Taking the functional derivative with respect to $\hat{A}(\boldsymbol{k})$, we find

$$(2\pi)^d\,\frac{\delta(E_{\rm gr}/N)}{\delta\hat{A}(\boldsymbol{k})} = \tfrac{1}{4}\,U - \frac{1-\gamma_{\boldsymbol{k}}}{\hat{A}^2(\boldsymbol{k})}\cdot zJ\,e^{-W}\quad , \tag{13.17}$$

where

$$W = \int \frac{d^d k}{(2\pi)^d} \, \frac{1 - \gamma_{\boldsymbol{k}}}{\hat{A}(\boldsymbol{k})} \quad . \tag{13.18}$$

We now have

$$\hat{A}(\boldsymbol{k}) = 2 \left(\frac{zJ}{U} \right)^{1/2} e^{-W/2} \sqrt{1 - \gamma_{\boldsymbol{k}}} \quad . \tag{13.19}$$

Inserting this into our expression for W, we obtain the self-consistent equation

$$W = r \, e^{W/2} \qquad ; \qquad r = C_d \left(\frac{U}{4zJ} \right)^{1/2} \qquad , \qquad C_d \equiv \int \frac{d^d k}{(2\pi)^d} \, \sqrt{1 - \gamma_{\boldsymbol{k}}} \quad . \tag{13.20}$$

One finds $C_{d=1} = 0.900316$ for the linear chain, $C_{d=2} = 0.958091$ for the square lattice, and $C_{d=3} = 0.974735$ on the cubic lattice.

The graphical solution to $W = r \, \exp\left(\frac{1}{2}W\right)$ is shown in Fig. 13.1. One sees that for $r > r_{\rm c} = 2/e \simeq 0.73576$, there is no solution. In this case, the variational wavefunction should be taken to be $\Psi = 1$, which is a product of $\psi_{n=0}$ states on each grain, corresponding to fixed charge $M_i = 0$ and maximally fluctuating phase. In this case we must restrict each $\phi_i \in [0, 2\pi]$. When $r < r_{\rm c}$, though, there are two solutions for W. The larger of the two is spurious, and the smaller one is the physical one. As J/U increases, *i.e.* r decreases, the size of $\hat{A}(\boldsymbol{k})$ increases, which means that A^{-1}_{ij} decreases in magnitude. This means that the correlation in Eqn. 13.12 is growing, and the phase variables are localized. The SCHA predicts a spurious first order phase transition; the real superfluid-insulator transition is continuous (second-order)[2].

13.1.3 Calculation of the Cooper pair hopping amplitude

Finally, let us compute J_{ij}. We do so by working to second order in perturbation theory in the *electron* hopping Hamiltonian

$$\hat{H}_{\rm hop} = -\frac{1}{(V_i \, V_j)^{1/2}} \sum_{\langle ij \rangle} \sum_{\boldsymbol{k}, \boldsymbol{k}', \sigma} \Big(t_{ij}(\boldsymbol{k}, \boldsymbol{k}') \, c^\dagger_{i, \boldsymbol{k}, \sigma} \, c_{j, \boldsymbol{k}', \sigma} + t^*_{ij}(\boldsymbol{k}, \boldsymbol{k}') \, c^\dagger_{j, \boldsymbol{k}', \sigma} \, c_{i, \boldsymbol{k}, \sigma} \Big) \quad . \tag{13.21}$$

Here $t_{ij}(\boldsymbol{k}, \boldsymbol{k}')$ is the amplitude for an electron of wavevector $\boldsymbol{k}'$ in grain j to hop to a state of wavevector $\boldsymbol{k}$ in grain i. To simplify matters we will assume the grains are identical in all respects other than their overall phases. We'll write the fermion destruction operators on grain i as $c_{\boldsymbol{k}\sigma}$ and those on grain j as $\tilde{c}_{\boldsymbol{k}\sigma}$. We furthermore assume $t_{ij}(\boldsymbol{k}, \boldsymbol{k}') = t$ is real and independent of $\boldsymbol{k}$ and $\boldsymbol{k}'$. Only spin polarization, and not momentum, is preserved in the hopping process. Then

$$\hat{H}_{\rm hop} = -\frac{t}{V} \sum_{\boldsymbol{k}, \boldsymbol{k}'} \left(c^\dagger_{\boldsymbol{k}\sigma} \, \tilde{c}_{\boldsymbol{k}'\sigma} + \tilde{c}^\dagger_{\boldsymbol{k}'\sigma} \, c_{\boldsymbol{k}\sigma} \right) \quad . \tag{13.22}$$

[2]That the SCHA gives a spurious first order transition was recognized by E. Pytte, *Phys. Rev. Lett.* **28**, 895 (1971).

Each grain is described by a BCS model. The respective Bogoliubov transformations are

$$\begin{aligned} c_{\boldsymbol{k}\sigma} &= \cos\vartheta_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\sin\vartheta_{\boldsymbol{k}}\,e^{i\phi}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma} \\ \tilde{c}_{\boldsymbol{k}\sigma} &= \cos\tilde{\vartheta}_{\boldsymbol{k}}\,\tilde{\gamma}_{\boldsymbol{k}\sigma} - \sigma\sin\tilde{\vartheta}_{\boldsymbol{k}}\,e^{i\tilde{\phi}}\,\tilde{\gamma}^\dagger_{-\boldsymbol{k}\,-\sigma} \quad . \end{aligned} \tag{13.23}$$

Second order perturbation says that the ground state energy $\mathcal{E}$ is

$$\mathcal{E} = \mathcal{E}_0 - \sum_n \frac{\big|\langle\, n\,|\,\hat{H}_{\rm hop}\,|\,{\rm G}\,\rangle\big|^2}{\mathcal{E}_n - \mathcal{E}_0} \quad , \tag{13.24}$$

where $|\,{\rm G}\,\rangle = |\,{\rm G}_i\,\rangle\otimes|\,{\rm G}_j\,\rangle$ is a product of BCS ground states on the two grains. Clearly the only intermediate states $|\,n\,\rangle$ which can couple to $|\,{\rm G}\,\rangle$ through a single application of $\hat{H}_{\rm hop}$ are states of the form

$$|\,\boldsymbol{k},\boldsymbol{k}',\sigma\,\rangle = \gamma^\dagger_{\boldsymbol{k}\sigma}\,\tilde{\gamma}^\dagger_{-\boldsymbol{k}'\,-\sigma}\,|\,{\rm G}\,\rangle \quad , \tag{13.25}$$

and for this state

$$\langle\,\boldsymbol{k},\boldsymbol{k}',\sigma\,|\,\hat{H}_{\rm hop}\,|\,{\rm G}\,\rangle = -\sigma\Big(\cos\vartheta_{\boldsymbol{k}}\sin\tilde{\vartheta}_{\boldsymbol{k}'}\,e^{i\tilde{\phi}} + \sin\vartheta_{\boldsymbol{k}}\cos\tilde{\vartheta}_{\boldsymbol{k}'}\,e^{i\phi}\Big) \tag{13.26}$$

The energy of this intermediate state is

$$E_{\boldsymbol{k},\boldsymbol{k}',\sigma} = E_{\boldsymbol{k}} + E_{\boldsymbol{k}'} + \frac{e^2}{C} \quad , \tag{13.27}$$

where we have included the contribution from the charging energy of each grain. Then we find[3]

$$\mathcal{E}^{(2)} = \mathcal{E}_0' - J\cos(\phi - \tilde{\phi}\,) \quad , \tag{13.28}$$

where

$$J = \frac{|t|^2}{V^2}\sum_{\boldsymbol{k},\boldsymbol{k}'}\frac{\Delta_{\boldsymbol{k}}}{E_{\boldsymbol{k}}}\cdot\frac{\Delta_{\boldsymbol{k}'}}{E_{\boldsymbol{k}'}}\cdot\frac{1}{E_{\boldsymbol{k}} + E_{\boldsymbol{k}'} + (e^2/C)} \quad . \tag{13.29}$$

For a general set of dissimilar grains,

$$J_{ij} = \frac{|t_{ij}|^2}{V_iV_j}\sum_{\boldsymbol{k},\boldsymbol{k}'}\frac{\Delta_{i,\boldsymbol{k}}}{E_{i,\boldsymbol{k}}}\cdot\frac{\Delta_{j,\boldsymbol{k}'}}{E_{j,\boldsymbol{k}'}}\cdot\frac{1}{E_{i,\boldsymbol{k}} + E_{j,\boldsymbol{k}'} + (e^2/2C_{ij})} \quad , \tag{13.30}$$

where $C_{ij}^{-1} = C_i^{-1} + C_j^{-1}$.

[3]There is no factor of two arising from a spin sum since we are summing over all $\boldsymbol{k}$ and $\boldsymbol{k}'$, and therefore summing over spin would overcount the intermediate states $|n\rangle$ by a factor of two.

13.2 Tunneling

We follow the very clear discussion in §9.3 of G. Mahan's *Many Particle Physics*. Consider two bulk samples, which we label left (L) and right (R). The Hamiltonian is taken to be

$$\hat{H} = \hat{H}_{\rm L} + \hat{H}_{\rm R} + \hat{H}_{\rm T} \quad , \tag{13.31}$$

where $\hat{H}_{\rm L,R}$ are the bulk Hamiltonians, and

$$\hat{H}_{\rm T} = -\sum_{i,j,\sigma} \left(T_{ij}\, c^\dagger_{{\rm L}\,i\,\sigma}\, c_{{\rm R}\,j\,\sigma} + T^*_{ij}\, c^\dagger_{{\rm R}\,j\,\sigma}\, c_{{\rm L}\,i\,\sigma} \right) \quad . \tag{13.32}$$

The indices i and j label single particle electron states (*not* Bogoliubov quasiparticles) in the two banks. As we shall discuss below, we can take them to correspond to Bloch wavevectors in a particular energy band. In a nonequilibrium setting we work in the grand canonical ensemble, with

$$\hat{K} = \hat{H}_{\rm L} - \mu_{\rm L}\,\hat{N}_{\rm L} + \hat{H}_{\rm R} - \mu_{\rm R}\,\hat{N}_{\rm R} + \hat{H}_{\rm T} \quad . \tag{13.33}$$

The difference between the chemical potentials is $\mu_{\rm R} - \mu_{\rm L} = eV$, where V is the voltage bias. The current flowing from left to right is

$$I(t) = e \left\langle \frac{d\hat{N}_{\rm L}}{dt} \right\rangle \quad . \tag{13.34}$$

Note that if $N_{\rm L}$ is increasing in time, this means an electron number current flows from right to left, and hence an electrical current (of fictitious positive charges) flows from left to right. We use perturbation theory in $\hat{H}_{\rm T}$ to compute $I(t)$. Note that expectations such as $\langle \Psi_{\rm L} | c_{{\rm L}i} | \Psi_{\rm L} \rangle$ vanish, while $\langle \Psi_{\rm L} |\, c_{{\rm L}i}\, c_{{\rm L}j}\, | \Psi_{\rm L} \rangle$ may not if $|\Psi_{\rm L}\rangle$ is a BCS state.

A few words on the labels i and j: We will assume the left and right samples can be described as perfect crystals, so i and j will represent crystal momentum eigenstates. The only exception to this characterization will be that we assume their respective surfaces are sufficiently rough to destroy conservation of momentum in the plane of the surface. Momentum perpendicular to the surface is also not conserved, since the presence of the surface breaks translation invariance in this direction. The matrix element T_{ij} will be dominated by the behavior of the respective single particle electron wavefunctions in the vicinity of their respective surfaces. As there is no reason for the respective wavefunctions to be coherent, they will in general disagree in sign in random fashion. We then expect the overlap to be proportional to $A^{1/2}$, where A is the junction area, on the basis of the Central Limit Theorem. Adding in the plane wave normalization factors, we therefore approximate

$$T_{ij} = T_{\boldsymbol{q},\boldsymbol{k}} \approx \left(\frac{A}{V_{\rm L} V_{\rm R}} \right)^{1/2} t\big(\xi_{{\rm L}\,\boldsymbol{q}}\,,\, \xi_{{\rm R}\,\boldsymbol{k}}\big) \quad , \tag{13.35}$$

where $\boldsymbol{q}$ and $\boldsymbol{k}$ are the wavevectors of the Bloch electrons on the left and right banks, respectively. Note that we presume spin is preserved in the tunneling process, although wavevector is not.

13.2.1 Perturbation theory

We begin by noting

$$\begin{aligned}\frac{d\hat{N}_{\rm L}}{dt} = \frac{i}{\hbar}\big[\hat{H}, \hat{N}_{\rm L}\big] &= \frac{i}{\hbar}\big[\hat{H}_{\rm T}\,,\,\hat{N}_{\rm L}\big] \\ &= -\frac{i}{\hbar}\sum_{i,j,\sigma}\big(T_{ij}\,c^\dagger_{{\rm L}\,i\,\sigma}\,c_{{\rm R}\,j\,\sigma} - T^*_{ij}\,c^\dagger_{{\rm R}\,j\,\sigma}\,c_{{\rm L}\,i\,\sigma}\big) \quad .\end{aligned} \tag{13.36}$$

First order perturbation theory then gives

$$|\,\Psi(t)\,\rangle = e^{-i\hat{H}_0(t-t_0)/\hbar}\,|\,\Psi(t_0)\,\rangle - \frac{i}{\hbar}\,e^{-i\hat{H}_0 t/\hbar}\int_{t_0}^{t} dt_1\;\hat{H}_{\rm T}(t_1)\,e^{i\hat{H}_0 t_0/\hbar}\,|\,\Psi(t_0)\,\rangle + \mathcal{O}\big(\hat{H}_{\rm T}^2\big) \quad , \tag{13.37}$$

where $\hat{H}_0 = \hat{H}_{\rm L} + \hat{H}_{\rm R}$ and

$$\hat{H}_{\rm T}(t) = e^{i\hat{H}_0 t/\hbar}\,\hat{H}_{\rm T}\,e^{-i\hat{H}_0 t/\hbar} \tag{13.38}$$

is the perturbation (hopping) Hamiltonian in the interaction representation. To lowest order in $\hat{H}_{\rm T}$, then,

$$\langle\,\Psi(t)\,|\,\hat{I}\,|\,\Psi(t)\,\rangle = -\frac{i}{\hbar}\int_{t_0}^{t} dt_1\,\langle\,\tilde{\Psi}(t_0)\,|\,\big[\hat{I}(t)\,,\,\hat{H}_{\rm T}(t_1)\big]\,|\,\tilde{\Psi}(t_0)\,\rangle \quad , \tag{13.39}$$

where $|\,\tilde{\Psi}(t_0)\,\rangle = e^{i\hat{H}_0 t_0/\hbar}\,|\,\Psi(t_0)\,\rangle$. Setting $t_0 = -\infty$, and averaging over a thermal ensemble of initial states, we have

$$I(t) = -\frac{i}{\hbar}\int_{-\infty}^{t} dt'\,\big\langle\big[\hat{I}(t)\,,\,\hat{H}_{\rm T}(t')\big]\big\rangle \quad , \tag{13.40}$$

where $\hat{I}(t) = e\dot{\hat{N}}_{\rm L}(t) = (+e)\,e^{i\hat{H}_0 t/\hbar}\,\dot{\hat{N}}_{\rm L}\,e^{-i\hat{H}_0 t/\hbar}$ is the charge current flowing from *right* to *left*. Note that it is the electron charge $-e$ that enters here and not the Cooper pair charge, since $\hat{H}_{\rm T}$ describes electron hopping.

There remains a *caveat* which we have already mentioned. The chemical potentials $\mu_{\rm L}$ and $\mu_{\rm R}$ differ according to

$$\mu_{\rm R} - \mu_{\rm L} = eV \quad , \tag{13.41}$$

where V is the bias voltage, *i.e.* the voltage *drop* from left to right. If $V > 0$, then $\mu_{\rm R} > \mu_{\rm L}$, which means an electron current flows from right to left, and an electrical current (*i.e.* the direction of positive charge flow) from left to right. We must work in an ensemble described by $\hat{K}_0$, where

$$\hat{K}_0 = \hat{H}_{\rm L} - \mu_{\rm L}\hat{N}_{\rm L} + \hat{H}_{\rm R} - \mu_{\rm R}\hat{N}_{\rm R} \quad . \tag{13.42}$$

We now separate $\hat{H}_{\rm T}$ into its component processes, writing $\hat{H}_{\rm T} = \hat{H}_{\rm T}^+ + \hat{H}_{\rm T}^-$, with

$$\hat{H}_{\rm T}^+ = -\sum_{i,j,\sigma} T_{ij}\, c^\dagger_{{\rm L}\,i\sigma}\, c_{{\rm R}\,j\sigma} \quad , \quad \hat{H}_{\rm T}^- = -\sum_{i,j,\sigma} T^*_{ij}\, c^\dagger_{{\rm R}\,j\sigma}\, c_{{\rm L}\,i\sigma} \quad . \tag{13.43}$$

Thus, $\hat{H}_{\rm T}^+$ describes hops from R to L, and $\hat{H}_{\rm T}^-$ from L to R. Note that $\hat{H}_{\rm T}^- = (\hat{H}_{\rm T}^+)^\dagger$. Therefore $\hat{H}_{\rm T}(t) = \hat{H}_{\rm T}^+(t) + \hat{H}_{\rm T}^-(t)$, where[4]

$$\begin{aligned}\hat{H}_{\rm T}^\pm(t) &= e^{i(\hat{K}_0+\mu_{\rm L}\hat{N}_{\rm L}+\mu_{\rm R}\hat{N}_{\rm R})t/\hbar}\,\hat{H}_{\rm T}^\pm\, e^{-i(\hat{K}_0+\mu_{\rm L}\hat{N}_{\rm L}+\mu_{\rm R}\hat{N}_{\rm R})t/\hbar} \\ &= e^{\mp ieVt/\hbar}\, e^{i\hat{K}_0 t/\hbar}\,\hat{H}_{\rm T}^\pm\, e^{-i\hat{K}_0 t/\hbar} \quad .\end{aligned} \tag{13.44}$$

Note that the current operator is

$$\hat{I} = \frac{ie}{\hbar}\left[\hat{H}_{\rm T}\,,\, N_{\rm L}\right] = \frac{ie}{\hbar}\left(\hat{H}_{\rm T}^- - \hat{H}_{\rm T}^+\right) \quad . \tag{13.45}$$

We then have

$$\begin{aligned}I(t) &= \frac{e}{\hbar^2}\int_{-\infty}^{t} dt' \left\langle \left[e^{ieVt/\hbar}\,\hat{H}_{\rm T}^-(t) - e^{-ieVt/\hbar}\,\hat{H}_{\rm T}^+(t)\,,\, e^{ieVt'/\hbar}\,\hat{H}_{\rm T}^-(t') + e^{-ieVt'/\hbar}\,\hat{H}_{\rm T}^+(t')\right]\right\rangle \\ &= I_{\rm N}(t) + I_{\rm J}(t) \quad ,\end{aligned} \tag{13.46}$$

where

$$I_{\rm N}(t) = \frac{e}{\hbar^2}\int_{-\infty}^{\infty} dt'\,\Theta(t-t')\left\{e^{+i\Omega(t-t')}\left\langle\left[\hat{H}_{\rm T}^-(t)\,,\,\hat{H}_{\rm T}^+(t')\right]\right\rangle - e^{-i\Omega(t-t')}\left\langle\left[\hat{H}_{\rm T}^+(t)\,,\,\hat{H}_{\rm T}^-(t')\right]\right\rangle\right\} \tag{13.47}$$

and

$$I_{\rm J}(t) = \frac{e}{\hbar^2}\int_{-\infty}^{\infty} dt'\,\Theta(t-t')\left\{e^{+i\Omega(t+t')}\left\langle\left[\hat{H}_{\rm T}^-(t)\,,\,\hat{H}_{\rm T}^-(t')\right]\right\rangle - e^{-i\Omega(t+t')}\left\langle\left[\hat{H}_{\rm T}^+(t)\,,\,\hat{H}_{\rm T}^+(t')\right]\right\rangle\right\}, \tag{13.48}$$

with $\Omega \equiv eV/\hbar$. $I_{\rm N}(t)$ is the usual *single particle tunneling current*, which is present both in normal metals as well as in superconductors. $I_{\rm J}(t)$ is the *Josephson pair tunneling current*, which is only present when the ensemble average is over states of indefinite particle number.

13.2.2 The single particle tunneling current $I_{\rm N}$

We now proceed to evaluate the so-called single-particle current $I_{\rm N}$ in Eqn. 13.47. This current is present, under voltage bias, between normal metal and normal metal, between normal metal

[4] We make use of the fact that $\hat{N}_{\rm L} + \hat{N}_{\rm R}$ commutes with $\hat{H}_{\rm T}^\pm$.

and superconductor, and between superconductor and superconductor. It is convenient to define the quantities

$$\begin{aligned}\mathcal{X}_{\mathsf{r}}(t-t') &\equiv -i\,\Theta(t-t')\left\langle\left[\hat{H}_{\rm T}^{-}(t)\,,\,\hat{H}_{\rm T}^{+}(t')\right]\right\rangle \\ \mathcal{X}_{\mathsf{a}}(t-t') &\equiv -i\,\Theta(t-t')\left\langle\left[\hat{H}_{\rm T}^{-}(t')\,,\,\hat{H}_{\rm T}^{+}(t)\right]\right\rangle \quad ,\end{aligned} \tag{13.49}$$

which differ by the order of the time values of the operators inside the commutator. We then have

$$\begin{aligned}I_{\rm N} &= \frac{ie}{\hbar^2}\int\limits_{-\infty}^{\infty}dt\,\left\{e^{+i\Omega t}\,\mathcal{X}_{\mathsf{r}}(t)+e^{-i\Omega t}\,\mathcal{X}_{\mathsf{a}}(t)\right\} \\ &= \frac{ie}{\hbar^2}\Big(\widetilde{\mathcal{X}}_{\mathsf{r}}(\Omega)+\widetilde{\mathcal{X}}_{\mathsf{a}}(-\Omega)\Big) \quad ,\end{aligned} \tag{13.50}$$

where $\widetilde{\mathcal{X}}_{\mathsf{a}}(\Omega)$ is the Fourier transform of $\mathcal{X}_{\mathsf{a}}(t)$ into the frequency domain. As we shall show presently, $\widetilde{\mathcal{X}}_{\mathsf{a}}(-\Omega)=-\widetilde{\mathcal{X}}_{\mathsf{r}}^{*}(\Omega)$, so we have

$$I_{\rm N}(V) = -\frac{2e}{\hbar^2}\,\mathsf{Im}\,\widetilde{\mathcal{X}}_{\mathsf{r}}(eV/\hbar) \quad . \tag{13.51}$$

Proof that $\tilde{\mathcal{X}}_{\mathsf{a}}(\Omega)=-\widetilde{\mathcal{X}}_{\mathsf{r}}^{*}(-\Omega)$

Consider the general case

$$\begin{aligned}\mathcal{X}_{\mathsf{r}}(t) &= -i\,\Theta(t)\left\langle\left[\hat{A}(t)\,,\,\hat{A}^{\dagger}(0)\right]\right\rangle \\ \mathcal{X}_{\mathsf{a}}(t) &= -i\,\Theta(t)\left\langle\left[\hat{A}(0)\,,\,\hat{A}^{\dagger}(t)\right]\right\rangle \quad .\end{aligned} \tag{13.52}$$

We now spectrally decompose these expressions, inserting complete sets of states in between products of operators. One finds

$$\begin{aligned}\widetilde{\mathcal{X}}_{\mathsf{r}}(\omega) &= -i\int\limits_{-\infty}^{\infty}dt\,\Theta(t)\sum_{m,n}P_m\left\{\big|\langle\,m\,|\,\hat{A}\,|\,n\,\rangle\big|^2e^{i(\omega_m-\omega_n)t}-\big|\langle\,m\,|\,\hat{A}^{\dagger}\,|\,n\,\rangle\big|^2e^{-i(\omega_m-\omega_n)t}\right\}e^{i\omega t} \\ &= \sum_{m,n}P_m\left\{\frac{\big|\langle\,m\,|\,\hat{A}\,|\,n\,\rangle\big|^2}{\omega+\omega_m-\omega_n+i\epsilon}-\frac{\big|\langle\,m\,|\,\hat{A}^{\dagger}\,|\,n\,\rangle\big|^2}{\omega-\omega_m+\omega_n+i\epsilon}\right\} \quad ,\end{aligned} \tag{13.53}$$

where the eigenvalues of $\hat{K}$ are $\hbar\omega_m$, and $P_m=e^{-\hbar\omega_m/k_{\rm B}T}\big/\,\Xi$ is the thermal probability for state $|\,m\,\rangle$, where Ξ is the grand partition function. The corresponding expression for $\widetilde{\mathcal{X}}_{\mathsf{a}}(\omega)$ is

$$\widetilde{\mathcal{X}}_{\mathsf{a}}(\omega) = \sum_{m,n}P_m\left\{\frac{\big|\langle\,m\,|\,\hat{A}\,|\,n\,\rangle\big|^2}{\omega-\omega_m+\omega_n+i\epsilon}-\frac{\big|\langle\,m\,|\,\hat{A}^{\dagger}\,|\,n\,\rangle\big|^2}{\omega+\omega_m-\omega_n+i\epsilon}\right\} \quad , \tag{13.54}$$

whence follows $\widetilde{\mathcal{X}}_{\rm a}(-\omega) = -\widetilde{\mathcal{X}}^*_{\rm r}(\omega)$. QED. Note that in general

$$\begin{aligned}\mathcal{Z}(t) = -i\,\Theta(t)\,\big\langle \hat{A}(t)\,\hat{B}(0)\big\rangle &= -i\,\Theta(t)\sum_{m,n} P_m\,\langle\, m\,|\,e^{i\hat{K}t/\hbar}\,\hat{A}\,e^{-i\hat{K}t/\hbar}\,|\,n\,\rangle\langle\, n\,|\,\hat{B}\,|\,m\,\rangle \\ &= -i\,\Theta(t)\sum_{m,n} P_m\,\langle\, m\,|\,\hat{A}\,|\,n\,\rangle\langle\, n\,|\,\hat{B}\,|\,m\,\rangle\,e^{i(\omega_m-\omega_n)t} \quad ,\end{aligned} \tag{13.55}$$

the Fourier transform of which is

$$\widetilde{\mathcal{Z}}(\omega) = \int\limits_{-\infty}^{\infty} dt\,e^{i\omega t}\,\mathcal{Z}(t) = \sum_{m,n} P_m\,\frac{\langle\, m\,|\,\hat{A}\,|\,n\,\rangle\langle\, n\,|\,\hat{B}\,|\,m\,\rangle}{\omega+\omega_m-\omega_n+i\epsilon} \quad . \tag{13.56}$$

If we define the *spectral density* $\rho(\omega)$ as

$$\rho(\omega) = 2\pi\sum_{m,n} P_m\langle\, m\,|\,\hat{A}\,|\,n\,\rangle\langle\, n\,|\,\hat{B}\,|\,m\,\rangle\,\delta(\omega+\omega_m-\omega_n) \quad , \tag{13.57}$$

then we have

$$\widetilde{\mathcal{Z}}(\omega) = \int\limits_{-\infty}^{\infty}\frac{d\nu}{2\pi}\,\frac{\rho(\nu)}{\omega-\nu+i\epsilon} \quad . \tag{13.58}$$

Note that $\rho(\omega)$ is real if $B = A^\dagger$.

Evaluation of $\widetilde{\mathcal{X}}_{\rm r}(\omega)$

We must compute

$$\begin{aligned}\mathcal{X}_{\rm r}(t) &= -i\,\Theta(t)\sum_{i,j,\sigma}\sum_{k,l,\sigma'} T^*_{kl}\,T_{ij}\,\Big\langle\Big[c^\dagger_{{\rm R}\,j\,\sigma}(t)\,c_{{\rm L}\,i\,\sigma}(t)\,,\,c^\dagger_{{\rm L}\,k\,\sigma'}(0)\,c_{{\rm R}\,l\,\sigma'}(0)\Big]\Big\rangle \\ &= -i\,\Theta(t)\sum_{\boldsymbol{q},\boldsymbol{k},\sigma}|T_{\boldsymbol{q},\boldsymbol{k}}|^2\,\Big\{\big\langle c^\dagger_{{\rm R}\,\boldsymbol{k}\,\sigma}(t)\,c_{{\rm R}\,\boldsymbol{k}\,\sigma}(0)\big\rangle\,\big\langle c_{{\rm L}\,\boldsymbol{q}\,\sigma}(t)\,c^\dagger_{{\rm L}\,\boldsymbol{q}\,\sigma}(0)\big\rangle \\ &\qquad\qquad - \big\langle c^\dagger_{{\rm L}\,\boldsymbol{q}\,\sigma}(0)\,c_{{\rm L}\,\boldsymbol{q}\,\sigma}(t)\big\rangle\,\big\langle c_{{\rm R}\,\boldsymbol{k}\,\sigma}(0)\,c^\dagger_{{\rm R}\,\boldsymbol{k}\,\sigma}(t)\big\rangle\Big\}\end{aligned} \tag{13.59}$$

Note how we have taken $j = l \to \boldsymbol{k}$ and $i = k \to \boldsymbol{q}$, since *in each bank* wavevector is assumed to be a good quantum number. We now invoke the Bogoliubov transformation,

$$c_{\boldsymbol{k}\sigma} = u_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\,v_{\boldsymbol{k}}\,e^{i\phi}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma} \quad , \tag{13.60}$$

where we write $u_{\boldsymbol{k}} = \cos\vartheta_{\boldsymbol{k}}$ and $v_{\boldsymbol{k}} = \sin\vartheta_{\boldsymbol{k}}$. We then have

$$\begin{aligned}
\big\langle c^\dagger_{\mathrm{R}\,\boldsymbol{k}\,\sigma}(t)\,c_{\mathrm{R}\,\boldsymbol{k}\,\sigma}(0)\big\rangle &= u^2_{\boldsymbol{k}}\,e^{iE_{\boldsymbol{k}}t/\hbar}\,f(E_{\boldsymbol{k}}) + v^2_{\boldsymbol{k}}\,e^{-iE_{\boldsymbol{k}}t/\hbar}\,\big[1-f(E_{\boldsymbol{k}})\big] \\
\big\langle c_{\mathrm{L}\,\boldsymbol{q}\,\sigma}(t)\,c^\dagger_{\mathrm{L}\,\boldsymbol{q}\,\sigma}(0)\big\rangle &= u^2_{\boldsymbol{q}}\,e^{-iE_{\boldsymbol{q}}t/\hbar}\,\big[1-f(E_{\boldsymbol{q}})\big] + v^2_{\boldsymbol{q}}\,e^{iE_{\boldsymbol{q}}t/\hbar}\,f(E_{\boldsymbol{q}}) \\
\big\langle c^\dagger_{\mathrm{L}\,\boldsymbol{q}\,\sigma}(0)\,c_{\mathrm{L}\,\boldsymbol{q}\,\sigma}(t)\big\rangle &= u^2_{\boldsymbol{q}}\,e^{-iE_{\boldsymbol{q}}t/\hbar}\,f(E_{\boldsymbol{q}}) + v^2_{\boldsymbol{q}}\,e^{iE_{\boldsymbol{q}}t/\hbar}\,\big[1-f(E_{\boldsymbol{q}})\big] \\
\big\langle c_{\mathrm{R}\,\boldsymbol{k}\,\sigma}(0)\,c^\dagger_{\mathrm{R}\,\boldsymbol{k}\,\sigma}(t)\big\rangle &= u^2_{\boldsymbol{k}}\,e^{iE_{\boldsymbol{k}}t/\hbar}\,\big[1-f(E_{\boldsymbol{k}})\big] + v^2_{\boldsymbol{k}}\,e^{-iE_{\boldsymbol{k}}t/\hbar}\,f(E_{\boldsymbol{k}}) \quad .
\end{aligned} \tag{13.61}$$

We now appeal to Eqn. 13.35 and convert the $\boldsymbol{q}$ and $\boldsymbol{k}$ sums to integrals over $\xi_{\mathrm{L}\,\boldsymbol{q}}$ and $\xi_{\mathrm{R}\,\boldsymbol{k}}$. Pulling out the DOS factors $g_{\mathrm{L}} \equiv g_{\mathrm{L}}(\mu_{\mathrm{L}})$ and $g_{\mathrm{R}} \equiv g_{\mathrm{R}}(\mu_{\mathrm{R}})$, as well as the hopping integral $t \equiv t\big(\xi_{\mathrm{L}\,\boldsymbol{q}}=0\,,\,\xi_{\mathrm{R}\,\boldsymbol{k}}=0\big)$ from the integrand, we have

$$\begin{aligned}
\mathcal{X}_{\mathsf{r}}(t) = -i\,\Theta(t) \times \tfrac{1}{2}\,g_{\mathrm{L}}\,g_{\mathrm{R}}\,|t|^2 A \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\xi' \times \\
\Big\{ \Big[u^2\,e^{-iEt/\hbar}\,(1-f) + v^2\,e^{iEt/\hbar}\,f\Big] \times \Big[u'^2\,e^{iE't/\hbar}\,f' + v'^2\,e^{-iE't/\hbar}\,(1-f')\Big] \\
- \Big[u^2\,e^{-iEt/\hbar}\,f + v^2\,e^{iEt/\hbar}\,(1-f)\Big] \times \Big[u'^2\,e^{iE't/\hbar}\,(1-f') + v'^2\,e^{-iE't/\hbar}\,f'\Big] \Big\} \quad ,
\end{aligned} \tag{13.62}$$

where unprimed quantities correspond to the left bank (L) and primed quantities to the right bank (R). The ξ and ξ' integrals are simplified by the fact that in $u^2 = (E+\xi)/2E$ and $v^2 = (E-\xi)/2E$, *etc.* The terms proportional to ξ and ξ' and to $\xi\xi'$ drop out because everything else in the integrand is even in ξ and ξ' separately. Thus, we may replace u^2, v^2, u'^2, and v'^2 all by $\frac{1}{2}$. We now compute the Fourier transform, and we can read off the results keeping in mind the integral,

$$\int_0^\infty dt\,e^{i\omega t}\,e^{i\Omega t}\,e^{-\epsilon t} = \frac{i}{\omega+\Omega+i\epsilon} \quad . \tag{13.63}$$

We then obtain

$$\begin{aligned}
\widetilde{\mathcal{X}}_{\mathsf{r}}(\omega) = \tfrac{1}{8}\,\hbar\,g_{\mathrm{L}}\,g_{\mathrm{R}}\,|t|^2 A \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\xi' \Big\{ \frac{2\,(f'-f)}{\hbar\omega+E'-E+i\epsilon} + \frac{1-f-f'}{\hbar\omega-E-E'+i\epsilon} \\
- \frac{1-f-f'}{\hbar\omega+E+E'+i\epsilon} \Big\} \quad .
\end{aligned} \tag{13.64}$$

Therefore,

$$I_{\rm N}(V,T) = -\frac{2e}{\hbar^2}\,{\rm Im}\,\widetilde{\mathcal{X}}_{\rm r}(eV/\hbar) \tag{13.65}$$
$$= \frac{\pi e}{\hbar}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\int_0^\infty d\xi\int_0^\infty d\xi'\,\Big\{(1-f-f')\Big[\delta(E+E'-eV)-\delta(E+E'+eV)\Big] + 2\,(f'-f)\,\delta(E'-E+eV)\Big\}\quad .$$

Single particle tunneling current in NIN junctions

We now evaluate $I_{\rm N}$ from Eqn. 13.65 for the case where both banks are normal metals. In this case, $E=\xi$ and $E'=\xi'$. (No absolute value symbol is needed since the ξ and ξ' integrals run over the positive real numbers.) At zero temperature, we have $f=0$ and thus

$$I_{\rm N}(V,T=0) = \frac{\pi e}{\hbar}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\int_0^\infty d\xi\int_0^\infty d\xi'\,\Big[\delta(\xi+\xi'-eV)-\delta(\xi+\xi'+eV)\Big]$$
$$= \frac{\pi e}{\hbar}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\int_0^{eV} d\xi = \frac{\pi e^2}{\hbar}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\,V\quad . \tag{13.66}$$

We thus identify the normal state conductance of the junction as

$$G_{\rm N} \equiv \frac{\pi e^2}{\hbar}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\quad . \tag{13.67}$$

Single particle tunneling current in NIS junctions

Consider the case where one of the banks is a superconductor and the other a normal metal. We will assume $V>0$ and work at $T=0$. From Eqn. 13.65, we then have

$$I_{\rm N}(V,T=0) = \frac{G_{\rm N}}{e}\int_0^\infty d\xi\int_0^\infty d\xi'\,\delta(\xi+E'-eV) = \frac{G_{\rm N}}{e}\int_0^\infty d\xi\,\Theta(eV-E)$$
$$= \frac{G_{\rm N}}{e}\int_\Delta^{eV} dE\,\frac{E}{\sqrt{E^2-\Delta^2}} = G_n\,\sqrt{V^2-(\Delta/e)^2}\quad . \tag{13.68}$$

The zero temperature conductance of the NIS junction is therefore

$$G_{\rm NIS}(V) = \frac{dI}{dV} = \frac{G_{\rm N}\,eV}{\sqrt{(eV)^2-\Delta^2}}\quad . \tag{13.69}$$

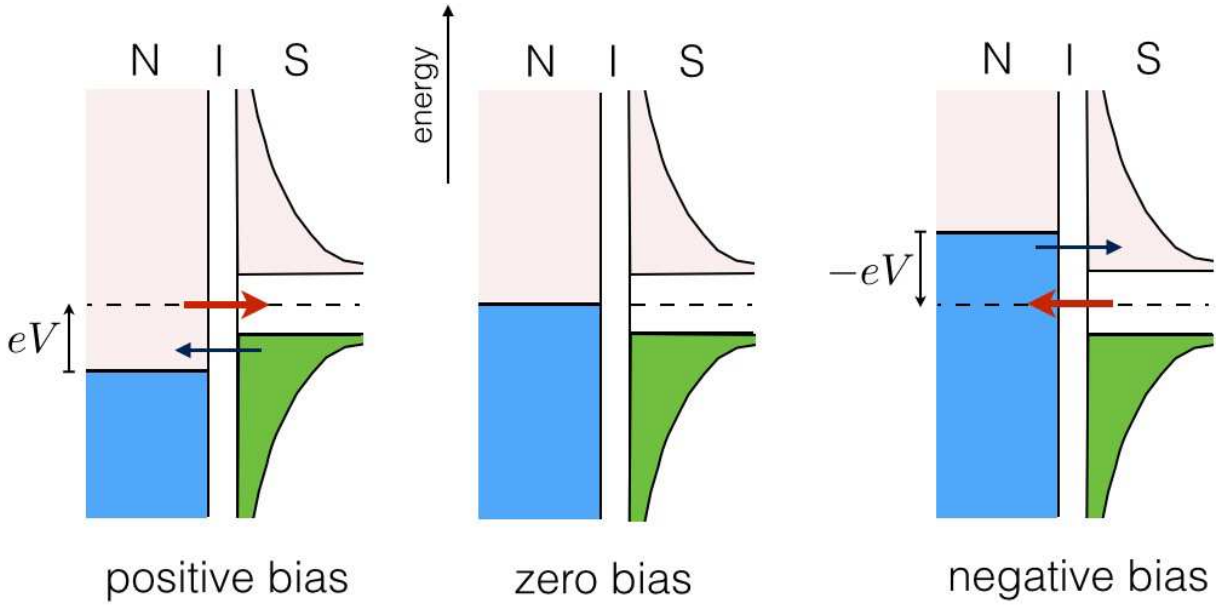


Figure 13.2: NIS tunneling for positive bias (left), zero bias (center), and negative bias (right). The left bank is maintained at an electrical potential V with respect to the right, hence $\mu_{\rm R} = \mu_{\rm L} + eV$. Blue regions indicate occupied fermionic states in the metal. Green regions indicate occupied electronic states in the superconductor. Light red regions indicate unoccupied states. Tunneling from or into the metal can only take place when its Fermi level lies outside the superconductor's gap region, meaning $|eV| > \Delta$, where V is the bias voltage. The arrow indicates the direction of electron number current. Black arrows indicate direction of electron current. Thick red arrows indicate direction of electrical current.

Hence the ratio $G_{\rm NIS}/G_{\rm NIN}$ is

$$\frac{G_{\rm NIS}(V)}{G_{\rm NIN}(V)} = \frac{eV}{\sqrt{(eV)^2 - \Delta^2}} \quad . \tag{13.70}$$

It is to be understood that these expressions are to be multiplied by $\mathrm{sgn}(V)\,\Theta\big(e|V|-\Delta\big)$ to obtain the full result valid at all voltages.

Superconducting density of states

We define

$$\begin{aligned} n_{\rm s}(E) &= 2\int \frac{d^3k}{(2\pi)^3}\,\delta(E - E_{\boldsymbol{k}}) \simeq g(\mu)\int_{-\infty}^{\infty} d\xi\;\delta\Big(E - \sqrt{\xi^2+\Delta^2}\Big) \\ &= g(\mu)\,\frac{2E}{\sqrt{E^2-\Delta^2}}\,\Theta(E-\Delta) \quad . \end{aligned} \tag{13.71}$$

This is the density of energy states per unit volume for elementary excitations in the superconducting state. Note that there is an *energy gap* of size Δ, and that the missing states from this

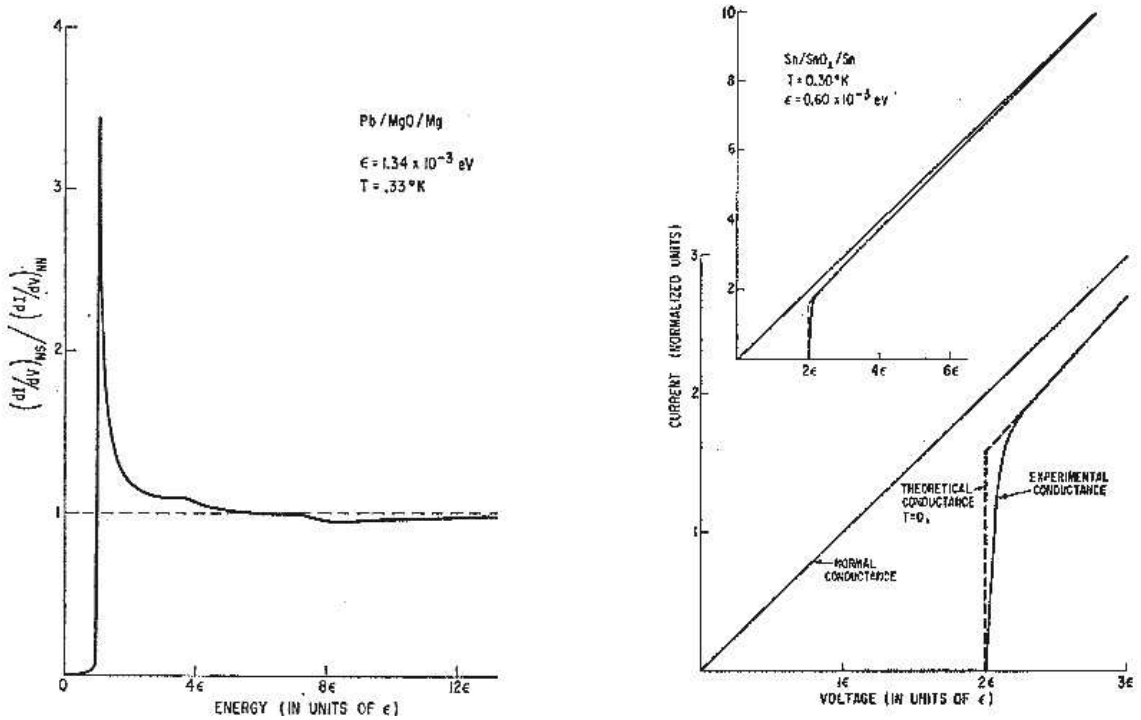


Figure 13.3: Tunneling data by Giaever *et al.* from *Phys. Rev.* **126**, 941 (1962). Left: normalized NIS tunneling conductance in a Pb/MgO/Mg sandwich junction. Pb is a superconductor for $T < T_{\rm c}^{\rm Pb} = 7.19\,{\rm K}$, and Mg is a metal. A thin MgO layer provides a tunnel barrier. Right: I-V characteristic for a SIS junction Sn/SnO_x/Sn. Sn is a superconductor for $T < T_{\rm c}^{\rm Sn} = 2.32\,{\rm K}$.

region pile up for $E \gtrsim \Delta$, resulting in a (integrable) divergence of $n_{\rm s}(E)$. In the limit $\Delta \to 0$, we have $n_{\rm s}(E) = 2\,g(\mu)\,\Theta(E)$. The factor of two arises because $n_{\rm s}(E)$ is the total density of states, which includes particle excitations above $k_{\rm F}$ as well as hole excitations below $k_{\rm F}$, both of which contribute $g(\mu)$. If $\Delta(\xi)$ is energy-dependent in the vicinity of $\xi = 0$, then we have

$$n(E) = g(\mu) \cdot \frac{E}{\xi} \cdot \left(1 + \frac{\Delta}{\xi}\frac{d\Delta}{d\xi}\right)^{-1}\Bigg|_{\xi=\sqrt{E^2-\Delta^2(\xi)}} \quad . \tag{13.72}$$

Here, $\xi = \sqrt{E^2 - \Delta^2(\xi)}$ is an implicit relation for $\xi(E)$.

The function $n_{\rm s}(E)$ vanishes for $E < 0$. We can, however, make a particle-hole transformation on the Bogoliubov operators, so that

$$\gamma_{\boldsymbol{k}\sigma} = \psi_{\boldsymbol{k}\sigma}\,\Theta(\xi_{\boldsymbol{k}}) + \psi^\dagger_{-\boldsymbol{k}\,-\sigma}\,\Theta(-\xi_{\boldsymbol{k}}) \quad . \tag{13.73}$$

We then have, up to constants,

$$\hat{K}_{\rm BCS} = \sum_{\boldsymbol{k}\sigma} \mathcal{E}_{\boldsymbol{k}\sigma}\,\psi^\dagger_{\boldsymbol{k}\sigma}\,\psi_{\boldsymbol{k}\sigma} \quad , \tag{13.74}$$

where

$$\mathcal{E}_{\boldsymbol{k}\sigma} = \begin{cases} +E_{\boldsymbol{k}\sigma} & \text{if } \xi_{\boldsymbol{k}} > 0 \\ -E_{\boldsymbol{k}\sigma} & \text{if } \xi_{\boldsymbol{k}} < 0 \end{cases} \quad . \tag{13.75}$$

The density of states for the ψ particles is then

$$\widetilde{n}_{\rm S}(\mathcal{E}) = \frac{g_{\rm S}\,|\mathcal{E}|}{\sqrt{\mathcal{E}^2 - \Delta^2}}\,\Theta\big(|\mathcal{E}| - \Delta\big) \quad , \tag{13.76}$$

were $g_{\rm S}$ is the metallic DOS at the Fermi level in the superconducting bank, *i.e.* above $T_{\rm c}$. Note that $\widetilde{n}_{\rm S}(-\mathcal{E}) = \widetilde{n}_{\rm S}(\mathcal{E})$ is now an even function of $\mathcal{E}$, and that half of the weight from $n_{\rm S}(E)$ has now been assigned to negative $\mathcal{E}$ states. The interpretation of Fig. 13.2 follows by writing

$$I_{\rm N}(V, T=0) = \frac{G_{\rm N}}{e g_{\rm S}} \int\limits_0^{eV} d\mathcal{E}\; n_{\rm S}(\mathcal{E}) \quad . \tag{13.77}$$

Note that this is properly odd under $V \to -V$. If $V > 0$, the tunneling current is proportional to the integral of the superconducting density of states from $\mathcal{E} = \Delta$ to $\mathcal{E} = eV$. Since $\widetilde{n}_{\rm S}(\mathcal{E})$ vanishes for $|\mathcal{E}| < \Delta$, the tunnel current vanishes if $|eV| < \Delta$.

Single particle tunneling current in SIS junctions

We now come to the SIS case, where both banks are superconducting. From Eqn. 13.65, we have ($T = 0$)

$$\begin{aligned} I_{\rm N}(V, T=0) &= \frac{G_{\rm N}}{e} \int\limits_0^\infty d\xi \int\limits_0^\infty d\xi'\; \delta(E + E' - eV) \\ &= \frac{G_{\rm N}}{e} \int\limits_0^\infty dE \int\limits_0^\infty dE'\; \frac{E}{\sqrt{E^2 - \Delta_{\rm L}^2}}\,\frac{E'}{\sqrt{E'^2 - \Delta_{\rm R}^2}} \Big\{ \delta(E + E' - eV) - \delta(E + E' + eV) \Big\} \quad . \end{aligned} \tag{13.78}$$

While this integral has no general analytic form, we see that $I_{\rm N}(V) = -I_{\rm N}(-V)$, and that the threshold voltage V^* below which $I_{\rm N}(V)$ vanishes is given by $eV^* = \Delta_{\rm L} + \Delta_{\rm R}$. For the special case $\Delta_{\rm L} = \Delta_{\rm R} \equiv \Delta$, one has

$$I_{\rm N}(V) = \frac{G_{\rm N}}{e} \left\{ \frac{(eV)^2}{eV + 2\Delta}\,\mathbb{K}(x) - (eV + 2\Delta)\,\Big(\mathbb{K}(x) - \mathbb{E}(x)\Big) \right\} \quad , \tag{13.79}$$

where $x = (eV - 2\Delta)/(eV + 2\Delta)$ and $\mathbb{K}(x)$ and $\mathbb{E}(x)$ are complete elliptic integrals of the first

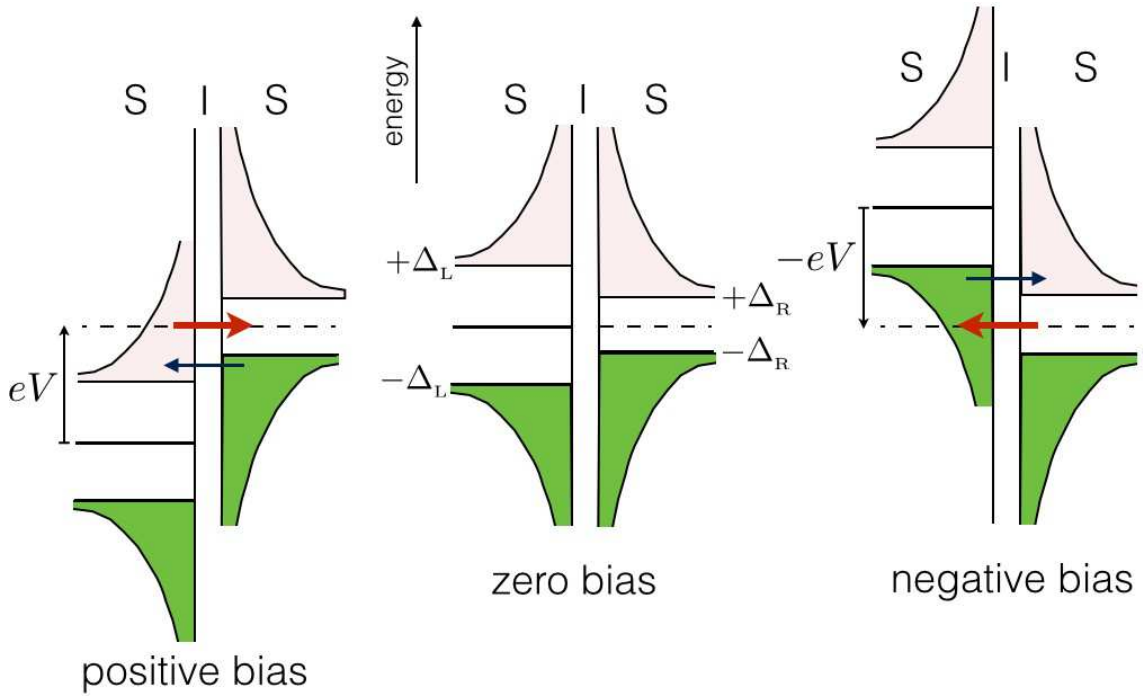


Figure 13.4: SIS tunneling for positive bias (left), zero bias (center), and negative bias (right). Green regions indicate occupied electronic states in each superconductor, where $\widetilde{n}_{\rm s}(\mathcal{E}) > 0$.

and second kinds, respectively:

$$\begin{aligned}\mathbb{K}(x) &= \int\limits_0^{\pi/2} \frac{d\theta}{\sqrt{1-x^2\sin^2\theta}} \\ \mathbb{E}(x) &= \int\limits_0^{\pi/2} d\theta\,\sqrt{1-x^2\sin^2\theta} \quad .\end{aligned} \tag{13.80}$$

We may also make progress by setting $eV = \Delta_{\rm L} + \Delta_{\rm R} + e\,\delta V$. One then has

$$I_{\rm N}(V^* + \delta V) = \frac{G_{\rm N}}{e}\int\limits_0^\infty d\xi_{\rm L}\int\limits_0^\infty d\xi_{\rm R}\;\delta\!\left(e\,\delta V - \frac{\xi_{\rm L}^2}{2\Delta_{\rm L}} - \frac{\xi_{\rm R}^2}{2\Delta_{\rm R}}\right) = \frac{\pi G_{\rm N}}{2e}\sqrt{\Delta_{\rm L}\,\Delta_{\rm R}} \quad . \tag{13.81}$$

Thus, the SIS tunnel current jumps discontinuously at $V = V^*$. At finite temperature, there is a smaller local maximum in $I_{\rm N}$ for $V = |\Delta_{\rm L} - \Delta_{\rm R}|/e$.

13.2.3 The Josephson pair tunneling current $I_{\rm J}$

Earlier we obtained the expression

$$I_{\rm J}(t) = \frac{e}{\hbar^2}\int\limits_{-\infty}^{\infty} dt'\,\Theta(t-t')\left\{e^{+i\Omega(t+t')}\left\langle\left[\hat H_{\rm T}^-(t)\,,\,\hat H_{\rm T}^-(t')\right]\right\rangle - e^{-i\Omega(t+t')}\left\langle\left[\hat H_{\rm T}^+(t)\,,\,\hat H_{\rm T}^+(t')\right]\right\rangle\right\} \quad . \tag{13.82}$$

Proceeding in analogy to the case for $I_{\rm N}$, define now the anomalous response functions,

$$\begin{aligned}\mathcal{Y}_{\sf r}(t-t') &= -i\,\Theta(t-t')\left\langle\left[\hat H_{\rm T}^+(t)\,,\,\hat H_{\rm T}^+(t')\right]\right\rangle \\ \mathcal{Y}_{\sf a}(t-t') &= -i\,\Theta(t-t')\left\langle\left[\hat H_{\rm T}^-(t')\,,\,\hat H_{\rm T}^-(t)\right]\right\rangle \quad .\end{aligned} \tag{13.83}$$

The spectral representations of these response functions are

$$\begin{aligned}\widetilde{\mathcal{Y}}_{\sf r}(\omega) &= \sum_{m,n} P_m\left\{\frac{\langle\, m\,|\,\hat H_{\rm T}^+\,|\,n\,\rangle\langle\, n\,|\,\hat H_{\rm T}^+\,|\,m\,\rangle}{\omega+\omega_m-\omega_n+i\epsilon} - \frac{\langle\, m\,|\,\hat H_{\rm T}^+\,|\,n\,\rangle\langle\, n\,|\,\hat H_{\rm T}^+\,|\,m\,\rangle}{\omega-\omega_m+\omega_n+i\epsilon}\right\} \\ \widetilde{\mathcal{Y}}_{\sf a}(\omega) &= \sum_{m,n} P_m\left\{\frac{\langle\, m\,|\,\hat H_{\rm T}^-\,|\,n\,\rangle\langle\, n\,|\,\hat H_{\rm T}^-\,|\,m\,\rangle}{\omega-\omega_m+\omega_n+i\epsilon} - \frac{\langle\, m\,|\,\hat H_{\rm T}^-\,|\,n\,\rangle\langle\, n\,|\,\hat H_{\rm T}^-\,|\,m\,\rangle}{\omega+\omega_m-\omega_n+i\epsilon}\right\} \quad ,\end{aligned} \tag{13.84}$$

from which we see $\widetilde{\mathcal{Y}}_{\sf a}(\omega) = -\widetilde{\mathcal{Y}}_{\sf r}^*(-\omega)$. The Josephson current is then given by

$$\begin{aligned}I_{\rm J}(t) &= -\frac{ie}{\hbar^2}\int\limits_{-\infty}^{\infty} dt'\left\{e^{-2i\Omega t}\,\mathcal{Y}_{\sf r}(t-t')\,e^{+i\Omega(t-t')} + e^{+2i\Omega t}\,\mathcal{Y}_{\sf a}(t-t')\,e^{-i\Omega(t-t')}\right\} \\ &= \frac{2e}{\hbar^2}\,\mathsf{Im}\Big[e^{-2i\Omega t}\,\widetilde{\mathcal{Y}}_{\sf r}(\Omega)\Big] \quad ,\end{aligned} \tag{13.85}$$

where $\Omega = eV/\hbar$.

Plugging in our expressions for $\hat H_{\rm T}^{\pm}$, we have

$$\begin{aligned}\mathcal{Y}_{\sf r}(t) &= -i\,\Theta(t)\sum_{\boldsymbol{k},\boldsymbol{q},\sigma} T_{\boldsymbol{k},\boldsymbol{q}}\,T_{-\boldsymbol{k},-\boldsymbol{q}}\left\langle\left[c^\dagger_{{\rm L}\,\boldsymbol{q}\,\sigma}(t)\,c_{{\rm R}\,\boldsymbol{k}\,\sigma}(t)\,,\,c^\dagger_{{\rm L}\,-\boldsymbol{q}\,-\sigma}(0)\,c_{{\rm R}\,-\boldsymbol{k}\,-\sigma}(0)\right]\right\rangle \\ &= 2i\,\Theta(t)\sum_{\boldsymbol{q},\boldsymbol{k}} T_{\boldsymbol{k},\boldsymbol{q}}\,T_{-\boldsymbol{k},-\boldsymbol{q}}\left\{\left\langle c^\dagger_{{\rm L}\,\boldsymbol{q}\uparrow}(t)\,c^\dagger_{{\rm L}\,-\boldsymbol{q}\downarrow}(0)\right\rangle\left\langle c_{{\rm R}\,\boldsymbol{k}\uparrow}(t)\,c_{{\rm R}\,-\boldsymbol{k}\downarrow}(0)\right\rangle\right. \\ &\qquad\qquad \left. - \left\langle c^\dagger_{{\rm L}\,-\boldsymbol{q}\downarrow}(0)\,c^\dagger_{{\rm L}\,\boldsymbol{q}\uparrow}(t)\right\rangle\left\langle c_{{\rm R}\,-\boldsymbol{k}\downarrow}(0)\,c_{{\rm R}\,\boldsymbol{k}\uparrow}(t)\right\rangle\right\} \quad .\end{aligned} \tag{13.86}$$

Again we invoke Bogoliubov,

$$c_{\boldsymbol{k}\uparrow} = u_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\uparrow} - v_{\boldsymbol{k}}\,e^{i\phi}\,\gamma^\dagger_{-\boldsymbol{k}\downarrow} \qquad\qquad c^\dagger_{\boldsymbol{k}\uparrow} = u_{\boldsymbol{k}}\,\gamma^\dagger_{\boldsymbol{k}\uparrow} - v_{\boldsymbol{k}}\,e^{-i\phi}\,\gamma_{-\boldsymbol{k}\downarrow} \tag{13.87}$$

$$c_{-\boldsymbol{k}\downarrow} = u_{\boldsymbol{k}}\,\gamma_{-\boldsymbol{k}\downarrow} + v_{\boldsymbol{k}}\,e^{i\phi}\,\gamma^\dagger_{\boldsymbol{k}\uparrow} \qquad\qquad c^\dagger_{-\boldsymbol{k}\downarrow} = u_{\boldsymbol{k}}\,\gamma^\dagger_{-\boldsymbol{k}\downarrow} + v_{\boldsymbol{k}}\,e^{-i\phi}\,\gamma_{\boldsymbol{k}\uparrow} \tag{13.88}$$

to obtain

$$\begin{aligned}
\big\langle c^\dagger_{{\rm L}\,\boldsymbol{q}\uparrow}(t)\,c^\dagger_{{\rm L}\,-\boldsymbol{q}\downarrow}(0)\big\rangle &= u_{\boldsymbol{q}}\,v_{\boldsymbol{q}}\,e^{-i\phi_{\rm L}}\left\{e^{iE_{\boldsymbol{q}}t/\hbar}\,f(E_{\boldsymbol{q}}) - e^{-iE_{\boldsymbol{q}}t/\hbar}\left[1-f(E_{\boldsymbol{q}})\right]\right\} \\
\big\langle c_{{\rm R}\,\boldsymbol{k}\uparrow}(t)\,c_{{\rm R}\,-\boldsymbol{k}\downarrow}(0)\big\rangle &= u_{\boldsymbol{k}}\,v_{\boldsymbol{k}}\,e^{+i\phi_{\rm R}}\left\{e^{-iE_{\boldsymbol{k}}t/\hbar}\left[1-f(E_{\boldsymbol{k}})\right] - e^{iE_{\boldsymbol{k}}t/\hbar}\,f(E_{\boldsymbol{k}})\right\} \\
\big\langle c^\dagger_{{\rm L}\,-\boldsymbol{q}\downarrow}(0)\,c^\dagger_{{\rm L}\,\boldsymbol{q}\uparrow}(t)\big\rangle &= u_{\boldsymbol{q}}\,v_{\boldsymbol{q}}\,e^{-i\phi_{\rm L}}\left\{e^{iE_{\boldsymbol{q}}t/\hbar}\left[1-f(E_{\boldsymbol{q}})\right] - e^{-iE_{\boldsymbol{q}}t/\hbar}\,f(E_{\boldsymbol{q}})\right\} \\
\big\langle c_{{\rm R}\,-\boldsymbol{k}\downarrow}(0)\,c_{{\rm R}\,\boldsymbol{k}\uparrow}(t)\big\rangle &= u_{\boldsymbol{k}}\,v_{\boldsymbol{k}}\,e^{+i\phi_{\rm R}}\left\{e^{-iE_{\boldsymbol{k}}t/\hbar}\,f(E_{\boldsymbol{k}}) - e^{iE_{\boldsymbol{k}}t/\hbar}\left[1-f(E_{\boldsymbol{k}})\right]\right\}
\end{aligned} \tag{13.89}$$

We then have

$$\begin{aligned}
\mathcal{Y}_{\sf r}(t) = i\,\Theta(t) \times \tfrac{1}{2}\,g_{\rm L}\,g_{\rm R}\,|t|^2 A\,e^{i(\phi_{\rm R}-\phi_{\rm L})}\int_{-\infty}^{\infty}\!\!d\xi\int_{-\infty}^{\infty}\!\!d\xi'\;u\,v\,u'\,v'\,\times \\
\left\{\left[e^{iEt/\hbar}\,f - e^{-iEt/\hbar}\,(1-f)\right]\times\left[e^{-iE't/\hbar}\,(1-f') - e^{iE't/\hbar}\,f'\right]\right. \\
\left. -\left[e^{iEt/\hbar}\,(1-f) - e^{-iEt/\hbar}\,f\right]\times\left[e^{-iE't/\hbar}\,f' - e^{iE't/\hbar}\,(1-f')\right]\right\} \quad ,
\end{aligned} \tag{13.90}$$

where once again primed and unprimed symbols refer respectively to left (L) and right (R) banks. Recall that the BCS coherence factors give $uv = \frac{1}{2}\sin(2\vartheta) = \Delta/2E$. Taking the Fourier transform, we have

$$\begin{aligned}
\widetilde{\mathcal{Y}}_{\sf r}(\omega) = \tfrac{1}{2}\,\hbar\,g_{\rm L}\,g_{\rm R}\,|t|^2\,e^{i(\phi_{\rm R}-\phi_{\rm L})}A\int_0^\infty\!\!d\xi\int_0^\infty\!\!d\xi'\;\frac{\Delta}{E}\,\frac{\Delta'}{E'}\left\{\frac{f-f'}{\hbar\omega+E-E'+i\epsilon} - \frac{f-f'}{\hbar\omega-E+E'+i\epsilon}\right. \\
\left. +\frac{1-f-f'}{\hbar\omega+E+E'+i\epsilon} - \frac{1-f-f'}{\hbar\omega-E-E'+i\epsilon}\right\} \quad .
\end{aligned} \tag{13.91}$$

Setting $T=0$, we have

$$\begin{aligned}
\widetilde{\mathcal{Y}}_{\sf r}(\omega) &= \frac{\hbar^2 G_{\rm N}}{2\pi e^2}\,e^{i(\phi_{\rm R}-\phi_{\rm L})}\int_0^\infty\!\!d\xi\int_0^\infty\!\!d\xi'\;\frac{\Delta\Delta'}{EE'}\left\{\frac{1}{\hbar\omega+E+E'+i\epsilon} - \frac{1}{\hbar\omega-E-E'+i\epsilon}\right\} \\
&= \frac{\hbar^2 G_{\rm N}}{2\pi e^2}\;e^{i(\phi_{\rm R}-\phi_{\rm L})}\int_\Delta^\infty\!\!dE\;\frac{\Delta}{\sqrt{E^2-\Delta^2}}\int_{\Delta'}^\infty\!\!dE'\;\frac{\Delta'}{\sqrt{E'^2-\Delta'^2}}\times\frac{2\,(E+E')}{(\hbar\omega)^2-(E+E')^2} \quad .
\end{aligned} \tag{13.92}$$

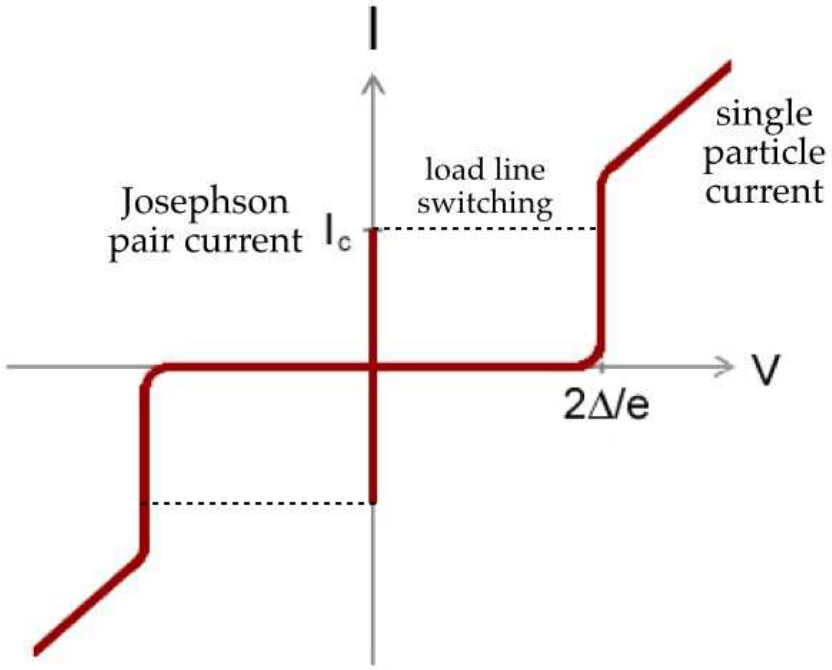


Figure 13.5: Current-voltage characteristics for a current-biased Josephson junction. Increasing current at zero bias voltage is possible up to $|I| = I_{\rm c}$, beyond which the voltage jumps along the dotted line. Subsequent reduction in current leads to hysteresis.

There is no general analytic form for this integral. However, for the special case $\Delta = \Delta'$, we have

$$\widetilde{\mathcal{Y}}_{\rm r}(\omega) = \frac{G_{\rm N}\hbar^2}{2e^2}\,\Delta\,\mathbb{K}\left(\frac{\hbar|\omega|}{4\Delta}\right)e^{i(\phi_{\rm R}-\phi_{\rm L})} \quad , \tag{13.93}$$

where $\mathbb{K}(x)$ is the complete elliptic integral of the first kind. Thus,

$$I_{\rm J}(t) = G_{\rm N}\cdot\frac{\Delta}{e}\,\mathbb{K}\left(\frac{e|V|}{4\Delta}\right)\,\sin\left(\phi_{\rm R}-\phi_{\rm L}-\frac{2eVt}{\hbar}\right) \quad . \tag{13.94}$$

With $V = 0$, one finds (at finite T),

$$I_{\rm J} = G_{\rm N}\cdot\frac{\pi\Delta}{2e}\,\tanh\left(\frac{\Delta}{2k_{\rm B}T}\right)\,\sin(\phi_{\rm R}-\phi_{\rm L}) \quad . \tag{13.95}$$

Thus, there is a spontaneous current flow in the absence of any voltage bias, provided the phases remain fixed. The maximum current which flows under these conditions is called the *critical current* of the junction, $I_{\rm c}$. Writing $R_{\rm N} = 1/G_{\rm N}$ for the normal state junction resistance, one has

$$I_{\rm c}R_{\rm N} = \frac{\pi\Delta}{2e}\,\tanh\left(\frac{\Delta}{2k_{\rm B}T}\right) \quad , \tag{13.96}$$

which is known as the *Ambegaokar-Baratoff relation*. Note that $I_{\rm c}$ agrees with what we found in Eqn. 13.81 for V just above $V^* = 2\Delta$. $I_{\rm c}$ is also the current flowing in a normal junction at bias voltage $V = \pi\Delta/2e$. Setting $I_{\rm c} = 2eJ/\hbar$ where J is the Josephson coupling, we find our $V = 0$ results here in complete agreement with those of Eqn. 13.29 when Coulomb charging energies of the grains are neglected.

Experimentally, one generally draws a current I across the junction and then measures the voltage difference. In other words, the junction is *current-biased*. Varying I then leads to a hysteretic voltage response, as shown in Fig. 13.5. The functional form of the oscillating current is then $I(t) = I_{\rm c}\sin(\phi_{\rm R} - \phi_{\rm L} - \Omega t)$, which gives no DC average. With $R_{\rm N} \approx 1.5\,\Omega$ and $\Delta = 1\,$meV, one obtains a critical current $I_{\rm c} = 1\,$mA. For a junction of area $A \sim 1\,{\rm mm}^2$, the critical current density is then $j_{\rm c} = I_{\rm c}/A \sim 10^3\,{\rm A/m}^2$. Current densities in bulk type I and type II materials can approach $j \sim 10^{11}\,{\rm A/m}^2$ and $10^9\,{\rm A/m}^2$, respectively.

13.3 The Josephson Effect

13.3.1 Two grain junction

In §13.1 we discussed a model for superconducting grains. Consider now only a single pair of grains, and write

$$\hat{K} = -J\cos(\phi_{\rm L} - \phi_{\rm R}) + \frac{2e^2}{C_{\rm L}}\,M_{\rm L}^2 + \frac{2e^2}{C_{\rm R}}\,M_{\rm R}^2 - 2\mu_{\rm L}M_{\rm L} - 2\mu_{\rm R}M_{\rm R} \quad , \tag{13.97}$$

where $M_{\rm L,R}$ is the number of Cooper pairs on each grain in excess of the background charge, which we assume here to be a multiple of $2e$. From the Heisenberg equations of motion, we have that

$$\dot{M}_{\rm L} = \frac{i}{\hbar}\left[\hat{K}, M_{\rm L}\right] = \frac{J}{\hbar}\,\sin(\phi_{\rm R} - \phi_{\rm L}) \quad , \tag{13.98}$$

which follows from the fact that $M_{\rm L} = -i\,\partial/\partial\phi_{\rm L}$. Similarly, we find $\dot{M}_{\rm R} = -\frac{J}{\hbar}\sin(\phi_{\rm R} - \phi_{\rm L})$. An electrical current $I = 2e\dot{M}_{\rm L} = -2e\dot{M}_{\rm R}$ then flows from left to right. The equations of motion for the phases are

$$\begin{aligned}\dot{\phi}_{\rm L} &= \frac{i}{\hbar}\left[\hat{K}\,,\,\phi_{\rm L}\right] = \frac{4e^2M_{\rm L}}{\hbar C_{\rm L}} - \frac{2\mu_{\rm L}}{\hbar} \\ \dot{\phi}_{\rm R} &= \frac{i}{\hbar}\left[\hat{K}\,,\,\phi_{\rm R}\right] = \frac{4e^2M_{\rm R}}{\hbar C_{\rm R}} - \frac{2\mu_{\rm R}}{\hbar} \quad .\end{aligned} \tag{13.99}$$

Let's assume the grains are large, so their self-capacitances are large too. In that case, we can neglect the Coulomb energy of each grain, and we obtain the *Josephson equations*

$$\frac{d\phi}{dt} = -\frac{2eV}{\hbar} \qquad , \qquad I(t) = I_{\rm c}\sin\phi(t) \quad , \tag{13.100}$$

where $eV = \mu_{\rm R} - \mu_{\rm L}$, $I_{\rm c} = 2eJ/\hbar$, and $\phi \equiv \phi_{\rm R} - \phi_{\rm L}$. When quasiparticle tunneling is accounted for, the second of the Josephson equations is modified to

$$I = I_{\rm c} \sin\phi + \big(G_0 + G_1 \cos\phi\big) V \quad , \tag{13.101}$$

where $G_0 \equiv G_{\rm N}$ is the quasiparticle contribution to the current, and G_1 accounts for higher order effects.

13.3.2 Effect of in-plane magnetic field

Thus far we have assumed that the effective hopping amplitude t between the L and R banks is real. This is valid in the absence of an external magnetic field, which breaks time-reversal. In the presence of an external magnetic field, t is replaced by $t \to t\, e^{i\gamma}$, where $\gamma = \frac{e}{\hbar c} \int_{\rm L}^{\rm R} \boldsymbol{A} \cdot d\boldsymbol{l}$ is the Aharonov-Bohm phase. Without loss of generality, we consider the junction interface to lie in the (x, y) plane, and we take $\boldsymbol{H} = H\hat{\boldsymbol{y}}$. We are then free to choose the gauge $\boldsymbol{A} = -Hx\hat{\boldsymbol{z}}$. Then

$$\gamma(x) = \frac{e}{\hbar c} \int_{\rm L}^{\rm R} \boldsymbol{A} \cdot d\boldsymbol{l} = -\frac{e}{\hbar c}\, H\, (\lambda_{\rm L} + \lambda_{\rm R} + d)\, x \quad , \tag{13.102}$$

where $\lambda_{\rm L,R}$ are the penetration depths for the two superconducting banks, and d is the junction separation. Typically $\lambda_{\rm L,R} \sim 100\,\text{Å} - 1000\,\text{Å}$, while $d \sim 10\,\text{Å}$, so usually we may neglect the junction separation in comparison with the penetration depth.

In the case of the single particle current $I_{\rm N}$, we needed the commutators $\big[\hat{H}_{\rm T}^+(t), \hat{H}_{\rm T}^-(0)\big]$ and $\big[\hat{H}_{\rm T}^-(t), \hat{H}_{\rm T}^+(0)\big]$. Since $\hat{H}_{\rm T}^+ \propto t$ while $\hat{H}_{\rm T}^- \propto t^*$, the result depends on the product $|t|^2$, which has no phase. Thus, $I_{\rm N}$ is unaffected by an in-plane magnetic field. For the Josephson pair tunneling current $I_{\rm J}$, however, we need $\big[\hat{H}_{\rm T}^+(t), \hat{H}_{\rm T}^+(0)\big]$ and $\big[\hat{H}_{\rm T}^-(t), \hat{H}_{\rm T}^-(0)\big]$. The former is proportional to t^2 and the latter to t^{*2}. Therefore the Josephson current density is

$$j_{\rm J}(x) = \frac{I_{\rm c}(T)}{A}\, \sin\!\left(\phi - \frac{2e}{\hbar c}\, H d_{\rm eff} x - \frac{2eVt}{\hbar}\right), \tag{13.103}$$

where $d_{\rm eff} \equiv \lambda_{\rm L} + \lambda_{\rm R} + d$ and $\phi = \phi_{\rm R} - \phi_{\rm L}$. Note that it is $2eHd_{\rm eff}/\hbar c = \arg(t^2)$ which appears in the argument of the sine. This may be interpreted as the Aharonov-Bohm phase accrued by a tunneling Cooper pair. We now assume our junction interface is a square of dimensions $L_x \times L_y$. At $V = 0$, the total Josephson current is then[5]

$$I_{\rm J} = \int_0^{L_x} dx \int_0^{L_y} dy\; j(x) = \frac{I_{\rm c}\phi_{\rm L}}{\pi\Phi}\, \sin(\pi\Phi/\phi_{\rm L}) \sin(\phi - \pi\Phi/\phi_{\rm L}) \quad , \tag{13.104}$$

[5]Take care not to confuse $\phi_{\rm L}$, the phase of the left superconducting bank, with $\phi_{\rm L}$, the London flux quantum $hc/2e$. To the untrained eye, these symbols look identical.

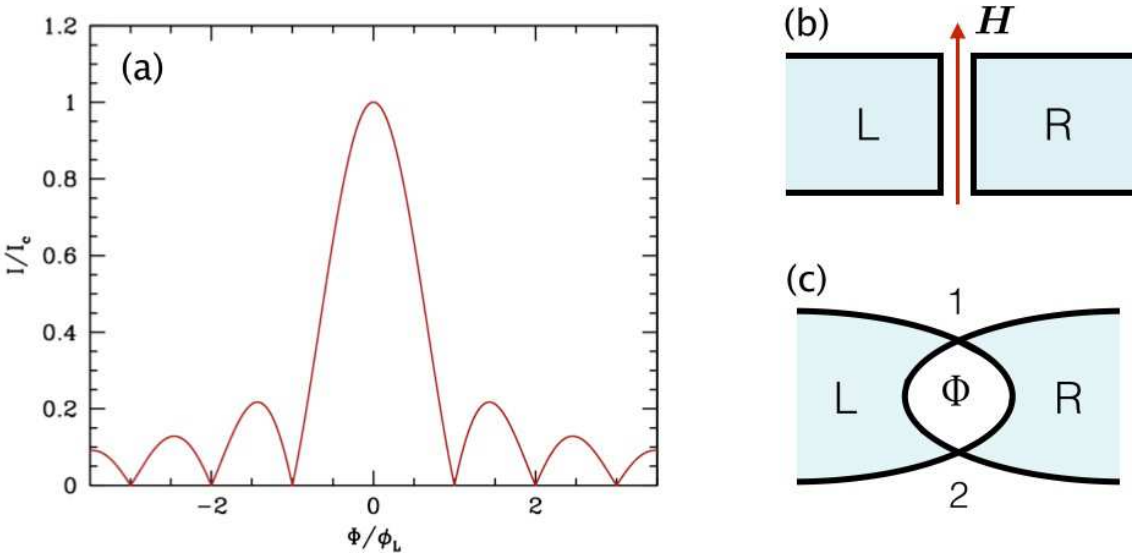


Figure 13.6: (a) Fraunhofer pattern of Josephson current *versus* flux due to in-plane magnetic field. (b) Sketch of Josephson junction experiment yielding (a). (c) Two-point superconducting quantum interferometer.

where $\Phi \equiv HL_x d_{\rm eff}$ is the total magnetic flux through the junction. The maximum current occurs when $\phi - \pi\Phi/\phi_{\rm L} = \pm\frac{1}{2}\pi$, where its magnitude is

$$I_{\rm max}(\Phi) = I_{\rm c}\left|\frac{\sin(\pi\Phi/\phi_{\rm L})}{\pi\Phi/\phi_{\rm L}}\right| \quad . \tag{13.105}$$

The shape $I_{\rm max}(\Phi)$ is precisely that of the single slit Fraunhofer pattern from geometrical optics! (See Fig. 13.6.)

13.3.3 Two-point quantum interferometer

Consider next the device depicted in Fig. 13.6(c) consisting of two weak links between superconducting banks. The current flowing from L to R is

$$I = I_{\rm c,1}\,\sin\phi_1 + I_{\rm c,2}\,\sin\phi_2 \quad . \tag{13.106}$$

where $\phi_1 \equiv \phi_{\rm L,1} - \phi_{\rm R,1}$ and $\phi_2 \equiv \phi_{\rm L,2} - \phi_{\rm R,2}$ are the phase differences across the two Josephson junctions. The total flux Φ inside the enclosed loop is

$$\phi_2 - \phi_1 = \frac{2\pi\Phi}{\phi_{\rm L}} \equiv 2\gamma \quad . \tag{13.107}$$

Writing $\phi_2 = \phi_1 + 2\gamma$, we extremize $I(\phi_1,\gamma)$ with respect to ϕ_1, and obtain

$$I_{\rm max}(\gamma) = \sqrt{(I_{\rm c,1}+I_{\rm c,2})^2\cos^2\gamma + (I_{\rm c,1}-I_{\rm c,2})^2\sin^2\gamma} \quad . \tag{13.108}$$

If $I_{c,1} = I_{c,2}$, we have $I_{max}(\gamma) = 2I_c\,|\cos\gamma\,|$. This provides for an extremely sensitive measurement of magnetic fields, since $\gamma = \pi\Phi/\phi_L$ and $\phi_L = 2.07 \times 10^{-7}\,\mathrm{G\,cm^2}$. Thus, a ring of area $1\,\mathrm{cm^2}$ allows for the detection of fields on the order of $10^{-7}\,\mathrm{G}$. This device is known as a Superconducting QUantum Interference Device, or SQUID. The limits of the SQUID's sensitivity are set by the noise in the SQUID or in the circuit amplifier.

13.3.4 RCSJ Model

In circuits, a Josephson junction, from a practical point of view, is always transporting current in parallel to some resistive channel. Josephson junctions also have electrostatic capacitance as well. Accordingly, consider the *resistively and capacitively shunted Josephson junction* (RCSJ), a sketch of which is provided in Fig. 13.8(c). The equations governing the RCSJ model are

$$\begin{aligned} I &= C\,\dot{V} + \frac{V}{R} + I_c \sin\phi \\ V &= \frac{\hbar}{2e}\,\dot{\phi} \quad , \end{aligned} \tag{13.109}$$

where we again take I to run from left to right. If the junction is *voltage-biased*, then integrating the second of these equations yields $\phi(t) = \phi_0 + \omega_J t$, where $\omega_J = 2eV/\hbar$ is the *Josephson frequency*. The current is then

$$I = \frac{V}{R} + I_c \sin(\phi_0 + \omega_J t) \quad . \tag{13.110}$$

If the junction is *current-biased*, then we substitute the second equation into the first, to obtain

$$\frac{\hbar C}{2e}\,\ddot{\phi} + \frac{\hbar}{2eR}\,\dot{\phi} + I_c \sin\phi = I \quad . \tag{13.111}$$

We adimensionalize by writing $s \equiv \omega_p t$, with $\omega_p = (2eI_c/\hbar C)^{1/2}$ is the *Josephson plasma frequency* (at zero current). We then have

$$\frac{d^2\phi}{ds^2} + \frac{1}{Q}\,\frac{d\phi}{ds} = j - \sin\phi \equiv -\frac{du}{d\phi} \quad , \tag{13.112}$$

where $Q = \omega_p \tau$ with $\tau = RC$, and $j = I/I_c$. The quantity Q^2 is called the *McCumber-Stewart parameter*. The resistance is $R(T \approx T_c) = R_N$, while $R(T \ll T_c) \approx R_N\,\exp(\Delta/k_B T)$. The dimensionless potential energy $u(\phi)$ is given by

$$u(\phi) = -j\phi - \cos\phi \tag{13.113}$$

and resembles a 'tilted washboard'; see Fig. 13.8(a,b). This is an $N = 2$ dynamical system on a cylinder. Writing $\omega \equiv \dot{\phi}$, we have

$$\frac{d}{ds}\begin{pmatrix}\phi\\ \omega\end{pmatrix} = \begin{pmatrix}\omega\\ j - \sin\phi - Q^{-1}\omega\end{pmatrix} \quad . \tag{13.114}$$

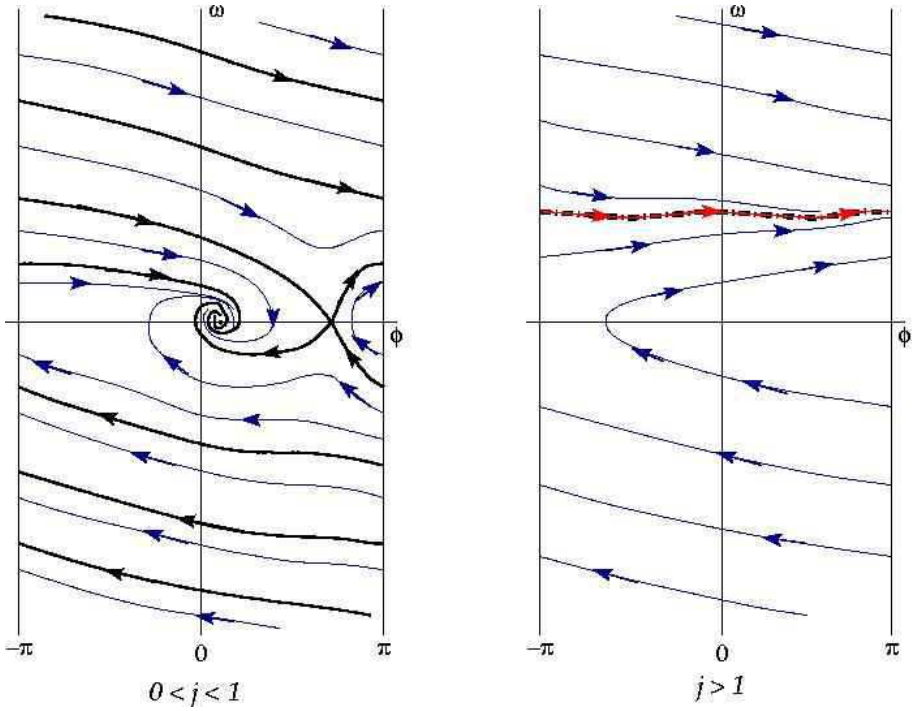


Figure 13.7: Phase flows for the equation $\ddot{\phi} + Q^{-1}\dot{\phi} + \sin\phi = j$. Left panel: $0 < j < 1$; note the separatrix (in black), which flows into the stable and unstable fixed points. Right panel: $j > 1$. The red curve overlying the thick black dot-dash curve is a *limit cycle*.

Note that $\phi \in [0, 2\pi]$ while $\omega \in (-\infty, \infty)$. Fixed points satisfy $\omega = 0$ and $j = \sin\phi$. Thus, for $|j| > 1$, there are no fixed points.

Strong damping

The RCSJ model dynamics are given by the second order ODE,

$$\partial_s^2\phi + Q^{-1}\partial_s\phi = -u'(\phi) = j - \sin\phi \quad . \tag{13.115}$$

The parameter $Q = \omega_{\rm p}\tau$ determines the damping, with large Q corresponding to small damping. Consider the large damping limit $Q \ll 1$. In this case the inertial term proportional to $\ddot{\phi}$ may be ignored, and what remains is a first order ODE. Restoring dimensions,

$$\frac{d\phi}{dt} = \Omega\,(j - \sin\phi) \qquad , \tag{13.116}$$

where $\Omega = \omega_{\rm p}^2 RC = 2eI_{\rm c}R/\hbar$. We are effectively setting $C \equiv 0$, hence this is known as the RSJ model. The above equation describes a $N = 1$ dynamical system on the circle. When $|j| < 1$, *i.e.* $|I| < I_c$, there are two fixed points, which are solutions to $\sin\phi^* = j$. The fixed point where $\cos\phi^* > 0$ is stable, while that with $\cos\phi^* < 0$ is unstable. The flow is toward the stable fixed

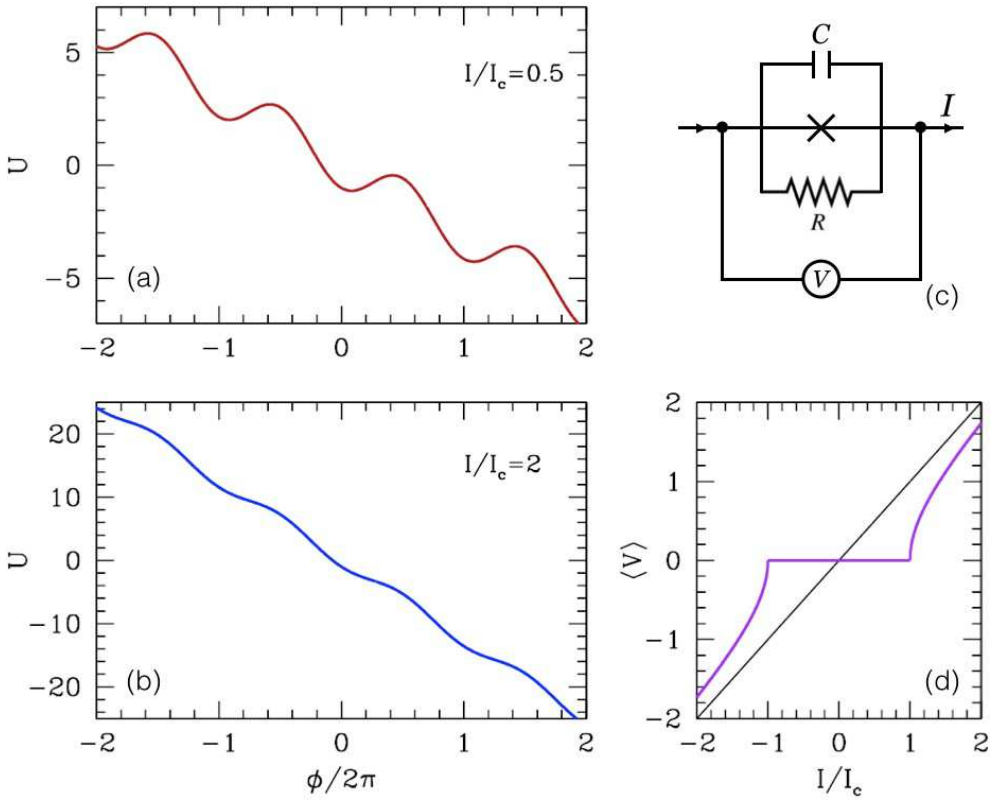


Figure 13.8: (a) Dimensionless washboard potential $u(\phi)$ for $I/I_{\rm c} = 0.5$. (b) $u(\phi)$ for $I/I_{\rm c} = 2.0$. (c) The resistively and capacitively shunted Josephson junction (RCSJ). (d) $\langle V\rangle$ *versus* I for the RSJ model.

point. At the fixed point, ϕ is constant, which means the voltage $V = \hbar\dot\phi/2e$ vanishes. There is current flow with no potential drop.

Consider the case $j > 1$. In this case there is a bottleneck in the ϕ evolution in the vicinity of $\phi = \frac{1}{2}\pi$, where $\dot\phi$ is smallest, but $\dot\phi > 0$ always. We compute the average voltage

$$\langle V\rangle = \frac{\hbar}{2e}\,\langle\dot\phi\rangle = \frac{\hbar}{2e}\cdot\frac{2\pi}{T} \quad , \tag{13.117}$$

where T is the rotational period for $\phi(t)$. We compute this using the equation of motion:

$$\Omega T = \int_0^{2\pi}\frac{d\phi}{j-\sin\phi} = \frac{2\pi}{\sqrt{j^2-1}} \quad . \tag{13.118}$$

Thus,

$$\langle V\rangle = \frac{\hbar}{2e}\sqrt{j^2-1}\cdot\frac{2eI_{\rm c}R}{\hbar} = R\sqrt{I^2-I_{\rm c}^2} \quad . \tag{13.119}$$

This behavior is sketched in Fig. 13.8(d).

Josephson plasma oscillations

When $I < I_{\rm c}$, the phase undergoes damped oscillations in the washboard minima. Expanding about the fixed point, we write $\phi = \sin^{-1} j + \delta\phi$, and obtain

$$\frac{d^2\delta\phi}{ds^2} + \frac{1}{Q}\frac{d\,\delta\phi}{ds} = -\sqrt{1-j^2}\,\delta\phi \quad . \tag{13.120}$$

This is the equation of a damped harmonic oscillator. With no damping ($Q = \infty$), the oscillation frequency is

$$\Omega(I) = \omega_{\rm p}\left(1 - \frac{I^2}{I_{\rm c}^2}\right)^{1/4} \quad . \tag{13.121}$$

When Q is finite, the frequency of the oscillations has an imaginary component, with solutions

$$\omega_\pm(I) = -\frac{i\,\omega_{\rm p}}{2Q} \pm \omega_{\rm p}\sqrt{\left(1 - \frac{I^2}{I_{\rm c}^2}\right)^{1/2} - \frac{1}{4Q^2}} \quad . \tag{13.122}$$

Retrapping current in underdamped junctions

The energy of the junction is given by

$$E = \tfrac{1}{2}CV^2 + \frac{\hbar I_{\rm c}}{2e}\,(1 - \cos\phi) \quad . \tag{13.123}$$

The first term may be thought of as a kinetic energy and the second as potential energy. Because the system is dissipative, energy is not conserved. Rather,

$$\dot E = CV\dot V + \frac{\hbar I_{\rm c}}{2e}\,\dot\phi\,\sin\phi = V\big(C\dot V + I_{\rm c}\sin\phi\big) = V\left(I - \frac{V}{R}\right) \quad . \tag{13.124}$$

Suppose the junction were completely undamped, *i.e.* $R = 0$. Then as the phase slides down the tilted washboard for $|I| < I_{\rm c}$, it moves from peak to peak, picking up speed as it moves along. When $R > 0$, there is energy loss, and $\phi(t)$ might not make it from one peak to the next. Suppose we start at a local maximum $\phi = \pi$ with $V = 0$. What is the energy when ϕ reaches 3π? To answer that, we assume that energy is almost conserved, so

$$E = \tfrac{1}{2}CV^2 + \frac{\hbar I_{\rm c}}{2e}\,(1 - \cos\phi) \approx \frac{\hbar I_{\rm c}}{e} \quad \Rightarrow \quad V = \left(\frac{e\hbar I_{\rm c}}{eC}\right)^{1/2}\big|\cos(\tfrac{1}{2}\phi)\big| \quad . \tag{13.125}$$

then

$$\begin{aligned}(\Delta E)_{\rm cycle} &= \int\limits_{-\infty}^{\infty} dt\, V\left(I - \frac{V}{R}\right) = \frac{\hbar}{2e}\int\limits_{-\pi}^{\pi} d\phi\,\left\{I - \frac{1}{R}\left(\frac{e\hbar I_{\rm c}}{eC}\right)^{1/2}\cos(\tfrac{1}{2}\phi)\right\} \\ &= \frac{\hbar}{2e}\left\{2\pi I - \frac{4}{R}\left(\frac{e\hbar I_{\rm c}}{eC}\right)^{1/2}\right\} = \frac{h}{2e}\left\{I - \frac{4I_{\rm c}}{\pi Q}\right\} \quad .\end{aligned} \tag{13.126}$$

Thus, we identify $I_{\rm r} \equiv 4I_{\rm c}/\pi Q \ll I_{\rm c}$ as the *retrapping current*. The idea here is to focus on the case where the phase evolution is on the cusp between trapped and free. If the system loses energy over the cycle, then subsequent motion will be attenuated, and the phase dynamics will flow to the zero voltage fixed point. Note that if the current I is reduced below $I_{\rm c}$ and then held fixed, eventually the junction will dissipate energy and enter the zero voltage state for any $|I| < I_{\rm c}$. But if the current is swept and $\dot I/I$ is faster than the rate of energy dissipation, the retrapping occurs at $I = I_{\rm r}$.

Thermal fluctuations

Restoring the proper units, the potential energy is $U(\phi) = (\hbar I_{\rm c}/2e)\,u(\phi)$. Thus, thermal fluctuations may be ignored provided

$$k_{\rm B}T \ll \frac{\hbar I_{\rm c}}{2e} = \frac{\hbar}{2eR_{\rm N}} \cdot \frac{\pi\Delta}{2e}\,\tanh\!\left(\frac{\Delta}{2k_{\rm B}T}\right) \quad , \tag{13.127}$$

where we have invoked the Ambegaokar-Baratoff formula, Eqn. 13.96. BCS theory gives $\Delta = 1.764\,k_{\rm B}T_{\rm c}$, so we require

$$k_{\rm B}T \ll \frac{h}{8R_{\rm N}e^2} \cdot (1.764\,k_{\rm B}T_{\rm c}) \cdot \tanh\!\left(\frac{0.882\,T_{\rm c}}{T}\right) \quad . \tag{13.128}$$

In other words,

$$\frac{R_{\rm N}}{R_{\rm K}} \ll \frac{0.22\,T_{\rm c}}{T}\,\tanh\!\left(\frac{0.882\,T_{\rm c}}{T}\right) \quad , \tag{13.129}$$

where $R_{\rm K} = h/e^2 = 25812.8\,\Omega$ is the quantum unit of resistance[6].

We can model the effect of thermal fluctuations by adding a noise term to the RCSJ model, writing

$$C\dot V + \frac{V}{R} + I_{\rm c}\sin\phi = I + \frac{V_{\rm f}}{R} \quad , \tag{13.130}$$

where $V_{\rm f}(t)$ is a stochastic term satisfying

$$\big\langle V_{\rm f}(t)\,V_{\rm f}(t')\big\rangle = 2k_{\rm B}TR\,\delta(t-t') \quad . \tag{13.131}$$

Adimensionalizing, we now have

$$\frac{d^2\phi}{ds^2} + \gamma\,\frac{d\phi}{ds} = -\frac{\partial u}{\partial\phi} + \eta(s) \quad , \tag{13.132}$$

where $s = \omega_{\rm p}t$, $\gamma = 1/\omega_{\rm p}RC$, $u(\phi) = -j\phi - \cos\phi$, $j = I/I_{\rm c}(T)$, and

$$\big\langle \eta(s)\,\eta(s')\big\rangle = \frac{2\omega_{\rm p}k_{\rm B}T}{I_{\rm c}^2R}\,\delta(s-s') \equiv 2\Theta\,\delta(s-s') \quad . \tag{13.133}$$

[6] $R_{\rm K}$ is called the *Klitzing* for Klaus von Klitzing, the discoverer of the integer quantum Hall effect.

Thus, $\Theta \equiv \omega_{\rm p} k_{\rm B} T / I_{\rm c}^2 R$ is a dimensionless measure of the temperature. Our problem is now that of a damped massive particle moving in the washboard potential and subjected to stochastic forcing due to thermal noise.

Writing $\omega = \partial_s \phi$, we have

$$\begin{aligned} \partial_s \phi &= \omega \\ \partial_s \omega &= -u'(\phi) - \gamma\omega + \sqrt{2\Theta}\,\eta(s) \quad . \end{aligned} \tag{13.134}$$

In this case, $W(s) = \int_0^s ds'\,\eta(s')$ describes a Wiener process: $\big\langle W(s)\,W(s')\big\rangle = \min(s,s')$. The probability distribution $P(\phi,\omega,s)$ then satisfies the Fokker-Planck equation[7],

$$\frac{\partial P}{\partial s} = -\frac{\partial}{\partial \phi}\big(\omega P\big) + \frac{\partial}{\partial \omega}\Big\{\big[u'(\phi) + \gamma\omega\big]P\Big\} + \Theta\,\frac{\partial^2 P}{\partial \omega^2} \quad . \tag{13.135}$$

We cannot make much progress beyond numerical work starting from this equation. However, if the mean drift velocity of the 'particle' is everywhere small compared with the thermal velocity $v_{\rm th} \propto \sqrt{\Theta}$, and the mean free path $\ell \propto v_{\rm th}/\gamma$ is small compared with the scale of variation of ϕ in the potential $u(\phi)$, then, following the classic treatment by Kramers, we can convert the Fokker-Planck equation for the distribution $P(\phi,\omega,t)$ to the Smoluchowski equation for the distribution $P(\phi,t)$[8]. These conditions are satisfied when the damping γ is large. To proceed along these lines, simply assume that ω relaxes quickly, so that $\partial_s \omega \approx 0$ at all times. This says $\omega = -\gamma^{-1} u'(\phi) + \gamma^{-1}\sqrt{2\Theta}\,\eta(s)$. Plugging this into $\partial_s \phi = \omega$, we have

$$\partial_s \phi = -\gamma^{-1} u'(\phi) + \gamma^{-1}\sqrt{2\Theta}\,\eta(s) \quad , \tag{13.136}$$

the Fokker-Planck equation for which is[9]

$$\frac{\partial P(\phi,s)}{\partial s} = \frac{\partial}{\partial \phi}\Big[\gamma^{-1} u'(\phi)\,P(\phi,s)\Big] + \gamma^{-2}\Theta\,\frac{\partial^2 P(\phi,s)}{\partial \phi^2} \quad , \tag{13.137}$$

which is called the Smoluchowski equation. Note that $-\gamma^{-1}u'(\phi)$ plays the role of a local drift velocity, and $\gamma^{-2}\Theta$ that of a diffusion constant. This may be recast as

$$\frac{\partial P}{\partial s} = -\frac{\partial W}{\partial \phi} \qquad , \qquad W(\phi,s) = -\gamma^{-1}\big(\partial_\phi u\big)P - \gamma^{-2}\,\Theta\,\partial_\phi P \quad . \tag{13.138}$$

[7]For the stochastic coupled ODEs $du_a = A_a\,dt + B_{ab}\,dW_b$ where each $W_a(t)$ is an independent Wiener process, *i.e.* $dW_a\,dW_b = \delta_{ab}\,dt$, then, using the Stratonovich stochastic calculus, one has the Fokker-Planck equation $\partial_t P = -\partial_a(A_a P) + \frac{1}{2}\partial_a\big[B_{ac}\,\partial_b(B_{bc}P)\big]$.

[8]See M. Ivanchenko and L. A. Zil'berman, *Sov. Phys. JETP* **28**, 1272 (1969) and, especially, V. Ambegaokar and B. I. Halperin, *Phys. Rev. Lett.* **22**, 1364 (1969).

[9]For the stochastic differential equation $dx = v_{\rm d}\,dt + \sqrt{2D}\,dW(t)$, where $W(t)$ is a Wiener process, the Fokker-Planck equation is $\partial_t P = -v_{\rm d}\,\partial_x P + D\,\partial_x^2 P$.

In steady state, we have that $\partial_s P = 0$, hence W must be a constant. We also demand $P(\phi, s) = P(\phi + 2\pi, s)$. To solve, define $F(\phi) \equiv e^{-\gamma\, u(\phi)/\Theta}$. In steady state, we then have

$$\frac{\partial}{\partial\phi}\left(\frac{P}{F}\right) = -\frac{\gamma^2 W}{\Theta}\cdot\frac{1}{F} \quad . \tag{13.139}$$

Integrating,

$$\begin{aligned}\frac{P(\phi)}{F(\phi)} - \frac{P(0)}{F(0)} &= -\frac{\gamma^2 W}{\Theta}\int\limits_0^\phi \frac{d\phi'}{F(\phi')} \\ \frac{P(2\pi)}{F(2\pi)} - \frac{P(\phi)}{F(\phi)} &= -\frac{\gamma^2 W}{\Theta}\int\limits_\phi^{2\pi} \frac{d\phi'}{F(\phi')} \quad .\end{aligned} \tag{13.140}$$

Multiply the first of these by $F(0)$ and the second by $F(2\pi)$, and then add, remembering that $P(2\pi) = P(0)$. One then obtains

$$P(\phi) = \frac{\gamma^2 W}{\Theta}\cdot\frac{F(\phi)}{F(2\pi) - F(0)}\cdot\left\{\int\limits_0^\phi d\phi'\,\frac{F(0)}{F(\phi')} + \int\limits_\phi^{2\pi} d\phi'\,\frac{F(2\pi)}{F(\phi')}\right\} \quad . \tag{13.141}$$

We now are in a position to demand that $P(\phi)$ be normalized. Integrating over the circle, we obtain

$$W = \frac{G(j,\gamma)}{\gamma} \tag{13.142}$$

where

$$\frac{1}{G(j,\gamma/\Theta)} = \frac{\gamma/\Theta}{\exp(\pi\gamma/\Theta) - 1}\left[\int\limits_0^{2\pi} d\phi\, f(\phi)\right]\left[\int\limits_0^{2\pi}\frac{d\phi'}{f(\phi')}\right] + \frac{\gamma}{\Theta}\int\limits_0^{2\pi} d\phi\, f(\phi)\int\limits_\phi^{2\pi}\frac{d\phi'}{f(\phi')} \quad , \tag{13.143}$$

where $f(\phi) \equiv F(\phi)/F(0) = e^{-\gamma\, u(\phi)/\Theta}\, e^{\gamma\, u(0)/\Theta}$ is normalized such that $f(0) = 1$.

It remains to relate the constant W to the voltage. For any function $g(\phi)$, we have

$$\frac{d}{dt}\left\langle g\big(\phi(s)\big)\right\rangle = \int\limits_0^{2\pi} d\phi\,\frac{\partial P}{\partial s}\,g(\phi) = -\int\limits_0^{2\pi} d\phi\,\frac{\partial W}{\partial\phi}\,g(\phi) = \int\limits_0^{2\pi} d\phi\, W(\phi)\, g'(\phi) \quad . \tag{13.144}$$

Technically we should restrict $g(\phi)$ to be periodic, but we can still make sense of this for $g(\phi) = \phi$, with

$$\left\langle\partial_s\phi\right\rangle = \int\limits_0^{2\pi} d\phi\, W(\phi) = 2\pi W \quad , \tag{13.145}$$

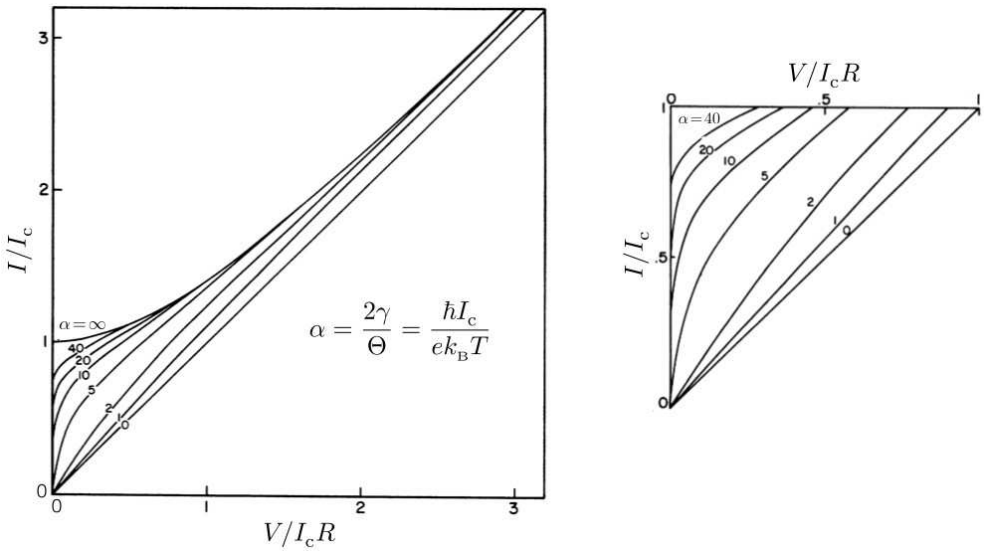


Figure 13.9: Left: scaled current bias $j = I/I_c$ *versus* scaled voltage $v = \langle V\rangle/I_cR$ for different values of the parameter γ/Θ, which is the ratio of damping to temperature. Right: detail of $j(v)$ plots. From Ambegaokar and Halperin (1969).

where the last expression on the RHS holds in steady state, where W is a constant. We could have chosen $g(\phi)$ to be a sawtooth type function, rising linearly on $\phi \in [0, 2\pi)$ then discontinuously dropping to zero, and only considered the parts where the integrands were smooth. Thus, after restoring physical units,

$$v \equiv \frac{\langle V\rangle}{I_cR} = \frac{\hbar\omega_p}{2eI_cR}\,\langle\partial_s\phi\rangle = 2\pi\,G(j,\gamma/\Theta) \quad . \quad . \tag{13.146}$$

AC Josephson effect

Suppose we add an AC bias to V, writing

$$V(t) = V_0 + V_1\sin(\omega_1 t) \quad . \tag{13.147}$$

Integrating the Josephson relation $\dot\phi = 2eV/\hbar$, we have

$$\phi(t) = \omega_J\,t + \frac{V_1}{V_0}\,\frac{\omega_J}{\omega_1}\,\cos(\omega_1 t) + \phi_0 \quad . \tag{13.148}$$

where $\omega_J = 2eV_0/\hbar$. Thus,

$$I_J(t) = I_c\,\sin\!\left(\omega_J t + \frac{V_1\,\omega_J}{V_0\,\omega_1}\,\cos(\omega_1 t) + \phi_0\right) \quad . \tag{13.149}$$

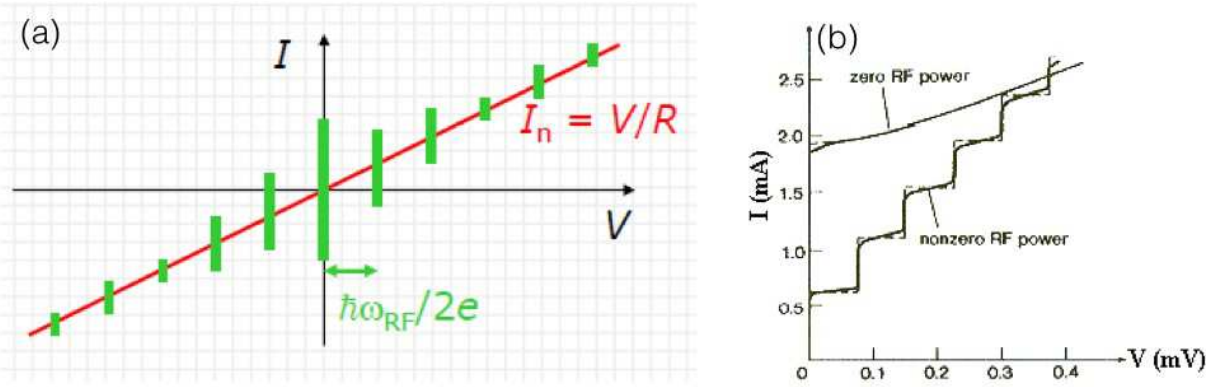


Figure 13.10: (a) Shapiro spikes in the voltage-biased AC Josephson effect. The Josephson current has a nonzero average only when $V_0 = n\hbar\omega_1/2e$, where ω_1 is the AC frequency. From *http://cmt.nbi.ku.dk/student_projects/bsc/heiselberg.pdf*. (b) Shapiro steps in the current-biased AC Josephson effect.

We now invoke the Bessel function generating relation,

$$e^{iz\cos\theta} = \sum_{n=-\infty}^{\infty} J_n(z)\, e^{-in\theta} \tag{13.150}$$

to write

$$I_{\rm J}(t) = I_{\rm c} \sum_{n=-\infty}^{\infty} J_n\!\left(\frac{V_1\,\omega_{\rm J}}{V_0\,\omega_1}\right) \sin\big[(\omega_{\rm J} - n\omega_1)\,t + \phi_0\big] \quad . \tag{13.151}$$

Thus, $I_{\rm J}(t)$ oscillates in time, except for terms for which

$$\omega_{\rm J} = n\omega_1 \quad \Rightarrow \quad V_0 = n\,\frac{\hbar\omega_1}{2e} \quad , \tag{13.152}$$

in which case

$$I_{\rm J}(t) = I_{\rm c}\, J_n\!\left(\frac{2eV_1}{\hbar\omega_1}\right) \sin\phi_0 \quad . \tag{13.153}$$

We now add back in the current through the resistor, to obtain

$$\begin{aligned} \big\langle I(t)\big\rangle &= \frac{V_0}{R} + I_{\rm c}\, J_n\!\left(\frac{2eV_1}{\hbar\omega_1}\right) \sin\phi_0 \\ &\in \left[\frac{V_0}{R} - I_{\rm c}\, J_n\!\left(\frac{2eV_1}{\hbar\omega_1}\right),\, \frac{V_0}{R} + I_{\rm c}\, J_n\!\left(\frac{2eV_1}{\hbar\omega_1}\right)\right] \quad . \end{aligned} \tag{13.154}$$

This feature, depicted in Fig. 13.10(a), is known as *Shapiro spikes*.

Current-biased AC Josephson effect

When the junction is current-biased, we must solve

$$\frac{\hbar C}{2e}\,\ddot\phi + \frac{\hbar}{2eR}\,\dot\phi + I_{\rm c}\sin\phi = I(t) \quad , \tag{13.155}$$

with $I(t) = I_0 + I_1\cos(\omega_1 t)$. This results in the *Shapiro steps* shown in Fig. 13.10(b). To analyze this equation, we write our phase space coordinates on the cylinder as $(x_1, x_2) = (\phi, \omega)$, and add the forcing term to Eqn. 13.114, *viz.*

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix}\phi\\ \omega\end{pmatrix} &= \begin{pmatrix}\omega\\ j - \sin\phi - Q^{-1}\omega\end{pmatrix} + \varepsilon\begin{pmatrix}0\\ \cos(\nu s)\end{pmatrix}\\ \frac{d\boldsymbol{x}}{ds} &= \boldsymbol{V}(\boldsymbol{x}) + \varepsilon\boldsymbol{f}(\boldsymbol{x}, s) \quad ,\end{aligned} \tag{13.156}$$

where $s = \omega_{\rm p}t$, $\nu = \omega_1/\omega_{\rm p}$, and $\varepsilon = I_1/I_{\rm c}$. As before, we have $j = I_0/I_{\rm c}$. When $\varepsilon = 0$, we have the RCSJ model, which for $|j| > 1$ has a stable limit cycle and no fixed points. The phase curves for the RCSJ model and the limit cycle for $|j| > 1$ are depicted in Fig. 13.7. In our case, the forcing term $\boldsymbol{f}(\boldsymbol{x}, s)$ has the simple form $f_1 = 0$, $f_2 = \cos(\nu s)$, but it could be more complicated and nonlinear in $\boldsymbol{x}$.

The phenomenon we are studying is called *synchronization*[10]. Linear oscillators perturbed by a harmonic force will oscillate with the forcing frequency once transients have damped out. Consider, for example, the equation $\ddot x + 2\beta\dot x + \omega_0^2 x = f_0\cos(\Omega t)$, where $\beta > 0$ is a damping coefficient. The solution is $x(t) = A(\Omega)\,\cos\big(\Omega t + \delta(\Omega)\big) + x_{\rm h}(t)$, where $x_{\rm h}(t)$ solves the homogeneous equation (*i.e.* with $f_0 = 0$) and decays to zero exponentially at large times. Nonlinear oscillators, such as the RCSJ model under study here, also can be synchronized to the external forcing, but not necessarily always. In the case of the Duffing oscillator, $\ddot x + 2\beta\dot x + x + \eta x^3$, with $\beta > 0$ and $\eta > 0$, the origin $(x = 0, \dot x = 0)$ is still a stable fixed point. In the presence of an external forcing $\varepsilon\, f_0\cos(\Omega t)$, with β, η, and ε all small, varying the detuning $\delta\Omega = \Omega - 1$ (also assumed small) can lead to hysteresis in the amplitude of the oscillations, but the oscillator is always entrained, *i.e.* synchronized with the external forcing.

The situation changes considerably if the nonlinear oscillator has no stable fixed point but rather a stable limit cycle. This is the case, for example, for the van der Pol equation $\ddot x + 2\beta(x^2 - 1)\dot x + x = 0$, and it is also the case for the RCSJ model. The limit cycle $\boldsymbol{x}_0(s)$ has a period, which we call T_0, so $\boldsymbol{x}(s + T_0) = \boldsymbol{x}(s)$. All points on the limit cycle (LC) are fixed under the T_0-advance map g_{T_0}, where $g_\tau\boldsymbol{x}(s) = \boldsymbol{x}(s + \tau)$. We may parameterize points along the LC by an angle θ which increases uniformly in s, so that $\dot\theta = \nu_0 = 2\pi/T_0$. Furthermore, since each point $\boldsymbol{x}_0(\theta)$ is a fixed point under g_{T_0}, and the LC is presumed to be attractive, we may define the *θ-isochrone* as the set of points $\{\boldsymbol{x}\}$ in phase space which flow to $\boldsymbol{x}_0(\theta)$ under repeated application of g_{T_0}. For an N-dimensional phase space, the isochrones are $(N-1)$-dimensional hypersurfaces. For the

[10]See A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization* (Cambridge, 2001).

RCSJ model, which has $N = 2$, the isochrones are curves $\theta = \theta(\phi, \omega)$ on the (ϕ, ω) cylinder. In particular, the θ-isochrone is a curve which intersects the LC at the point $\boldsymbol{x}_0(\theta)$. We then have

$$\begin{aligned}\frac{d\theta}{ds} &= \sum_{j=1}^{N} \frac{\partial\theta}{\partial x_j}\,\frac{dx_j}{ds} \\ &= \nu_0 + \varepsilon \sum_{j=1}^{N} \frac{\partial\theta}{\partial x_j}\, f_j\big(\boldsymbol{x}(s), s\big) \quad .\end{aligned} \tag{13.157}$$

If we are close to the LC, we may replace $\boldsymbol{x}(s)$ on the RHS above with $\boldsymbol{x}_0(\theta)$, yielding

$$\frac{d\theta}{ds} = \nu_0 + \varepsilon F(\theta, s) \quad , \tag{13.158}$$

where

$$F(\theta, s) = \sum_{j=1}^{N} \left.\frac{\partial\theta}{\partial x_j}\right|_{\boldsymbol{x}_0(\theta)} f_j\big(\boldsymbol{x}_0(\theta), s\big) \quad . \tag{13.159}$$

OK, so now here's the thing. The function $F(\theta, s)$ is separately periodic in both its arguments, so we may write

$$F(\theta, s) = \sum_{k,l} F_{k,l}\, e^{i(k\theta + l\nu s)} \quad , \tag{13.160}$$

where $\boldsymbol{f}\big(\boldsymbol{x}, s + \frac{2\pi}{\nu}\big) = \boldsymbol{f}(\boldsymbol{x}, s)$, *i.e.* ν is the forcing frequency. The unperturbed solution has $\dot\theta = \nu_0$, hence the forcing term in Eqn. 13.158 is resonant when $k\nu_0 + l\nu \approx 0$. This occurs when $\nu \approx \frac{p}{q}\,\nu_0$, where p and q are relatively prime integers. The resonance condition is satisfied when $k = rp$ and $l = -rq$ for any integer r

.

We now separate the resonant from nonresonant terms in the (k, l) sum, writing

$$\dot\theta = \nu_0 + \varepsilon \sum_{r=-\infty}^{\infty} F_{rp,-rq}\, e^{ir(p\theta - q\nu s)} + \text{NRT} \quad , \tag{13.161}$$

where NRT stands for "non-resonant terms". We next average over short time scales to eliminate these nonresonant terms, and focus on the dynamics of the average phase $\langle\theta\rangle$. Defining $\psi \equiv p\,\langle\theta\rangle - q\,\nu s$, we have

$$\begin{aligned}\dot\psi &= p\,\langle\dot\theta\rangle - q\nu \\ &= (p\nu_0 - q\nu) + \varepsilon p \sum_{r=-\infty}^{\infty} F_{rp,-rq}\, e^{ir\psi} \\ &= -\delta + \varepsilon\, G(\psi) \quad ,\end{aligned} \tag{13.162}$$

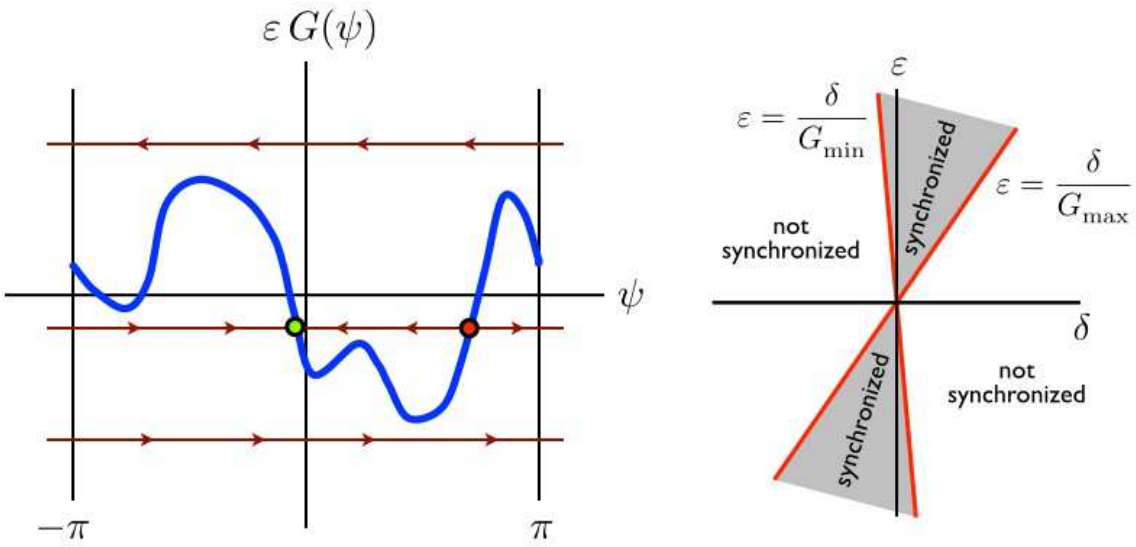


Figure 13.11: Left: graphical solution of $\dot\psi = -\delta + \varepsilon\, G(\psi)$. Fixed points are only possible if $-\varepsilon\, G_{\rm min} \leqslant \delta \leqslant G_{\rm max}$. Right: synchronization region, shown in grey, in the (δ, ε) plane.

where $\delta \equiv q\nu - p\nu_0$ is the detuning, and $G(\psi) \equiv p \sum_r F_{rp,-rq}\, e^{ir\psi}$ is the sum over resonant terms. This last equation is that of a simple $N = 1$ dynamical system on the circle! If the detuning δ falls within the range $\big[\varepsilon G_{\rm min}\,,\, \varepsilon G_{\rm max}\big]$, then ψ flows to a stable fixed point where $\delta = \varepsilon\, G(\psi^*)$. The oscillator is then synchronized with the forcing, because $\langle\dot\theta\rangle \to \frac{q}{p}\,\nu$. If the detuning is too large and lies outside this range, then there is no synchronization. Rather, $\psi(s)$ increases linearly with the time s, and $\langle\theta(t)\rangle = \theta_0 + \frac{q}{p}\,\nu s + \frac{1}{p}\,\psi(s)$, where

$$dt = \frac{d\psi}{\varepsilon\, G(\psi) - \delta} \qquad \Longrightarrow \qquad T_\psi = \int_0^{2\pi} \frac{d\psi}{\varepsilon\, G(\psi) - \delta} \quad . \tag{13.163}$$

For weakly forced, weakly nonlinear oscillators, resonance occurs only for $\nu = \pm\nu_0$, but in the case of weakly forced, strongly nonlinear oscillators, the general resonance condition is $\nu = \frac{p}{q}\,\nu_0$. The reason is that in the case of weakly nonlinear oscillators, the limit cycle is itself harmonic to zeroth order. There are then only two frequencies in its Fourier decomposition, *i.e.* $\pm\nu_0$. In the strongly nonlinear case, the limit cycle is decomposed into a fundamental frequency ν_0 plus all its harmonics. In addition, the forcing $\boldsymbol{f}(\boldsymbol{x}, s)$ can itself can be a general periodic function of s, involving multiples of the fundamental forcing frequency ν. For the case of the RCSJ, the forcing function is harmonic and independent of $\boldsymbol{x}$. This means that only the $l = \pm 1$ terms enter in the above analysis.

13.4 Ultrasonic Attenuation

Recall the electron-phonon Hamiltonian,

$$\hat{H}_{\rm el-ph} = \frac{1}{\sqrt{V}} \sum_{\substack{\boldsymbol{k},\boldsymbol{k}' \\ \sigma,\lambda}} g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \left(a^\dagger_{\boldsymbol{k}'-\boldsymbol{k},\lambda} + a_{\boldsymbol{k}-\boldsymbol{k}',\lambda} \right) c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma} \tag{13.164}$$

$$= \frac{1}{\sqrt{V}} \sum_{\substack{\boldsymbol{k},\boldsymbol{k} \\ \sigma,\lambda}} g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \big(a^\dagger_{\boldsymbol{k}'-\boldsymbol{k},\lambda} + a_{\boldsymbol{k}-\boldsymbol{k}',\lambda} \big) \big(u_{\boldsymbol{k}} \gamma^\dagger_{\boldsymbol{k}\sigma} - \sigma\, e^{-i\phi}\, v_{\boldsymbol{k}}\, \gamma_{-\boldsymbol{k}\,-\sigma} \big) \big(u_{\boldsymbol{k}'} \gamma_{\boldsymbol{k}'\sigma} - \sigma\, e^{i\phi}\, v_{\boldsymbol{k}'}\, \gamma^\dagger_{-\boldsymbol{k}'\,-\sigma} \big) \ .$$

Let's now compute the phonon lifetime using Fermi's Golden Rule[11]. In the phonon absorption process, a phonon of wavevector $\boldsymbol{q}$ is absorbed by an electron of wavevector $\boldsymbol{k}$, converting it into an electron of wavevector $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{q}$. The net absorption rate of $(\boldsymbol{q}, \lambda)$ phonons is then is given by the rate of

$$\Gamma^{\rm abs}_{\boldsymbol{q}\lambda} = \frac{2\pi n_{\boldsymbol{q},\lambda}}{V} \sum_{\boldsymbol{k},\boldsymbol{k}',\sigma} \left| g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \right|^2 \left(u_{\boldsymbol{k}} u_{\boldsymbol{k}'} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'} \right)^2 f_{\boldsymbol{k}\sigma} \left(1 - f_{\boldsymbol{k}'\sigma} \right) \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}} - \hbar\omega_{\boldsymbol{q}\lambda})\, \delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}\,{\rm mod}\,\boldsymbol{G}} \quad . \tag{13.165}$$

Here $n_{\boldsymbol{q}\lambda}$ is the Bose function and $f_{\boldsymbol{k}\sigma}$ the Fermi function, and we have assumed that the phonon frequencies are all smaller than 2Δ, so we may ignore quasiparticle pair creation and pair annihilation processes. Note that the electron Fermi factors yield the probability that the state $|\boldsymbol{k}\sigma\rangle$ is occupied while $|\boldsymbol{k}'\sigma\rangle$ is vacant. *Mutatis mutandis*, the emission rate of these phonons is[12]

$$\Gamma^{\rm em}_{\boldsymbol{q}\lambda} = \frac{2\pi (n_{\boldsymbol{q},\lambda} + 1)}{V} \sum_{\boldsymbol{k},\boldsymbol{k}',\sigma} \left| g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \right|^2 \left(u_{\boldsymbol{k}} u_{\boldsymbol{k}'} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'} \right)^2 f_{\boldsymbol{k}'\sigma} \left(1 - f_{\boldsymbol{k}\sigma} \right) \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}} - \hbar\omega_{\boldsymbol{q}\lambda})\, \delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}\,{\rm mod}\,\boldsymbol{G}} \quad . \tag{13.166}$$

We then have

$$\frac{dn_{\boldsymbol{q}\lambda}}{dt} = -\alpha_{\boldsymbol{q}\lambda}\, n_{\boldsymbol{q}\lambda} + s_{\boldsymbol{q}\lambda} \quad , \tag{13.167}$$

where

$$\alpha_{\boldsymbol{q}\lambda} = \frac{4\pi}{V} \sum_{\boldsymbol{k},\boldsymbol{k}'} \left| g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \right|^2 \left(u_{\boldsymbol{k}} u_{\boldsymbol{k}'} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'} \right)^2 \left(f_{\boldsymbol{k}} - f_{\boldsymbol{k}'} \right) \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}} - \hbar\omega_{\boldsymbol{q}\lambda})\, \delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}\,{\rm mod}\,\boldsymbol{G}} \tag{13.168}$$

is the attenuation rate, and $s_{\boldsymbol{q}\lambda}$ is due to spontaneous emission,

$$s_{\boldsymbol{q}\lambda} = \frac{4\pi}{V} \sum_{\boldsymbol{k},\boldsymbol{k}'} \left| g_{\boldsymbol{k}\boldsymbol{k}'\lambda} \right|^2 \left(u_{\boldsymbol{k}} u_{\boldsymbol{k}'} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'} \right)^2 f_{\boldsymbol{k}'} \left(1 - f_{\boldsymbol{k}} \right) \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}} - \hbar\omega_{\boldsymbol{q}\lambda})\, \delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}\,{\rm mod}\,\boldsymbol{G}} \quad . \tag{13.169}$$

[11] Here we follow §3.4 of J. R. Schrieffer, *Theory of Superconductivity* (Benjamin-Cummings, 1964).

[12] Note the factor of $n + 1$ in the emission rate, where the additional 1 is due to spontaneous emission. The absorption rate includes only a factor of n.

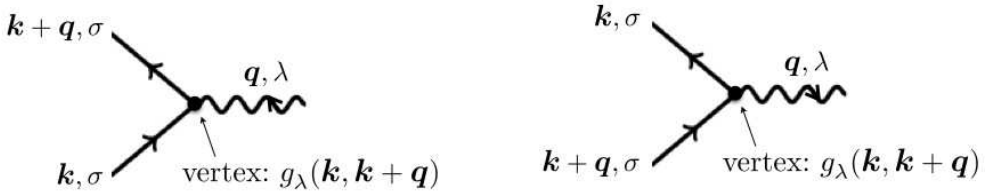


Figure 13.12: Phonon absorption and emission processes.

We now expand about the Fermi surface, writing

$$\frac{1}{V}\sum_{\boldsymbol{k},\boldsymbol{k}'} F(\xi_{\boldsymbol{k}},\xi_{\boldsymbol{k}'})\,\delta_{\boldsymbol{k}',\boldsymbol{k}+\boldsymbol{q}} = \tfrac{1}{4}\,g^2(\mu)\int\limits_{-\infty}^{\infty}\!d\xi\int\limits_{-\infty}^{\infty}\!d\xi'\,F(\xi,\xi')\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,\delta(k_{\rm F}\hat{\boldsymbol{k}}'-k_{\rm F}\hat{\boldsymbol{k}}-\boldsymbol{q})\;. \tag{13.170}$$

for any function $F(\xi,\xi')$. The integrals over $\hat{\boldsymbol{k}}$ and $\hat{\boldsymbol{k}}'$ give

$$\int\frac{d\hat{\boldsymbol{k}}}{4\pi}\int\frac{d\hat{\boldsymbol{k}}'}{4\pi}\,\delta(k_{\rm F}\hat{\boldsymbol{k}}'-k_{\rm F}\hat{\boldsymbol{k}}-\boldsymbol{q}) = \frac{1}{4\pi k_{\rm F}^3}\cdot\frac{k_{\rm F}}{2q}\cdot\Theta(2k_{\rm F}-q)\quad . \tag{13.171}$$

The step function appears naturally because the constraint $k_{\rm F}\hat{\boldsymbol{k}}' = k_{\rm F}\hat{\boldsymbol{k}}+\boldsymbol{q}$ requires that $\boldsymbol{q}$ connect two points which lie on the metallic Fermi surface, so the largest $|\boldsymbol{q}|$ can be is $2k_{\rm F}$. We will drop the step function in the following expressions, assuming $q<2k_{\rm F}$, but it is good to remember that it is implicitly present. Thus, ignoring *Umklapp* processes, we have

$$\alpha_{\boldsymbol{q}\lambda} = \frac{g^2(\mu)\,|g_{\boldsymbol{q}\lambda}|^2}{8\,k_{\rm F}^2\,q}\int\limits_{-\infty}^{\infty}\!d\xi\int\limits_{-\infty}^{\infty}\!d\xi'\,(uu'-vv')^2\,(f-f')\,\delta(E'-E-\hbar\omega_{\boldsymbol{q}\lambda})\quad . \tag{13.172}$$

We now use

$$(uu'\pm vv')^2 = \left(\sqrt{\frac{E+\xi}{2E}}\sqrt{\frac{E'+\xi'}{2E'}}\pm\sqrt{\frac{E-\xi}{2E}}\sqrt{\frac{E'-\xi'}{2E'}}\,\right)^2 = \frac{EE'+\xi\xi'\pm\Delta^2}{EE'} \tag{13.173}$$

and change variables $\big(\,\xi = E\,dE/\sqrt{E^2-\Delta^2}\,\big)$ to write

$$\alpha_{\boldsymbol{q}\lambda} = \frac{g^2(\mu)\,|g_{\boldsymbol{q}\lambda}|^2}{2\,k_{\rm F}^2\,q}\int\limits_{\Delta}^{\infty}\!dE\int\limits_{\Delta}^{\infty}\!dE'\,\frac{(EE'-\Delta^2)(f-f')}{\sqrt{E^2-\Delta^2}\,\sqrt{E'^2-\Delta^2}}\,\delta(E'-E-\hbar\omega_{\boldsymbol{q}\lambda})\quad . \tag{13.174}$$

We now satisfy the Dirac delta function, which means we eliminate the E' integral and set $E' = E+\hbar\omega_{\boldsymbol{q}\lambda}$ everywhere else in the integrand. Clearly the $f-f'$ term will be first order in the smallness of $\hbar\omega_{\boldsymbol{q}\prime}$ so in all other places we may set $E'=E$ to lowest order. This simplifies the above expression considerably, and we are left with

$$\alpha_{\boldsymbol{q}\lambda} = \frac{g^2(\mu)\,|g_{\boldsymbol{q}\lambda}|^2\,\hbar\omega_{\boldsymbol{q}\lambda}}{2\,k_{\rm F}^2\,q}\int\limits_{\Delta}^{\infty}\!dE\left(-\frac{\partial f}{\partial E}\right) = \frac{g^2(\mu)\,|g_{\boldsymbol{q}\lambda}|^2\,\hbar\omega_{\boldsymbol{q}\lambda}}{2\,k_{\rm F}^2\,q}\,f(\Delta)\quad , \tag{13.175}$$

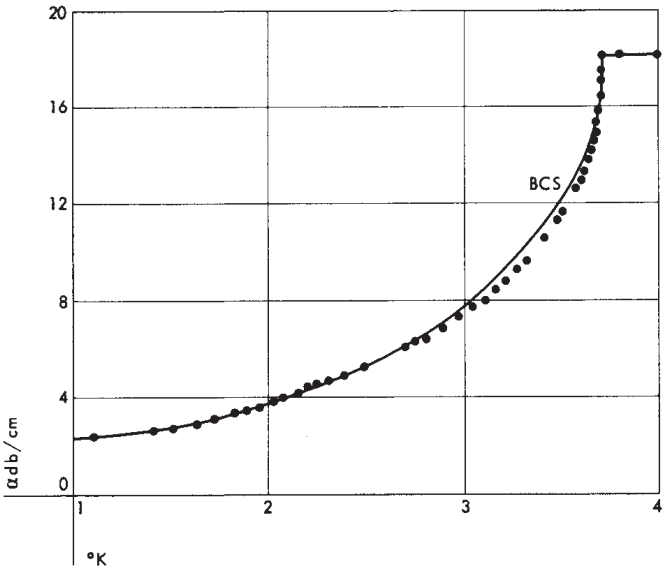


Figure 13.13: Ultrasonic attenuation in tin, compared with predictions of the BCS theory. From R. W. Morse, *IBM Jour. Res. Dev.* **6**, 58 (1963).

where $q < 2k_{\rm F}$ is assumed. For $q \to 0$, we have $\omega_{\boldsymbol{q}\lambda}/q \to c_\lambda(\hat{\boldsymbol{q}})$, the phonon velocity.

We may now write the ratio of the phonon attenuation rate in the superconducting and normal states as

$$\frac{\alpha_{\rm S}(T)}{\alpha_{\rm N}(T)} = \frac{f(\Delta)}{f(0)} = \frac{2}{\exp\left(\frac{\Delta(T)}{k_{\rm B}T}\right)+1} \quad . \tag{13.176}$$

The ratio naturally goes to unity at $T = T_{\rm c}$, where Δ vanishes. Results from early experiments on superconducting Sn are shown in Fig. 13.13.

13.5 Nuclear Magnetic Relaxation

We start with the hyperfine Hamiltonian,

$$\hat{H}_{\rm HF} = A \sum_{\boldsymbol{k},\boldsymbol{k}'} \sum_{\boldsymbol{R}} \varphi^*_{\boldsymbol{k}}(\boldsymbol{R})\, \varphi_{\boldsymbol{k}'}(\boldsymbol{R}) \left[J^+_{\boldsymbol{R}}\, c^\dagger_{\boldsymbol{k}\downarrow}\, c_{\boldsymbol{k}'\uparrow} + J^-_{\boldsymbol{R}}\, c^\dagger_{\boldsymbol{k}\uparrow}\, c_{\boldsymbol{k}'\downarrow} + J^z_{\boldsymbol{R}} \left(c^\dagger_{\boldsymbol{k}\uparrow}\, c_{\boldsymbol{k}'\uparrow} - c^\dagger_{\boldsymbol{k}\downarrow}\, c_{\boldsymbol{k}'\downarrow} \right) \right] \tag{13.177}$$

where $\boldsymbol{J}_{\boldsymbol{R}}$ is the nuclear spin operator on nuclear site $\boldsymbol{R}$, satisfying

$$\left[J^\mu_{\boldsymbol{R}}\,,\, J^\nu_{\boldsymbol{R}'} \right] = i\, \epsilon_{\mu\nu\lambda}\, J^\lambda_{\boldsymbol{R}}\, \delta_{\boldsymbol{R},\boldsymbol{R}'} \quad , \tag{13.178}$$

and where $\varphi_{\boldsymbol{k}}(\boldsymbol{R})$ is the amplitude of the electronic Bloch wavefunction (with band index suppressed) on the nuclear site $\boldsymbol{R}$. Using

$$c_{\boldsymbol{k}\sigma} = u_{\boldsymbol{k}}\, \gamma_{\boldsymbol{k}\sigma} - \sigma\, v_{\boldsymbol{k}}\, e^{i\phi}\, \gamma^\dagger_{-\boldsymbol{k}\,-\sigma} \tag{13.179}$$

we have for $\boldsymbol{S}_{\boldsymbol{k}\boldsymbol{k}'} = \frac{1}{2}\, c^\dagger_{\boldsymbol{k}\mu}\, \boldsymbol{\sigma}_{\mu\nu}\, c_{\boldsymbol{k}'\nu}$,

$$\begin{aligned}
S^+_{\boldsymbol{k}\boldsymbol{k}'} &= u_{\boldsymbol{k}} u_{\boldsymbol{k}'} \gamma^\dagger_{\boldsymbol{k}\uparrow}\gamma_{\boldsymbol{k}'\downarrow} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, \gamma_{-\boldsymbol{k}\downarrow}\gamma^\dagger_{-\boldsymbol{k}'\uparrow} + u_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, e^{i\phi}\, \gamma^\dagger_{\boldsymbol{k}\uparrow}\gamma^\dagger_{-\boldsymbol{k}'\uparrow} - u_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, e^{-i\phi}\, \gamma_{-\boldsymbol{k}\downarrow}\gamma_{\boldsymbol{k}'\downarrow} \\
S^-_{\boldsymbol{k}\boldsymbol{k}'} &= u_{\boldsymbol{k}} u_{\boldsymbol{k}'} \gamma^\dagger_{\boldsymbol{k}\downarrow}\gamma_{\boldsymbol{k}'\uparrow} - v_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, \gamma_{-\boldsymbol{k}\uparrow}\gamma^\dagger_{-\boldsymbol{k}'\downarrow} - u_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, e^{i\phi}\, \gamma^\dagger_{\boldsymbol{k}\downarrow}\gamma^\dagger_{-\boldsymbol{k}'\downarrow} + u_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, e^{-i\phi}\, \gamma_{-\boldsymbol{k}\uparrow}\gamma_{\boldsymbol{k}'\uparrow} \qquad (13.180) \\
S^z_{\boldsymbol{k}\boldsymbol{k}'} &= \tfrac{1}{2}\sum_\sigma \Big(u_{\boldsymbol{k}} u_{\boldsymbol{k}'}\, \gamma^\dagger_{\boldsymbol{k}\sigma}\gamma_{\boldsymbol{k}'\sigma} + v_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, \gamma_{-\boldsymbol{k}\,-\sigma}\gamma^\dagger_{-\boldsymbol{k}'\,-\sigma} - \sigma\, u_{\boldsymbol{k}} v_{\boldsymbol{k}'}\, e^{i\phi}\, \gamma^\dagger_{\boldsymbol{k}\sigma}\gamma^\dagger_{-\boldsymbol{k}'\,-\sigma} - \sigma\, v_{\boldsymbol{k}} u_{\boldsymbol{k}'}\, e^{-i\phi}\, \gamma_{-\boldsymbol{k}\,-\sigma}\gamma_{\boldsymbol{k}'\sigma} \Big) .
\end{aligned}$$

Let's assume our nuclei are initially spin polarized, and let us calculate the rate $1/T_1$ at which the J^z component of the nuclear spin relaxes. Again appealing to the Golden Rule,

$$\frac{1}{T_1} = 2\pi\,|A|^2 \sum_{\boldsymbol{k},\boldsymbol{k}'} |\varphi_{\boldsymbol{k}}(0)|^2\, |\varphi_{\boldsymbol{k}'}(0)|^2 \left(u_{\boldsymbol{k}} u_{\boldsymbol{k}'} + v_{\boldsymbol{k}} v_{\boldsymbol{k}'}\right)^2 f_{\boldsymbol{k}}\big(1 - f_{\boldsymbol{k}'}\big)\, \delta(E_{\boldsymbol{k}'} - E_{\boldsymbol{k}} - \hbar\omega) \qquad (13.181)$$

where ω is the nuclear spin precession frequency in the presence of internal or external magnetic fields. Assuming $\varphi_{\boldsymbol{k}}(\boldsymbol{R}) = C/\sqrt{V}$, we write $V^{-1}\sum_{\boldsymbol{k}} \to \frac{1}{2} g(\mu)\int d\xi$ and we appeal to Eqn. 13.173. Note that the coherence factors in this case give $(uu' + vv')^2$, as opposed to $(uu' - vv')^2$ as we found in the case of ultrasonic attenuation (more on this below). What we then obtain is

$$\frac{1}{T_1} = 2\pi\,|A|^2\,|C|^4\, g^2(\mu) \int\limits_\Delta^\infty dE\, \frac{E(E+\hbar\omega) + \Delta^2}{\sqrt{E^2-\Delta^2}\,\sqrt{(E+\hbar\omega)^2 - \Delta^2}}\, f(E)\,\big[1 - f(E+\hbar\omega)\big] \quad . \qquad (13.182)$$

Let's first evaluate this expression for normal metals, where $\Delta = 0$. We have

$$\frac{1}{T_{1,\mathrm{N}}} = 2\pi\,|A|^2\,|C|^4\, g^2(\mu) \int\limits_0^\infty d\xi\; f(\xi)\big[1 - f(\xi+\hbar\omega)\big] = \pi\,|A|^2\,|C|^4\, g^2(\mu)\, k_{\mathrm{B}} T \quad , \qquad (13.183)$$

where we have assumed $\hbar\omega \ll k_{\mathrm{B}} T$, and used $f(\xi)\big[1 - f(\xi)\big] = -k_{\mathrm{B}} T\, f'(\xi)$. The assumption $\omega \to 0$ is appropriate because the nuclear magneton is so tiny: $\mu_{\mathrm{N}}/k_{\mathrm{B}} = 3.66 \times 10^{-4}\mathrm{K/T}$, so the nuclear splitting is on the order of mK even at fields as high as 10 T. The NMR relaxation rate is thus proportional to temperature, a result known as the *Korringa law*.

Now let's evaluate the ratio of NMR relaxation rates in the superconducting and normal states. Assuming $\hbar\omega \ll \Delta$, we have

$$\frac{T^{-1}_{1,\mathrm{S}}}{T^{-1}_{1,\mathrm{N}}} = 2\int\limits_\Delta^\infty dE\, \frac{E(E+\hbar\omega) + \Delta^2}{\sqrt{E^2-\Delta^2}\,\sqrt{(E+\hbar\omega)^2 - \Delta^2}} \left(-\frac{\partial f}{\partial E}\right) \quad . \qquad (13.184)$$

We dare not send $\omega \to 0$ in the integrand, because this would lead to a logarithmic divergence. Numerical integration shows that for $\hbar\omega \lesssim \frac{1}{2} k_{\mathrm{B}} T_{\mathrm{c}}$, the above expression has a peak just below $T = T_{\mathrm{c}}$. This is the famous *Hebel-Slichter peak*.

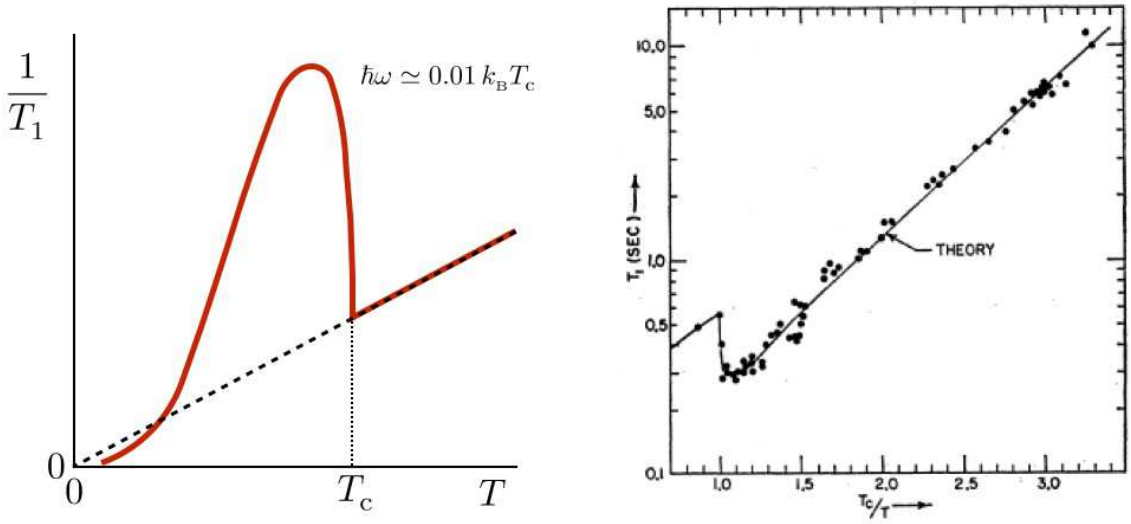


Figure 13.14: Left: Sketch of NMR relaxation rate $1/T_1$ *versus* temperature as predicted by BCS theory, with $\hbar\omega \approx 0.01\, k_{\rm B} T_{\rm c}$, showing the Hebel-Slichter peak. Right: T_1 *versus* $T_{\rm c}/T$ in a powdered aluminum sample, from Y. Masuda and A. G. Redfield, *Phys. Rev.* **125**, 159 (1962). The Hebel-Slichter peak is seen here as a dip.

These results for acoustic attenuation and spin relaxation exemplify so-called *case I* and *case II* responses of the superconductor, respectively. In case I, the transition matrix element is proportional to $uu' - vv'$, which vanishes at $\xi = 0$. In case II, the transition matrix element is proportional to $uu' + vv'$.

13.6 General Theory of BCS Linear Response

Consider a general probe of the superconducting state described by the perturbation Hamiltonian

$$\hat{V}(t) = \sum_{\boldsymbol{k},\sigma} \sum_{\boldsymbol{k}',\sigma'} \Big[B\big(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma'\big)\, e^{-i\omega t} + B^*\big(\boldsymbol{k}'\sigma' \,|\, \boldsymbol{k}\sigma\big)\, e^{+i\omega t} \Big] c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma'} \quad . \tag{13.185}$$

An example would be ultrasonic attenuation, where

$$\hat{V}_{\rm ultra}(t) = U \sum_{\boldsymbol{k},\boldsymbol{k}',\sigma} \phi_{\boldsymbol{k}'-\boldsymbol{k}}(t)\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma'} \quad . \tag{13.186}$$

Here $\phi(\boldsymbol{r}) = \nabla \cdot \boldsymbol{u}$ is the deformation of the lattice and U is the deformation potential, with the interaction of the local deformation with the electrons given by $U\phi(\boldsymbol{r})\, n(\boldsymbol{r})$, where $n(\boldsymbol{r})$ is the total electron number density at $\boldsymbol{r}$. Another example is interaction with microwaves. In this

case, the bare dispersion is corrected by $\boldsymbol{p} \to \boldsymbol{p} + \frac{e}{c}\boldsymbol{A}$, hence

$$\hat{V}_{\mu\text{wave}}(t) = \frac{e\hbar}{2m^*c} \sum_{\boldsymbol{k},\boldsymbol{k}',\sigma} (\boldsymbol{k} + \boldsymbol{k}') \cdot \boldsymbol{A}_{\boldsymbol{k}'-\boldsymbol{k}}(t)\, c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma'} \quad , \tag{13.187}$$

where m^* is the band mass.

Consider now a general perturbation Hamiltonian of the form

$$\hat{V} = -\sum_i \big(\phi_i(t)\, C_i^\dagger + \phi_i^*(t)\, C_i\big) \tag{13.188}$$

where C_i are operators labeled by i. We write

$$\phi_i(t) = \int\limits_{-\infty}^{\infty} \frac{d\omega}{2\pi}\, \hat{\phi}_i(\omega)\, e^{-i\omega t} \quad . \tag{13.189}$$

According to the general theory of linear response formulated in chapter 9, the power dissipation due to this perturbation is given by

$$\begin{aligned} P(\omega) = &-i\omega\, \hat{\phi}_i^*(\omega)\, \hat{\phi}_j(\omega)\, \hat{\chi}_{C_iC_j^\dagger}(\omega) + i\omega\, \hat{\phi}_i(\omega)\, \hat{\phi}_j^*(\omega)\, \hat{\chi}_{C_i^\dagger C_j}(-\omega) \\ &- i\omega\, \hat{\phi}_i^*(\omega)\, \hat{\phi}_j^*(-\omega)\, \hat{\chi}_{C_iC_j}(\omega) + i\omega\, \hat{\phi}_i(\omega)\, \hat{\phi}_j(-\omega)\, \hat{\chi}_{C_i^\dagger C_j^\dagger}(-\omega) \quad . \end{aligned} \tag{13.190}$$

where $\hat{H} = \hat{H}_0 + \hat{V}$ and $C_i(t) = e^{i\hat{H}_0 t/\hbar}\, C_i\, e^{-i\hat{H}_0 t/\hbar}$ is the operator C_i in the interaction representation.

$$\hat{\chi}_{AB}(\omega) = \frac{i}{\hbar} \int\limits_0^\infty dt\, e^{-i\omega t} \left\langle \big[A(t)\,, B(0)\big] \right\rangle \tag{13.191}$$

For our application, we have $i \equiv (\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma')$ and $j \equiv (\boldsymbol{p}\mu \,|\, \boldsymbol{p}'\mu')$, with $C_i^\dagger = c^\dagger_{\boldsymbol{k}\sigma}\, c_{\boldsymbol{k}'\sigma'}$ and $C_j = c^\dagger_{\boldsymbol{p}'\mu'} c_{\boldsymbol{p}\mu}$, *etc.* So we need to compute the response function,

$$\hat{\chi}_{C_iC_j^\dagger}(\omega) = \frac{i}{\hbar} \int\limits_0^\infty dt \left\langle \Big[c^\dagger_{\boldsymbol{k}'\sigma'}(t)\, c_{\boldsymbol{k}\sigma}(t)\,,\, c^\dagger_{\boldsymbol{p}\mu}(0)\, c_{\boldsymbol{p}'\mu'}(0) \Big] \right\rangle e^{i\omega t} \quad . \tag{13.192}$$

OK, so strap in, because this is going to be a bit of a bumpy ride.

We evaluate the commutator in real time and then Fourier transform to the frequency domain. Using Wick's theorem for fermions[13],

$$\langle c_1^\dagger\, c_2\, c_3^\dagger\, c_4 \rangle = \langle c_1^\dagger\, c_2 \rangle \langle c_3^\dagger\, c_4 \rangle - \langle c_1^\dagger\, c_3^\dagger \rangle \langle c_2\, c_4 \rangle + \langle c_1^\dagger\, c_4 \rangle \langle c_2\, c_3^\dagger \rangle \quad , \tag{13.193}$$

[13] Wick's theorem is valid when taking expectation values in Slater determinant states.

we have

$$\begin{aligned}\chi_{C_i C_j^\dagger}(t) &= \frac{i}{\hbar}\Big\langle \Big[c^\dagger_{\boldsymbol{k}'\sigma'}(t)\, c_{\boldsymbol{k}\sigma}(t)\,,\, c^\dagger_{\boldsymbol{p}\mu}(0)\, c_{\boldsymbol{p}'\mu'}(0)\Big]\Big\rangle\, \Theta(t) \\ &= -\frac{i}{\hbar}\Big[F^a_{\boldsymbol{k}'\sigma'}(t)\, F^b_{\boldsymbol{k}\sigma}(t) - F^c_{\boldsymbol{k}\sigma}(t)\, F^d_{\boldsymbol{k}'\sigma'}(t)\Big]\,\delta_{\boldsymbol{p},\boldsymbol{k}}\,\delta_{\boldsymbol{p}',\boldsymbol{k}'}\,\delta_{\mu,\sigma}\,\delta_{\mu',\sigma'} \\ &\qquad + \frac{i}{\hbar}\Big[G^a_{\boldsymbol{k}'\sigma'}(t)\, G^b_{\boldsymbol{k}\sigma}(t) - G^c_{\boldsymbol{k}\sigma}(t)\, G^d_{\boldsymbol{k}'\sigma'}(t)\Big]\,\sigma\sigma'\,\delta_{\boldsymbol{p},-\boldsymbol{k}'}\,\delta_{\boldsymbol{p}',-\boldsymbol{k}}\,\delta_{\mu,-\sigma'}\,\delta_{\mu',-\sigma}\quad ,\end{aligned} \tag{13.194}$$

where, using the Bogoliubov transformation,

$$\begin{aligned} c_{\boldsymbol{k}\sigma} &= u_{\boldsymbol{k}}\,\gamma_{\boldsymbol{k}\sigma} - \sigma\, v_{\boldsymbol{k}}\, e^{+i\phi}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma} \\ c^\dagger_{-\boldsymbol{k}\,-\sigma} &= u_{\boldsymbol{k}}\,\gamma^\dagger_{-\boldsymbol{k}\,-\sigma} + \sigma\, v_{\boldsymbol{k}}\, e^{-i\phi}\,\gamma_{\boldsymbol{k}\sigma}\quad ,\end{aligned} \tag{13.195}$$

we find

$$\begin{aligned} F^a_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c^\dagger_{\boldsymbol{q}\nu}(t)\, c_{\boldsymbol{q}\nu}(0)\big\rangle = -i\,\Theta(t)\,\Big\{u^2_{\boldsymbol{q}}\, e^{iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}}) + v^2_{\boldsymbol{q}}\, e^{-iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big]\Big\} \\ F^b_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c_{\boldsymbol{q}\nu}(t)\, c^\dagger_{\boldsymbol{q}\nu}(0)\big\rangle = -i\,\Theta(t)\,\Big\{u^2_{\boldsymbol{q}}\, e^{-iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big] + v^2_{\boldsymbol{q}}\, e^{iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}})\Big\} \\ F^c_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c^\dagger_{\boldsymbol{q}\nu}(0)\, c_{\boldsymbol{q}\nu}(t)\big\rangle = -i\,\Theta(t)\,\Big\{u^2_{\boldsymbol{q}}\, e^{-iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}}) + v^2_{\boldsymbol{q}}\, e^{iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big]\Big\} \\ F^d_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c_{\boldsymbol{q}\nu}(0)\, c^\dagger_{\boldsymbol{q}\nu}(t)\big\rangle = -i\,\Theta(t)\,\Big\{u^2_{\boldsymbol{q}}\, e^{iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big] + v^2_{\boldsymbol{q}}\, e^{-iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}})\Big\}\end{aligned} \tag{13.196}$$

and

$$\begin{aligned} G^a_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c^\dagger_{\boldsymbol{q}\nu}(t)\, c^\dagger_{-\boldsymbol{q}\,-\nu}(0)\big\rangle = -i\,\Theta(t)\, u_{\boldsymbol{q}}\, v_{\boldsymbol{q}}\, e^{-i\phi}\Big\{e^{iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}}) - e^{-iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big]\Big\} \\ G^b_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c_{\boldsymbol{q}\nu}(t)\, c_{-\boldsymbol{q}\,-\nu}(0)\big\rangle = -i\,\Theta(t)\, u_{\boldsymbol{q}}\, v_{\boldsymbol{q}}\, e^{+i\phi}\Big\{e^{-E_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big] - e^{-iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}})\Big\} \\ G^c_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c^\dagger_{\boldsymbol{q}\nu}(0)\, c^\dagger_{-\boldsymbol{q}\,-\nu}(t)\big\rangle = -i\,\Theta(t)\, u_{\boldsymbol{q}}\, v_{\boldsymbol{q}}\, e^{-i\phi}\Big\{e^{iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big] - e^{-iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}})\Big\} \\ G^d_{\boldsymbol{q}\nu}(t) &= -i\,\Theta(t)\,\big\langle c^\dagger_{\boldsymbol{q}\nu}(0)\, c^\dagger_{-\boldsymbol{q}\,-\nu}(t)\big\rangle = -i\,\Theta(t)\, u_{\boldsymbol{q}}\, v_{\boldsymbol{q}}\, e^{+i\phi}\Big\{e^{-iE_{\boldsymbol{q}}t/\hbar}\, f(E_{\boldsymbol{q}}) - e^{iE_{\boldsymbol{q}}t/\hbar}\big[1-f(E_{\boldsymbol{q}})\big]\Big\}\quad .\end{aligned} \tag{13.197}$$

Taking the Fourier transforms, we have[14]

$$\hat{F}^a(\omega) = \frac{u^2 f}{\omega + E + i\epsilon} + \frac{v^2\,(1-f)}{\omega - E + i\epsilon}\qquad ,\qquad \hat{F}^c(\omega) = \frac{u^2 f}{\omega - E + i\epsilon} + \frac{v^2\,(1-f)}{\omega + E + i\epsilon} \tag{13.198}$$

$$\hat{F}^b(\omega) = \frac{u^2\,(1-f)}{\omega - E + i\epsilon} + \frac{v^2 f}{\omega + E + i\epsilon}\qquad ,\qquad \hat{F}^d(\omega) = \frac{u^2\,(1-f)}{\omega + E + i\epsilon} + \frac{v^2 f}{\omega - E + i\epsilon} \tag{13.199}$$

[14]Here we are being somewhat loose and have set $\hbar = 1$ to avoid needless notational complication. We shall restore the proper units at the end of our calculation.

and

$$\hat{G}^a(\omega) = u\,v\,e^{-i\phi}\left(\frac{f}{\omega+E+i\epsilon} - \frac{1-f}{\omega-E+i\epsilon}\right) \quad , \quad \hat{G}^c(\omega) = u\,v\,e^{+i\phi}\left(\frac{1-f}{\omega-E+i\epsilon} - \frac{f}{\omega+E+i\epsilon}\right) \tag{13.200}$$

$$\hat{G}^b(\omega) = u\,v\,e^{+i\phi}\left(\frac{1-f}{\omega+E+i\epsilon} - \frac{f}{\omega-E+i\epsilon}\right) \quad , \quad \hat{G}^d(\omega) = u\,v\,e^{+i\phi}\left(\frac{f}{\omega+E+i\epsilon} - \frac{1-f}{\omega-E+i\epsilon}\right) \quad . \tag{13.201}$$

Using the result that the Fourier transform of a product is a convolution of Fourier transforms, we have from Eqn. 13.194,

$$\hat{\chi}_{C_i C_j^\dagger}(\omega) = \frac{i}{\hbar}\,\delta_{\boldsymbol{p},\boldsymbol{k}}\,\delta_{\boldsymbol{p}',\boldsymbol{k}'}\,\delta_{\mu,\sigma}\,\delta_{\mu',\sigma'}\int\limits_{-\infty}^{\infty}\frac{d\nu}{2\pi}\left[\hat{F}^c_{\boldsymbol{k}\sigma}(\nu)\,\hat{F}^d_{\boldsymbol{k}'\sigma'}(\omega-\nu) - \hat{F}^a_{\boldsymbol{k}'\sigma'}(\nu)\,\hat{F}^b_{\boldsymbol{k}\sigma}(\omega-\nu)\right] \tag{13.202}$$

$$+\frac{i}{\hbar}\,\delta_{\boldsymbol{p},-\boldsymbol{k}'}\,\delta_{\boldsymbol{p}',-\boldsymbol{k}}\,\delta_{\mu,-\sigma'}\,\delta_{\mu',-\sigma}\int\limits_{-\infty}^{\infty}\frac{d\nu}{2\pi}\left[\hat{G}^a_{\boldsymbol{k}\sigma}(\nu)\,\hat{G}^b_{\boldsymbol{k}'\sigma'}(\omega-\nu) - \hat{G}^c_{\boldsymbol{k}'\sigma'}(\nu)\,\hat{G}^d_{\boldsymbol{k}\sigma}(\omega-\nu)\right] \quad .$$

The integrals are easily done via the contour method. For example, one has

$$i\int\limits_{-\infty}^{\infty}\frac{d\nu}{2\pi}\,\hat{F}^c_{\boldsymbol{k}\sigma}(\nu)\,\hat{F}^d_{\boldsymbol{k}'\sigma'}(\omega-\nu) = -\int\limits_{-\infty}^{\infty}\frac{d\nu}{2\pi i}\left(\frac{u^2\,f}{\nu-E+i\epsilon} + \frac{v^2\,(1-f)}{\nu+E+i\epsilon}\right)\left(\frac{u'^2\,(1-f')}{\omega-\nu+E'+i\epsilon} + \frac{v'^2\,f'}{\omega-\nu-E'+i\epsilon}\right)$$

$$= \frac{u^2\,u'^2\,(1-f)\,f'}{\omega+E-E'+i\epsilon} + \frac{v^2\,u'^2\,ff'}{\omega-E-E'+i\epsilon} + \frac{u^2\,v'^2\,(1-f)(1-f')}{\omega+E+E'+i\epsilon} + \frac{v^2\,v'^2\,f(1-f')}{\omega-E+E'+i\epsilon}\,. \tag{13.203}$$

One then finds (with proper units restored),

$$\hat{\chi}_{C_i C_j^\dagger}(\omega) = \delta_{\boldsymbol{p},\boldsymbol{k}}\,\delta_{\boldsymbol{p}',\boldsymbol{k}'}\,\delta_{\mu,\sigma}\,\delta_{\mu',\sigma'}\left(\frac{u^2u'^2\,(f-f')}{\hbar\omega-E+E'+i\epsilon} - \frac{v^2v'^2\,(f-f')}{\hbar\omega+E-E'+i\epsilon}\right. \tag{13.204}$$

$$\left. + \frac{u^2v'^2\,(1-f-f')}{\hbar\omega+E+E'+i\epsilon} - \frac{v^2u'^2\,(1-f-f')}{\hbar\omega-E-E'+i\epsilon}\right)$$

$$+\,\delta_{\boldsymbol{p},-\boldsymbol{k}'}\,\delta_{\boldsymbol{p}',-\boldsymbol{k}}\,\delta_{\mu,-\sigma'}\,\delta_{\mu',-\sigma}\left(\frac{f'-f}{\hbar\omega-E+E'+i\epsilon} - \frac{f'-f}{\hbar\omega+E-E'+i\epsilon}\right.$$

$$\left. + \frac{1-f-f'}{\hbar\omega+E+E'+i\epsilon} - \frac{1-f-f'}{\hbar\omega-E-E'+i\epsilon}\right)uvu'v'\sigma\sigma' \quad .$$

We are almost done. Note that $C_i = c^\dagger_{\boldsymbol{k}'\sigma'}c_{\boldsymbol{k}\sigma}$ means $C_i^\dagger = c^\dagger_{\boldsymbol{k}\sigma}c_{\boldsymbol{k}'\sigma'}$, hence once we have $\hat{\chi}_{C_iC_j^\dagger}(\omega)$ we can easily obtain from it $\hat{\chi}_{C_i^\dagger C_j^\dagger}(\omega)$ and the other response functions in Eqn. 13.190, simply by permuting the wavevector and spin labels.

13.6.1 Case I and case II probes

The last remaining piece in the derivation is to note that, for virtually all cases of interest,

$$\sigma\sigma' B(-\boldsymbol{k}' \; -\sigma' \,|\, -\boldsymbol{k} \; -\sigma) = \eta \, B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \quad , \tag{13.205}$$

where $B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma')$ is the transition matrix element in the original fermionic (*i.e.* 'pre-Bogoliubov') representation, from Eqn. 13.185, and where $\eta = +1$ (case I) or $\eta = -1$ (case II). The eigenvalue η tells us how the perturbation Hamiltonian transforms under the combined operations of time reversal and particle-hole transformation. The action of time reversal is

$$\mathcal{T} \,|\, \boldsymbol{k}\,\sigma \,\rangle = \sigma \,|\, -\boldsymbol{k} \; -\sigma \,\rangle \quad \Rightarrow \quad c^\dagger_{\boldsymbol{k}\sigma} \to \sigma \, c^\dagger_{-\boldsymbol{k}\,-\sigma} \tag{13.206}$$

The particle-hole transformation sends $c^\dagger_{\boldsymbol{k}\sigma} \to c_{\boldsymbol{k}\sigma}$. Thus, under the combined operation,

$$\begin{aligned} \sum_{\boldsymbol{k},\sigma} \sum_{\boldsymbol{k}',\sigma'} B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \, c^\dagger_{\boldsymbol{k}\sigma} \, c_{\boldsymbol{k}'\sigma'} \;\; &\to \;\; -\sum_{\boldsymbol{k},\sigma} \sum_{\boldsymbol{k}',\sigma'} \sigma\sigma' \, B(-\boldsymbol{k}' \; -\sigma' \,|\, -\boldsymbol{k} \; -\sigma) \, c^\dagger_{\boldsymbol{k}\sigma} \, c_{\boldsymbol{k}'\sigma'} + \text{const.} \\ &\to \;\; -\eta \sum_{\boldsymbol{k},\sigma} \sum_{\boldsymbol{k}',\sigma'} B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \, c^\dagger_{\boldsymbol{k}\sigma} \, c_{\boldsymbol{k}'\sigma'} + \text{const.} \quad . \end{aligned} \tag{13.207}$$

If we can write $B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') = B_{\sigma\sigma'}(\xi_{\boldsymbol{k}}, \xi_{\boldsymbol{k}'})$, then, further assuming that our perturbation corresponds to a definite η, we have that the power dissipated is

$$\begin{aligned} P = \tfrac{1}{2}\, g^2(\mu) \sum_{\sigma,\sigma'} \int_{-\infty}^{\infty} d\omega \, \omega \int_{-\infty}^{\infty} d\xi \int_{-\infty}^{\infty} d\xi' \, \big| B_{\sigma\sigma'}(\xi, \xi'; \omega) \big|^2 \times \\ \Big\{ \big(uu' - \eta vv'\big)^2 \, (f - f') \, \Big[\delta(\hbar\omega + E - E') + \delta(\hbar\omega + E' - E) \Big] \\ + \tfrac{1}{2} (uv' + \eta vu')^2 \, (1 - f - f') \, \Big[\delta(\hbar\omega - E - E') - \delta(\hbar\omega + E + E') \Big] \Big\} \quad . \end{aligned} \tag{13.208}$$

The coherence factors entering the above expression are

$$\begin{aligned} \tfrac{1}{2}(uu' - \eta vv')^2 &= \frac{1}{2} \left(\sqrt{\frac{E+\xi}{2E}} \sqrt{\frac{E'+\xi'}{2E'}} - \eta \sqrt{\frac{E-\xi}{2E}} \sqrt{\frac{E'-\xi'}{2E'}} \right)^2 = \frac{EE' + \xi\xi' - \eta\Delta^2}{2EE'} \\ \tfrac{1}{2}(uv' + \eta vu')^2 &= \frac{1}{2} \left(\sqrt{\frac{E+\xi}{2E}} \sqrt{\frac{E'-\xi'}{2E'}} + \eta \sqrt{\frac{E-\xi}{2E}} \sqrt{\frac{E'+\xi'}{2E'}} \right)^2 = \frac{EE' - \xi\xi' + \eta\Delta^2}{2EE'} \quad . \end{aligned} \tag{13.209}$$

Integrating over ξ and ξ' kills the $\xi\xi'$ terms, and we define the coherence factors

$$F(E, E', \Delta) \equiv \frac{EE' - \eta\Delta^2}{2EE'} \qquad , \qquad \widetilde{F}(E, E', \Delta) \equiv \frac{EE' + \eta\Delta^2}{2EE'} = 1 - F \quad . \tag{13.210}$$

case	$\hbar\omega \ll 2\Delta$	$\hbar\omega \gg 2\Delta$	$\hbar\omega \approx 2\Delta$	$\hbar\omega \gg 2\Delta$
I ($\eta = +1$)	$F \approx 0$	$F \approx \frac{1}{2}$	$\widetilde{F} \approx 1$	$\widetilde{F} \approx \frac{1}{2}$
II ($\eta = -1$)	$F \approx 1$	$F \approx \frac{1}{2}$	$\widetilde{F} \approx 0$	$\widetilde{F} \approx \frac{1}{2}$

Table 13.1: Frequency dependence of the BCS coherence factors $F(E, E+\hbar\omega, \Delta)$ and $\widetilde{F}(E, \hbar\omega - E, \Delta)$ for $E \approx \Delta$.

The behavior of $F(E, E', \Delta)$ is summarized in Tab. 13.1. If we approximate $B_{\sigma\sigma'}(\xi, \xi'; \omega) \approx B_{\sigma\sigma'}(0, 0\,; \omega)$, and we define $|\mathcal{B}(\omega)|^2 = \sum_{\sigma,\sigma'} \left| B_{\sigma\sigma'}(0, 0\,; \omega) \right|^2$, then we have

$$P = \int\limits_{-\infty}^{\infty} d\omega \, |\mathcal{B}(\omega)|^2 \, \mathcal{P}(\omega) \quad , \tag{13.211}$$

where

$$\mathcal{P}(\omega) \equiv \omega \int\limits_{\Delta}^{\infty} dE \int\limits_{\Delta}^{\infty} dE' \, \widetilde{n}_{\rm s}(E) \, \widetilde{n}_{\rm s}(E') \left\{ F(E, E', \Delta) \, (f - f') \left[\delta(\hbar\omega + E - E') + \delta(\hbar\omega + E' - E) \right] \right.$$
$$\left. + \, \widetilde{F}(E, E', \Delta) \, (1 - f - f') \left[\delta(\hbar\omega - E - E') - \delta(\hbar\omega + E + E') \right] \right\} \quad , \tag{13.212}$$

with

$$\widetilde{n}_{\rm s}(E) = \frac{g(\mu) \, |E|}{\sqrt{E^2 - \Delta^2}} \, \Theta(E^2 - \Delta^2) \quad , \tag{13.213}$$

which is the superconducting density of states from Eqn. 13.76. Note that the coherence factor for quasiparticle scattering is F, while that for quasiparticle pair creation or annihilation is $\widetilde{F} = 1 - F$.

13.6.2 Electromagnetic absorption

The interaction of light and matter is given in Eqn. 13.187. We have

$$B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') = \frac{e\hbar}{2mc} \, (\boldsymbol{k} + \boldsymbol{k}') \cdot \boldsymbol{A}_{\boldsymbol{k}-\boldsymbol{k}'} \, \delta_{\sigma\sigma'} \quad , \tag{13.214}$$

from which we see

$$\sigma\sigma' \, B(-\boldsymbol{k}' \, -\sigma' \,|\, -\boldsymbol{k} \, -\sigma) = -B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \quad , \tag{13.215}$$

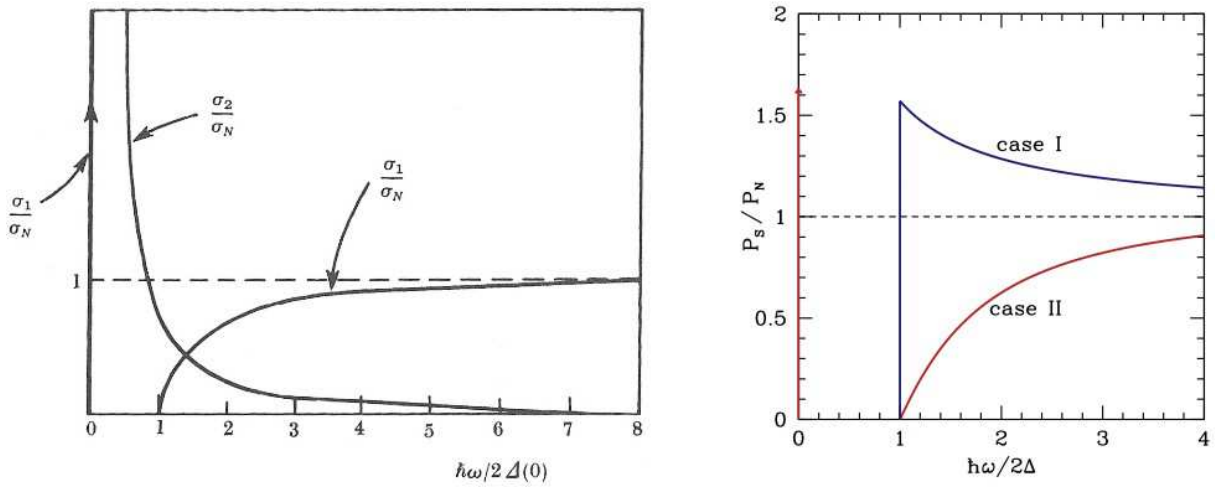


Figure 13.15: Left: real (σ_1) and imaginary (σ_2) parts of the conductivity of a superconductor, normalized by the metallic value of σ_1 just above T_c. From J. R. Schrieffer, *Theory of Superconductivity*. Right: ratio of $\mathcal{P}_s(\omega)/\mathcal{P}_N(\omega)$ for case I (blue) and case II (red) probes.

hence we have $\eta = -1$, *i.e.* case II. Let's set $T = 0$, so $f = f' = 0$. We see from Eqn. 13.212 that $\mathcal{P}(\omega) = 0$ for $\omega < 2\Delta$. We then have

$$\mathcal{P}(\omega) = \tfrac{1}{2}\, g^2(\mu) \int_\Delta^{\hbar\omega-\Delta} dE\, \frac{E(\hbar\omega - E) - \Delta^2}{\sqrt{(E^2-\Delta^2)\left((\hbar\omega - E)^2 - \Delta^2\right)}} \quad . \tag{13.216}$$

If we set $\Delta = 0$, we obtain $\mathcal{P}_N(\omega) = \frac{1}{2}\omega^2$. The ratio between superconducting and normal values is

$$\frac{\sigma_{1,s}(\omega)}{\sigma_{1,N}(\omega)} = \frac{\mathcal{P}_s(\omega)}{\mathcal{P}_N(\omega)} = \frac{1}{\omega}\int_\Delta^{\hbar\omega-\Delta} dE\, \frac{E(\hbar\omega - E) - \Delta^2}{\sqrt{(E^2-\Delta^2)\left((\hbar\omega - E)^2 - \Delta^2\right)}} \quad , \tag{13.217}$$

where $\sigma_1(\omega)$ is the real (dissipative) part of the conductivity. The result can be obtained in closed form in terms of elliptic integrals[15], and is

$$\frac{\sigma_{1,s}(\omega)}{\sigma_{1,N}(\omega)} = \left(1 + \frac{1}{x}\right)\mathbb{E}\left(\frac{1-x}{1+x}\right) - \frac{2}{x}\,\mathbb{K}\left(\frac{1-x}{1+x}\right) \quad , \tag{13.218}$$

where $x = \hbar\omega/2\Delta$. The imaginary part $\sigma_{2,s}(\omega)$ may then be obtained by Kramers-Kronig transform, and is

$$\frac{\sigma_{2,s}(\omega)}{\sigma_{1,N}(\omega)} = \frac{1}{2}\left(1 + \frac{1}{x}\right)\mathbb{E}\left(\frac{2\sqrt{x}}{1+x}\right) - \frac{1}{2}\left(1 - \frac{1}{x}\right)\mathbb{K}\left(\frac{2\sqrt{x}}{1+x}\right) \quad . \tag{13.219}$$

[15] See D. C. Mattis and J. Bardeen, *Phys. Rev.* **111**, 412 (1958).

The conductivity sum rule,

$$\int_0^\infty d\omega\,\sigma_1(\omega) = \frac{\pi n e^2}{2m} \quad , \tag{13.220}$$

is satisfied in translation-invariant systems[16]. In a superconductor, when the gap opens, the spectral weight in the region $\omega \in (0, 2\Delta)$ for case I probes shifts to the $\omega > 2\Delta$ region. One finds $\lim_{\omega\to 2\Delta^+} \mathcal{P}_{\rm S}(\omega)/\mathcal{P}_{\rm N}(\omega) = \frac{1}{2}\pi$. Case II probes, however, lose spectral weight in the $\omega > 2\Delta$ region in addition to developing a spectral gap. The missing spectral weight emerges as a delta function peak at zero frequency. The London equation $\boldsymbol{j} = -(c/4\pi\lambda_{\rm L})\,\boldsymbol{A}$ gives

$$-i\omega\,\sigma(\omega)\,\boldsymbol{E}(\omega) = -i\omega\,\boldsymbol{j}(\omega) = -\frac{c^2}{4\pi\lambda_{\rm L}^2}\,\boldsymbol{E}(\omega) \quad , \tag{13.221}$$

which says

$$\sigma(\omega) = \frac{c^2}{4\pi\lambda_{\rm L}^2}\,\frac{i}{\omega} + Q\,\delta(\omega) \quad , \tag{13.222}$$

where Q is as yet unknown[17]. We can determine the value of Q via Kramers-Kronig, *viz.*

$$\sigma_2(\omega) = -\,{\rm P}\int_{-\infty}^{\infty}\frac{d\nu}{\pi}\,\frac{\sigma_1(\nu)}{\nu-\omega} \quad , \tag{13.223}$$

where P denotes principal part. Thus,

$$\frac{c^2}{4\pi\lambda_{\rm L}^2\omega} = -Q\int_{-\infty}^{\infty}\frac{d\nu}{\pi}\,\frac{\delta(\nu)}{\nu-\omega} = \frac{Q}{\pi} \qquad \Rightarrow \qquad Q = \frac{c^2}{4\lambda_{\rm L}} \quad . \tag{13.224}$$

Thus, the full London $\sigma(\omega) = \sigma_1(\omega) + i\sigma_2(\omega)$ may be written as

$$\sigma(\omega) = \lim_{\epsilon\to 0^+}\frac{c^2}{4\lambda_{\rm L}}\,\frac{1}{\epsilon - i\pi\omega} = \frac{c^2}{4\lambda_{\rm L}}\left\{\delta(\omega) + \frac{i}{\pi\omega}\right\} \quad . \tag{13.225}$$

Note that the London form for $\sigma_1(\omega)$ includes only the delta-function and none of the structure due to thermally excited quasiparticles ($\omega < 2\Delta$) or pair-breaking ($\omega > 2\Delta$). *Nota bene: while the real part of the conductivity $\sigma_1(\omega)$ includes a $\delta(\omega)$ piece which is finite below 2Δ, because it lies at zero frequency, it does not result in any energy dissipation.* It is also important to note that the electrodynamic response in London theory is purely local. The actual electromagnetic response kernel $K_{\mu\nu}(\boldsymbol{q},\omega)$ computed using BCS theory is $\boldsymbol{q}$-dependent, even at $\omega = 0$. This says that a magnetic field $\boldsymbol{B}(\boldsymbol{x})$ will induce screening currents at positions $\boldsymbol{x}'$ which are not too

[16] Neglecting interband transitions, the conductivity sum rule is satisfied under replacement of the electron mass m by the band mass m^*.

[17] Note that $\omega\,\delta(\omega) = 0$ when multiplied by any nonsingular function in an integrand.

distant from $\boldsymbol{x}$. The relevant length scale here turns out to be the *coherence length* $\xi_0 = \hbar v_{\rm F}/\pi\Delta_0$ (at zero temperature).

At finite temperature, $\sigma_1(\omega, T)$ exhibits a Hebel-Slichter peak, also known as the *coherence peak*. Examples from two presumably non-s-wave superconductors are shown in Fig. 13.16.

Impurities and translational invariance

Observant students may notice that our derivation of $\sigma(\omega)$ makes no sense. The reason is that $B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \propto (\boldsymbol{k} + \boldsymbol{k}') \cdot \boldsymbol{A}_{\boldsymbol{k}-\boldsymbol{k}'}$, which is not of the form $B_{\sigma\sigma'}(\xi_{\boldsymbol{k}}, \xi_{\boldsymbol{k}'})$. For an electromagnetic field of frequency ω, the wavevector $q = \omega/c$ may be taken to be $q \to 0$, since the wavelength of light in the relevant range (optical frequencies and below) is enormous on the scale of the Fermi wavelength of electrons in the metallic phase. We then have that $\boldsymbol{k} = \boldsymbol{k}' + \boldsymbol{q}$, in which case the coherence factor $u_{\boldsymbol{k}}v_{\boldsymbol{k}'} - v_{\boldsymbol{k}}u_{\boldsymbol{k}'}$ vanishes as $\boldsymbol{q} \to 0$ and $\sigma_1(\omega)$ vanishes as well! This is because in the absence of translational symmetry breaking due to impurities, the current operator $\boldsymbol{j}$ commutes with the Hamiltonian, hence matrix elements of the perturbation $\boldsymbol{j} \cdot \boldsymbol{A}$ cannot cause any electronic transitions, and therefore there can be no dissipation. But this is not quite right, because the crystalline potential itself breaks translational invariance. What is true is this: *with no disorder, the dissipative conductivity $\sigma_1(\omega)$ vanishes on frequency scales below those corresponding to interband transitions*. Of course, this is also true in the metallic phase as well.

As shown by Mattis and Bardeen, if we relax the condition of momentum conservation, which is appropriate in the presence of impurities which break translational invariance, then we basically arrive back at the condition $B(\boldsymbol{k}\sigma \,|\, \boldsymbol{k}'\sigma') \approx B_{\sigma\sigma'}(\xi_{\boldsymbol{k}}, \xi_{\boldsymbol{k}'})$. One might well wonder whether we should be classifying perturbation operators by the η parity in the presence of impurities, but provided $\Delta\tau \ll \hbar$, the Mattis-Bardeen result, which we have derived above, is correct.

13.7 Electromagnetic Response of Superconductors

Here we follow chapter 8 of Schrieffer, *Theory of Superconductivity*. In chapter 10 of the lecture notes, we derived the linear response result (Eqn. 10.200)

$$\big\langle j_\mu(\boldsymbol{x}, t)\big\rangle = -\frac{c}{4\pi}\int d^3x' \int dt'\; K_{\mu\nu}(\boldsymbol{x}, t\;\; \boldsymbol{x}', t')\, A^\nu(\boldsymbol{x}', t') \quad , \tag{13.226}$$

where $\boldsymbol{j}(\boldsymbol{x}, t)$ is the electrical current density, which is a sum of paramagnetic and diamagnetic contributions, *viz.*

$$\begin{aligned}
\big\langle j^{\rm p}_\mu(\boldsymbol{x}, t)\big\rangle &= \frac{i}{\hbar c}\int d^3x' \int dt'\, \big\langle \big[j^{\rm p}_\mu(\boldsymbol{x}, t), j^{\rm p}_\nu(\boldsymbol{x}', t')\big]\big\rangle\, \Theta(t - t')\, A^\nu(\boldsymbol{x}', t') \\
\big\langle j^{\rm d}_\mu(\boldsymbol{x}, t)\big\rangle &= -\frac{e}{mc^2}\,\big\langle j^{\rm p}_0(\boldsymbol{x}, t)\big\rangle\, A^\mu(\boldsymbol{x}, t)\,(1 - \delta_{\mu 0}) \quad ,
\end{aligned} \tag{13.227}$$

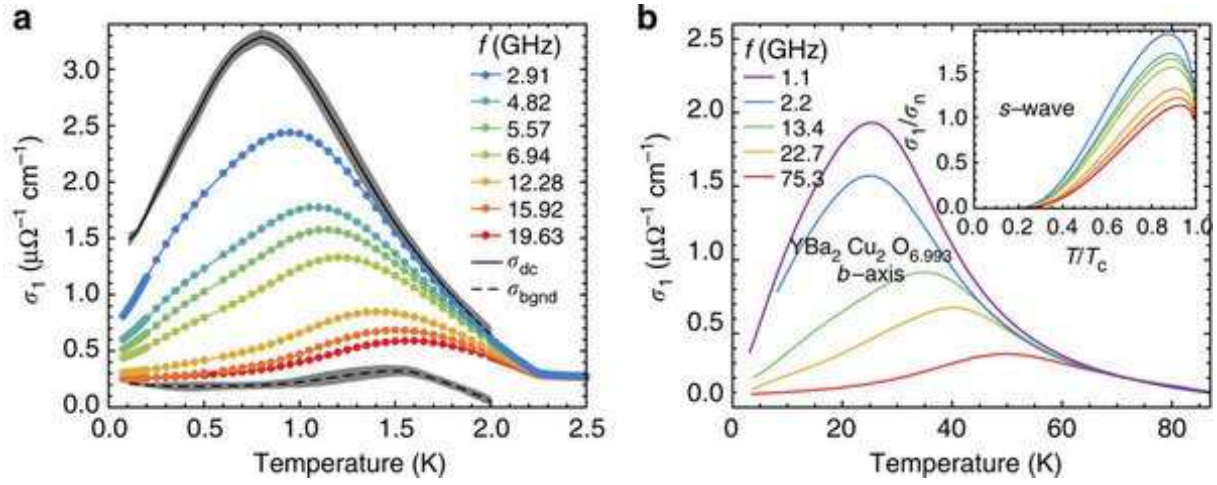


Figure 13.16: Real part of the conductivity $\sigma_1(\omega, T)$ in $CeCoIn_5$ (left; $T_c = 2.25\,\mathrm{K}$) and in $YBa_2Cu_3O_{6.993}$ (right; $T_c = 89\,\mathrm{K}$), each showing a coherence peak *versus* temperature over a range of low frequencies. Inset at right shows predictions for s-wave BCS superconductors. Both these materials are believed to involve a more exotic pairing structure. From C. J. S. Truncik *et al.*, *Nature Comm.* **4**, 2477 (2013).

with $j_0^{\rm p}(\boldsymbol{x}) = ce\, n(\boldsymbol{x})$. We then conclude[18]

$$K_{\mu\nu}(\boldsymbol{x}t;\boldsymbol{x}'t') = \frac{4\pi}{i\hbar c^2}\left\langle\left[j_\mu^{\rm p}(\boldsymbol{x},t), j_\nu^{\rm p}(\boldsymbol{x}',t')\right]\right\rangle\Theta(t-t') + \frac{4\pi e}{mc^3}\left\langle j_0^{\rm p}(\boldsymbol{x},t)\right\rangle\delta(\boldsymbol{x}-\boldsymbol{x}')\,\delta(t-t')\,\delta_{\mu\nu}\,(1-\delta_{\mu 0}) \quad . \tag{13.228}$$

In Fourier space, we may write

$$K_{\mu\nu}(\boldsymbol{q},t) = \overbrace{\frac{4\pi}{i\hbar c^2 V}\left\langle\left[j_\mu^{\rm p}(\boldsymbol{q},t), j_\nu^{\rm p}(-\boldsymbol{q},0)\right]\right\rangle\Theta(t)}^{K^{\rm p}_{\mu\nu}(\boldsymbol{q},t)} + \overbrace{\frac{4\pi ne^2}{mc^2}\,\delta(t)\,\delta_{\mu\nu}\,(1-\delta_{\mu 0})}^{K^{\rm d}_{\mu\nu}(\boldsymbol{q},t)} \quad , \tag{13.229}$$

where the paramagnetic current operator is

$$j_0^{\rm p}(\boldsymbol{q}) = ce\sum_{\boldsymbol{k},\sigma} c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}+\boldsymbol{q}\,\sigma} \qquad , \qquad \boldsymbol{j}^{\rm p}(\boldsymbol{q}) = -\frac{e\hbar}{m}\sum_{\boldsymbol{k},\sigma}\left(\boldsymbol{k}+\tfrac{1}{2}\boldsymbol{q}\right)c^\dagger_{\boldsymbol{k}\sigma}\,c_{\boldsymbol{k}+\boldsymbol{q}\,\sigma} \quad . \tag{13.230}$$

13.7.1 Computation of the electromagnetic response kernel

The calculation of the electromagnetic response kernel $K_{\mu\nu}(\boldsymbol{q},\omega)$ is tedious, but it yields all we need to know about the electromagnetic response of superconductors. For example, if we work

[18] We use a Minkowski metric $g^{\mu\nu} = g_{\mu\nu} = \mathsf{diag}(-,+,+,+)$ to raise and lower indices.

in a gauge where $A^0 = 0$, we have $\boldsymbol{E}(\omega) = i\omega \boldsymbol{A}(\omega)/c$ and hence the conductivity tensor is

$$\sigma_{ij}(\boldsymbol{q},\omega) = \frac{ic^2}{4\pi\omega}\, K_{ij}(\boldsymbol{q},\omega) \quad , \tag{13.231}$$

where i and j are spatial indices. Using the results of §13.6, the paramagnetic response kernel at $\omega = 0$ is

$$K^{\rm p}_{ij}(\boldsymbol{q},\omega=0) = -\frac{8\pi\hbar^2 e^2}{m^2c^2}\int \frac{d^3k}{(2\pi)^3}\,\big(k_i + \tfrac{1}{2}q_i\big)\big(k_j + \tfrac{1}{2}q_j\big)\,L(\boldsymbol{k},\boldsymbol{q}) \quad , \tag{13.232}$$

where

$$\begin{aligned} L(\boldsymbol{k},\boldsymbol{q}) = &\left(\frac{E_{\boldsymbol{k}}E_{\boldsymbol{k}+\boldsymbol{q}} - \xi_{\boldsymbol{k}}\xi_{\boldsymbol{k}+\boldsymbol{q}} - \Delta_{\boldsymbol{k}}\Delta_{\boldsymbol{k}+\boldsymbol{q}}}{2E_{\boldsymbol{k}}E_{\boldsymbol{k}+\boldsymbol{q}}}\right)\left(\frac{1 - f(E_{\boldsymbol{k}}) - f(E_{\boldsymbol{k}+\boldsymbol{q}})}{E_{\boldsymbol{k}} + E_{\boldsymbol{k}+\boldsymbol{q}} + i\epsilon}\right) \\ &+ \left(\frac{E_{\boldsymbol{k}}E_{\boldsymbol{k}+\boldsymbol{q}} + \xi_{\boldsymbol{k}}\xi_{\boldsymbol{k}+\boldsymbol{q}} + \Delta_{\boldsymbol{k}}\Delta_{\boldsymbol{k}+\boldsymbol{q}}}{2E_{\boldsymbol{k}}E_{\boldsymbol{k}+\boldsymbol{q}}}\right)\left(\frac{f(E_{\boldsymbol{k}+\boldsymbol{q}}) - f(E_{\boldsymbol{k}})}{E_{\boldsymbol{k}} - E_{\boldsymbol{k}+\boldsymbol{q}} + i\epsilon}\right) \quad . \end{aligned} \tag{13.233}$$

At $T = 0$, we have $f(E_{\boldsymbol{k}}) = f(E_{\boldsymbol{k}+\boldsymbol{q}}) = 0$, and only the first term contributes. As $\boldsymbol{q} \to 0$, we have $L(\boldsymbol{k}, \boldsymbol{q} \to 0) = 0$ because the coherence factor vanishes while the energy denominator remains finite. Thus, only the diamagnetic response remains, and at $T = 0$ we therefore have

$$\lim_{\boldsymbol{q}\to 0} K_{ij}(\boldsymbol{q},0) = \frac{\delta_{ij}}{\lambda_{\rm L}^2(0)} \quad , \tag{13.234}$$

where $\lambda_{\rm L}(0) = \sqrt{mc^2/4\pi ne^2}$ is London's penetration depth. According to §10.6.5, this should be purely transverse, but it is not – a defect of our mean field calculation. This can be repaired, but for our purposes it suffices to take the transverse part, *i.e.*

$$\lim_{\boldsymbol{q}\to 0} K_{ij}(\boldsymbol{q},0) = \frac{\delta_{ij} - \hat{q}_i\,\hat{q}_j}{\lambda_{\rm L}^2(0)} \quad . \tag{13.235}$$

Thus, as long as $\lambda_{\rm L}(0)$ is finite, the $\omega \to 0$ conductivity diverges.

At finite temperature, we have

$$\lim_{\boldsymbol{q}\to 0} L(\boldsymbol{k},\boldsymbol{q}) = -\left.\frac{\partial f}{\partial E}\right|_{E=E_{\boldsymbol{k}}} = \frac{1}{k_{\rm B}T}\,f(E_{\boldsymbol{k}})\big[1 - f(E_{\boldsymbol{k}})\big] \quad , \tag{13.236}$$

hence

$$\lim_{\boldsymbol{q}\to 0} K^{\rm p}_{ij}(\boldsymbol{q},\omega=0) = -\frac{8\pi\hbar^2e^2}{m^2c^2k_{\rm B}T}\int\frac{d^3k}{(2\pi)^3}\,k_i\,k_j\,\frac{e^{E_{\boldsymbol{k}}/k_{\rm B}T}}{\big(e^{E_{\boldsymbol{k}}/k_{\rm B}T}+1\big)^2} = -\frac{n_{\rm n}(T)}{n}\,\frac{1}{\lambda_{\rm L}^2(0)}\,\delta_{ij} \quad , \tag{13.237}$$

were $n_{\rm n}(T)$ is the number density of normal electrons,

$$n_{\rm n}(T) = \frac{\hbar^2}{3\pi^2 m}\int\limits_0^\infty dk\,k^4\left(-\frac{\partial f}{\partial E}\right)_{E=E_{\boldsymbol{k}}} \quad . \tag{13.238}$$

Expanding about $k = k_{\rm F}$, where $-\partial f/\partial E$ is sharply peaked at low temperatures, we find

$$\begin{aligned} n_{\rm n}(T) &= \frac{\hbar^2}{3m} \cdot 2 \int \frac{d^3k}{(2\pi)^3}\, k^2 \left(-\frac{\partial f}{\partial E} \right) \\ &= \frac{\hbar^2 k_{\rm F}^2}{3m}\, g(\varepsilon_{\rm F}) \cdot 2 \int\limits_0^\infty d\xi \left(-\frac{\partial f}{\partial E} \right) = 2n \int\limits_0^\infty d\xi \left(-\frac{\partial f}{\partial E} \right) \quad , \end{aligned} \tag{13.239}$$

which agrees precisely with what we found in Eqn. 3.136. Note that when the gap vanishes at $T_{\rm c}$, the integral yields $\frac{1}{2}$, and thus $n_{\rm n}(T_{\rm c}) = n$, as expected.

Still, Eqn. 13.237 is not fully correct, due to the failure of the mean field theory to respect gauge invariance. The correct paramagnetic response kernel in this limit is given by

$$K^{\rm p}_{ij}(\boldsymbol{q} \to 0, \omega = 0, T) = -\frac{n_{\rm n}(T)}{n} \frac{1}{\lambda_{\rm L}^2(0)}\, \delta_{ij} - \frac{n_{\rm s}(T)}{n} \frac{1}{\lambda_{\rm L}^2(0)}\, \hat{q}_i\, \hat{q}_j \tag{13.240}$$

while $K^{\rm d}_{ij}(\boldsymbol{q}, \omega, T) = \delta_{ij}/\lambda_{\rm L}^2(0)$. Here, $n_{\rm s}(T) = n - n_{\rm n}(T)$ is the density of condensed (superconducting) electrons. Thus, the spatial part of the response tensor is

$$K_{ij}(\boldsymbol{q} \to 0, \omega = 0, T) = \frac{n_{\rm s}(T)}{n} \frac{1}{\lambda_{\rm L}^2(0)} \left(\delta_{ij} - \hat{q}_i\, \hat{q}_j \right) \equiv \frac{1}{\lambda_{\rm L}^2(T)} \left(\delta_{ij} - \hat{q}_i \hat{q}_j \right) \quad , \tag{13.241}$$

which is properly transverse.

There is a slick argument, due to Landau, which yields this result. Suppose a superflow is established at some velocity $\boldsymbol{v}$. In steady state, any normal current will be damped out, and the electrical current will be $\boldsymbol{j} = -en_{\rm s}\boldsymbol{v}$. Now hop on a frame moving with the supercurrent. The superflow in the moving frame is stationary, so the current is due to normal electrons (quasiparticles), and $\boldsymbol{j}' = -en_{\rm n}(-\boldsymbol{v}) = +en_{\rm n}\boldsymbol{v}$. That is, the normal particles which were at rest in the lab frame move with velocity $-\boldsymbol{v}$ in the frame of the superflow, which we denote with a prime. The quasiparticle distribution in this primed frame is

$$f'_{\boldsymbol{k}\sigma} = \frac{1}{e^{\beta(E_{\boldsymbol{k}} + \hbar \boldsymbol{v} \cdot \boldsymbol{k})} + 1} \quad , \tag{13.242}$$

since, for a Galilean-invariant system, which we are assuming, the energy is

$$\begin{aligned} E' &= E + \boldsymbol{v} \cdot \boldsymbol{P} + \tfrac{1}{2} M \boldsymbol{v}^2 \\ &= \sum_{\boldsymbol{k},\sigma} \left(E_{\boldsymbol{k}} + \hbar \boldsymbol{k} \cdot \boldsymbol{v} \right) n_{\boldsymbol{k}\sigma} + \tfrac{1}{2} M \boldsymbol{v}^2 \quad . \end{aligned} \tag{13.243}$$

Expanding now in $\boldsymbol{v}$,

$$\begin{aligned} \boldsymbol{j}' &= -\frac{e\hbar}{mV} \sum_{\boldsymbol{k},\sigma} f'_{\boldsymbol{k}\sigma}\, \boldsymbol{k} = -\frac{e\hbar}{mV} \sum_{\boldsymbol{k},\sigma} \boldsymbol{k} \left\{ f(E_{\boldsymbol{k}}) + \hbar \boldsymbol{k} \cdot \boldsymbol{v}\, \frac{\partial f(E)}{\partial E} \bigg|_{E=E_{\boldsymbol{k}}} + \ldots \right\} \\ &= \frac{2\hbar^2 e \boldsymbol{v}}{3m} \int \frac{d^3k}{(2\pi)^3}\, k^2 \left(-\frac{\partial f}{\partial E} \right)_{E=E_{\boldsymbol{k}}} = \frac{\hbar^2 e \boldsymbol{v}}{3\pi^2 m} \int\limits_0^\infty dk\, k^4 \left(-\frac{\partial f}{\partial E} \right)_{E=E_{\boldsymbol{k}}} = \; en_{\rm n}\boldsymbol{v} \quad , \end{aligned} \tag{13.244}$$

yielding the exact same expression for $n_{\rm n}(T)$. So we conclude that $\lambda_{\rm L}^2(T) = mc^2/4\pi n_{\rm s}(T)e^2$, with $n_{\rm s}(T=0) = n$ and $n_{\rm s}(T \geqslant T_{\rm c}) = 0$. The difference $n_{\rm s}(0) - n_{\rm s}(T)$ is exponentially small in $\Delta_0/k_{\rm B}T$ for small T.

Microwave absorption measurements usually focus on the quantity $\lambda_{\rm L}(T) - \lambda_{\rm L}(0)$. A piece of superconductor effectively changes the volume – and hence the resonant frequency – of the cavity in which it is placed. Measuring the cavity resonance frequency shift $\Delta\omega_{\rm res}$ as a function of temperature allows for a determination of the difference $\Delta\lambda_{\rm L}(T) \propto \Delta\omega_{\rm res}(T)$.

Note that anything but an exponential dependence of $\Delta \ln \lambda_{\rm L}$ on $1/T$ indicates that there are low-lying quasiparticle excitations. The superconducting density of states is then replaced by

$$g_{\rm s}(E) = g_{\rm n} \int \frac{d\hat{\boldsymbol{k}}}{4\pi} \frac{E}{\sqrt{E^2 - \Delta^2(\hat{\boldsymbol{k}})}} \, \Theta\big(E^2 - \Delta^2(\hat{\boldsymbol{k}})\big) \quad , \tag{13.245}$$

where the gap $\Delta(\hat{\boldsymbol{k}})$ depends on direction in $\boldsymbol{k}$-space. If $g(E) \propto E^\alpha$ as $E \to 0$, then

$$n_{\rm n}(T) \propto \int_0^\infty dE \, g_{\rm s}(E) \left(-\frac{\partial f}{\partial E}\right) \propto T^\alpha \quad , \tag{13.246}$$

in contrast to the exponential $\exp(-\Delta_0/k_{\rm B}T)$ dependence for the s-wave (full gap) case. For example, if

$$\Delta(\hat{\boldsymbol{k}}) = \Delta_0 \, \sin^n\theta \, e^{in\varphi} \propto \Delta_0 \, Y_{nn}(\theta, \varphi) \quad , \tag{13.247}$$

then we find $g_{\rm s}(E) \propto E^{2/n}$. For $n = 2$ we would then predict a linear dependence of $\Delta \ln \lambda_{\rm L}(T)$ on T at low temperatures. Of course it is also possible to have *line nodes* of the gap function, *e.g.* $\Delta(\hat{\boldsymbol{k}}) = \Delta_0 \, (3\cos^2\theta - 1) \propto \Delta_0 \, Y_{20}(\theta, \varphi)$.

EXERCISE: Compute the energy dependence of $g_{\rm s}(E)$ when the gap function has line nodes.

13.7.2 Finite q response and the true penetration depth

Let us continue to keep $\omega = 0$ and write

$$K_{ij}(\boldsymbol{q}, \omega = 0, T) = K(\boldsymbol{q}, T) \, (\delta_{ij} - \hat{q}_i \, \hat{q}_j) \quad . \tag{13.248}$$

For $q\xi_0 \ll 1$, where $\xi_0 \equiv \hbar v_{\rm F}/\pi\Delta_0$ is the superconducting coherence length, one finds $K(\boldsymbol{q}, T) = \lambda_{\rm L}^{-2}(T)$, as we obtained above. In the limit $q\xi_0 \gg 1$, Khalatnikov and Abrikosov obtained[19]

$$K(q, T) = \frac{3\pi}{4\lambda_{\rm L}^2} \frac{\Delta(T)}{\Delta_0} \frac{1}{q\xi_0} \, \tanh\left(\frac{\Delta(T)}{2k_{\rm B}T}\right) \tag{13.249}$$

[19] I. M. Khalatnikov and A. A. Abrikosov, *Adv. Phys.* **8**, 45 (1959).

Note that this has the same q^{-1} tail as Pippard's kernel discussed in §11.3.2. The generalized penetration depth is defined to be[20]

$$\lambda(T) \equiv \frac{1}{\pi}\int\limits_{-\infty}^{\infty} dq\,\frac{1}{q^2 + K(q)} \quad , \tag{13.250}$$

from which one obtains[21]

$$\frac{\lambda(T)}{\lambda_0} = \left[\frac{\Delta_0}{\Delta(T)}\,\mathrm{ctnh}\left(\frac{\Delta(T)}{2k_{\rm B}T}\right)\right]^{1/2} \quad , \tag{13.251}$$

with

$$\lambda_0 = \frac{8\cdot 3^{1/6}}{9\cdot(2\pi)^{1/3}}\left[\xi_0\,\lambda_{\rm L}^2(0)\right]^{1/3} \quad . \tag{13.252}$$

[20] See the discussion in §49 of Fetter and Walecka.
[21] See Fetter and Walecka, §52.

Chapter 14

Magnetism

14.1 Introduction

Magnetism arises from two sources. One is the classical magnetic moment due to a current density $\boldsymbol{j}$:

$$\boldsymbol{m} = \frac{1}{2c}\int d^3r\, \boldsymbol{r}\times\boldsymbol{j} \quad . \tag{14.1}$$

The other is the intrinsic spin $\boldsymbol{S}$ of a quantum-mechanical particle (typically the electron):

$$\boldsymbol{m} = g\mu_0\boldsymbol{S}/\hbar \qquad ; \qquad \mu_0 = \frac{q\hbar}{2mc} = \text{magneton}, \tag{14.2}$$

where g is the *g-factor* (duh!). For the electron, $q=-e$ and $\mu_0 = -\mu_{\rm B}$, where $\mu_{\rm B} = e\hbar/2mc$ is the Bohr magneton.

The Hamiltonian for a single electron is

$$\hat{H} = \frac{\boldsymbol{\pi}^2}{2m} + V(\boldsymbol{r}) + \frac{e\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{H} + \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}V\times\boldsymbol{\pi} + \frac{\hbar^2}{8m^2c^2}\boldsymbol{\nabla}^2V + \frac{(\boldsymbol{\pi}^2)^2}{8m^3c^2} + \ldots \quad , \tag{14.3}$$

where $\boldsymbol{\pi} = \boldsymbol{p} + \frac{e}{c}\boldsymbol{A}$. Where did this come from? From the Dirac equation,

$$i\hbar\frac{\partial\Psi}{\partial t} = \begin{pmatrix} mc^2+V & c\,\boldsymbol{\sigma}\cdot\boldsymbol{\pi} \\ c\,\boldsymbol{\sigma}\cdot\boldsymbol{\pi} & -mc^2+V\end{pmatrix}\Psi = E\Psi \quad . \tag{14.4}$$

The wavefunction Ψ is a four-component Dirac spinor. Since mc^2 is the largest term for our applications, the upper two components of Ψ are essentially the positive energy components. However, the Dirac Hamiltonian mixes the upper two and lower two components of Ψ. One can 'unmix' them by making a canonical transformation,

$$\hat{H} \longrightarrow \hat{H}' \equiv e^{iS}\,\hat{H}\,e^{-iS} \quad , \tag{14.5}$$

where S is Hermitian, to render $\hat{H}'$ block diagonal. With $E = mc^2 + \varepsilon$, the effective Hamiltonian is given by (14.3). This is known as the Foldy-Wouthuysen transformation, the details of which may be found in many standard books on relativistic quantum mechanics and quantum field theory (*e.g.* Bjorken and Drell, Itzykson and Zuber, *etc.*) and are recited in §14.11 below. Note that the Dirac equation leads to $g = 2$. If we go beyond "tree level" and allow for radiative corrections within QED, we obtain a perturbative expansion,

$$g = 2\left\{1 + \frac{\alpha}{2\pi} + \mathcal{O}(\alpha^2)\right\} \quad , \tag{14.6}$$

where $\alpha = e^2/\hbar c \approx 1/137$ is the fine structure constant.[1]

There are two terms in (14.3) which involve the electron's spin:

$$\begin{aligned} \text{Zeeman interaction} \;:\; \hat{H}_{\rm z} &= \frac{e\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{H} \\ \text{Spin-orbit interaction} \;:\; \hat{H}_{\rm so} &= \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}V\times\left(\boldsymbol{p}+\tfrac{e}{c}\boldsymbol{A}\right) \quad . \end{aligned} \tag{14.7}$$

The numerical value for $\mu_{\rm B}$ is

$$\begin{aligned} \mu_{\rm B} &= \frac{e\hbar}{2mc} = 5.788\times10^{-9}\,{\rm eV/G} \\ \mu_{\rm B}/k_{\rm B} &= 6.717\times10^{-5}\,{\rm K/G} \quad . \end{aligned} \tag{14.8}$$

So on the scale of electron volts, laboratory scale fields ($H \lesssim 10^6$ G) are rather small. (And $\sim$2000 times smaller for nucleons!).

The thermodynamic magnetization density is defined through

$$\boldsymbol{M} = -\frac{1}{V}\frac{\partial F}{\partial \boldsymbol{H}} \quad , \tag{14.9}$$

where $F(T, V, \boldsymbol{H}, N)$ is the Helmholtz free energy. The susceptibility is then

$$\chi_{\alpha\beta}(\boldsymbol{r}\,|\,\boldsymbol{r}') = -\frac{1}{V}\,\frac{\partial^2 F}{\partial H^\alpha(\boldsymbol{r})\,\partial H^\beta(\boldsymbol{r}')} \quad . \tag{14.10}$$

When the field $\boldsymbol{H}(\boldsymbol{r}, t)$ is time-dependent, we must use time-dependent perturbation theory to compute the time-dependent susceptibility function,

$$\chi_{\alpha\beta}(\boldsymbol{r}, t\,|\,\boldsymbol{r}', t') = \frac{\delta\big\langle M^\alpha(\boldsymbol{r}, t)\big\rangle}{\delta H^\beta(\boldsymbol{r}', t')} \quad , \tag{14.11}$$

where F is replaced by a suitable generating function in the nonequilibrium case. Note that $\boldsymbol{M}$ has the dimensions of $\boldsymbol{H}$.

[1] Note that with $\mu_{\rm n} = e\hbar/2m_{\rm p}c$ for the nuclear magneton, $g_{\rm p} = 2.793$ and $g_{\rm n} = -1.913$. These results immediately suggest that there is composite structure to the nucleons, *i.e.* quarks.

14.1.1 Absence of orbital magnetism within classical physics

It is amusing to note that classical statistical mechanics cannot account for orbital magnetism. This is because the partition function is independent of the vector potential, which may be seen by simply shifting the origin of integration for the momentum $\boldsymbol{p}$:

$$\begin{aligned} Z(\boldsymbol{A}) = \mathsf{Tr}\; e^{-\beta\hat{H}} &= \int \frac{d^N\boldsymbol{r}\, d^N\boldsymbol{p}}{(2\pi\hbar)^{dN}}\, e^{-\beta\hat{H}\left(\{\boldsymbol{p}_i - \frac{q}{c}\boldsymbol{A}(\boldsymbol{r}_i),\boldsymbol{r}_i\}\right)} \\ &= \int \frac{d^N\boldsymbol{r}\, d^N\boldsymbol{p}}{(2\pi\hbar)^{dN}}\, e^{-\beta\hat{H}(\{\boldsymbol{p}_i,\boldsymbol{r}_i\})} = Z(\boldsymbol{A}=0) \quad . \end{aligned} \tag{14.12}$$

Thus, the free energy must be independent of $\boldsymbol{A}$ and hence independent of $\boldsymbol{H} = \boldsymbol{\nabla}\times\boldsymbol{A}$, and $\boldsymbol{M} = -\partial F/\partial\boldsymbol{H} = 0$. This inescapable result is known as the Bohr-von Leeuwen theorem. Of course, classical statistical mechanics can describe magnetism due to intrinsic spin, *e.g.*

$$\begin{aligned} Z_{\text{Heisenberg}}(\boldsymbol{H}) &= \prod_i \int \frac{d\hat{\boldsymbol{\Omega}}_i}{4\pi}\, e^{\beta J\sum_{\langle ij\rangle}\hat{\boldsymbol{\Omega}}_i\cdot\hat{\boldsymbol{\Omega}}_j}\; e^{\beta g\mu_0\boldsymbol{H}\cdot\sum_i\hat{\boldsymbol{\Omega}}_i} \\ Z_{\text{Ising}}(H) &= \sum_{\{\sigma_i\}} e^{\beta J\sum_{\langle ij\rangle}\sigma_i\,\sigma_j}\; e^{\beta g\mu_0 H\sum_i\sigma_i} \quad . \end{aligned} \tag{14.13}$$

Theories of magnetism generally fall into two broad classes: localized and itinerant. In the localized picture, we imagine a set of individual local moments $\boldsymbol{m}_i$ localized at different points in space (typically, though not exclusively, on lattice sites). In the itinerant picture, we focus on delocalized Bloch states which also carry electron spin.

14.2 Basic Atomic Physics

14.2.1 Single electron Hamiltonian

We start with the single-electron Hamiltonian,

$$\hat{H} = \frac{1}{2m}\left(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}\right)^2 + V(\boldsymbol{r}) + g\mu_{\text{B}}\boldsymbol{H}\cdot\boldsymbol{s}/\hbar + \frac{1}{2m^2c^2}\,\boldsymbol{s}\cdot\boldsymbol{\nabla}V\times\left(\boldsymbol{p}+\tfrac{e}{c}\boldsymbol{A}\right) \quad , \tag{14.14}$$

where the single electron spin operator is $\boldsymbol{s} = \frac{1}{2}\hbar\boldsymbol{\sigma}$. For a single atom or ion in a crystal, let us initially neglect effects due to its neighbors. In that case the potential $V(\boldsymbol{r})$ may be taken to be spherically symmetric, so with $\boldsymbol{l} = \boldsymbol{r}\times\boldsymbol{p}$, the first term in the spin-orbit part of the Hamiltonian becomes

$$\hat{H}_{\text{so}} = \frac{1}{2m^2c^2}\,\boldsymbol{s}\cdot\boldsymbol{\nabla}V\times\boldsymbol{p} = \frac{1}{2m^2c^2}\,\frac{1}{r}\frac{\partial V}{\partial r}\boldsymbol{s}\cdot\boldsymbol{l} \quad , \tag{14.15}$$

with $\boldsymbol{\nabla} V = \hat{\boldsymbol{r}}(\partial V/\partial r)$. We adopt the gauge $\boldsymbol{A} = \frac{1}{2}\boldsymbol{H} \times \boldsymbol{r}$ so that

$$\frac{1}{2m}\left(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}\right)^2 = \frac{\boldsymbol{p}^2}{2m} + \frac{e}{2mc}\,\boldsymbol{H}\cdot\boldsymbol{l} + \frac{e^2}{8mc^2}\,(\boldsymbol{H}\times\boldsymbol{r})^2 \quad . \tag{14.16}$$

Finally, restoring the full SO term, we have

$$\begin{aligned}\hat{H} = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r}) + \frac{1}{\hbar}\mu_{\rm B}(\boldsymbol{l} + 2\boldsymbol{s})\cdot\boldsymbol{H} + \frac{1}{2m^2c^2}\,\frac{1}{r}\,\frac{\partial V}{\partial r}\,\boldsymbol{l}\cdot\boldsymbol{s} \\ + \frac{e^2}{8mc^2}\,(\boldsymbol{H}\times\boldsymbol{r})^2 + \frac{\mu_{\rm B}}{\hbar}\,\frac{rV'(r)}{4mc^2}\,2\boldsymbol{s}\cdot\big[\boldsymbol{H} - \hat{\boldsymbol{r}}(\boldsymbol{H}\cdot\hat{\boldsymbol{r}})\big] \quad .\end{aligned} \tag{14.17}$$

The last term is usually negligible because $rV'(r)$ is on the scale of electron volts, while $mc^2 = 511\,\mathrm{keV}$ is the electron mass[2]. The $(\boldsymbol{H}\times\boldsymbol{r})^2$ breaks the rotational symmetry of an isolated ion, so in principal we cannot describe states by total angular momentum J. However, this effect is of order $\boldsymbol{H}^2$, so if we only desire energies to order $\boldsymbol{H}^2$, we needn't perturb the wavefunctions themselves with this term, *i.e.* we can simply treat it within first order perturbation theory, leading to an energy shift $\frac{e^2}{8mc^2}\,\langle\, n\,|\,(\boldsymbol{H}\times\boldsymbol{r})^2\,|\,n\,\rangle$ in state $|\,n\,\rangle$.

14.2.2 The Darwin term

If $V(\boldsymbol{r}) = -Ze^2/r$, then from $\boldsymbol{\nabla}^2(1/r) = -4\pi\delta(\boldsymbol{r})$ we have

$$\frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2 V = \frac{Z\pi e^2\hbar^2}{2m^2c^2}\,\delta(\boldsymbol{r}) \quad , \tag{14.18}$$

which is centered at the nucleus. This leads to an energy shift for s-wave states,

$$\Delta E_{s-\rm wave} = \frac{Z\pi e^2\hbar^2}{2m^2c^2}\,\big|\psi(0)\big|^2 = \frac{\pi}{2}\,Z\,\alpha^2 a_{\rm B}^3\,\big|\psi(0)\big|^2\cdot\frac{e^2}{a_{\rm B}} \quad , \tag{14.19}$$

where $\alpha = \frac{e^2}{\hbar c} \approx \frac{1}{137}$ is the fine structure constant and $a_{\rm B} = \frac{\hbar^2}{me^2} \approx 0.529\,\text{Å}$ is the Bohr radius. For large Z atoms and ions, the Darwin term contributes a significant contribution to the total energy.

14.2.3 Many electron Hamiltonian

The full N-electron atomic Hamiltonian, for nuclear charge Ze, is then

$$\begin{aligned}\hat{H} = \sum_{i=1}^N\left[\frac{\boldsymbol{p}_i^2}{2m} - \frac{Ze^2}{r_i}\right] + \sum_{i<j}^N\frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{i=1}^N\zeta(r_i)\,\boldsymbol{l}_i\cdot\boldsymbol{s}_i \\ + \sum_{i=1}^N\left\{\frac{\mu_{\rm B}}{\hbar}\,(\boldsymbol{l}_i + 2\boldsymbol{s}_i)\cdot\boldsymbol{H} + \frac{e^2}{8mc^2}(\boldsymbol{H}\times\boldsymbol{r}_i)^2\right\} \quad ,\end{aligned} \tag{14.20}$$

[2]Exercise: what happens in the case of high Z atoms?

where $\boldsymbol{l}_i = \boldsymbol{r}_i \times \boldsymbol{p}_i$ and

$$\zeta(r) = \frac{Ze^2}{2m^2c^2}\,\frac{1}{r^3} = \frac{Z}{\hbar^2}\left(\frac{e^2}{\hbar c}\right)^2 \frac{e^2}{2a_{\rm B}}\left(\frac{a_{\rm B}}{r}\right)^3 \quad . \tag{14.21}$$

The total orbital and spin angular momentum are $\boldsymbol{L} = \sum_i \boldsymbol{l}_i$ and $\boldsymbol{S} = \sum_i \boldsymbol{s}_i$, respectively.

The full many-electron atom is too difficult a problem to solve exactly. Generally progress is made by using the Hartree-Fock method to reduce the many-body problem to an effective one-body problem. One starts with the interacting Hamiltonian

$$\hat{H} = \sum_{i=1}^{N}\left[\frac{\boldsymbol{p}_i^2}{2m} - \frac{Ze^2}{r_i}\right] + \sum_{i<j}^{N}\frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \quad , \tag{14.22}$$

and treats $\hat{H}_{\rm so}$ as a perturbation, and writes the best possible single Slater determinant state:

$$\Psi_{\sigma_1\cdots\sigma_N}(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N) = \mathcal{A}\Big[\varphi_{1\sigma_1}(\boldsymbol{r}_1)\cdots\varphi_{N\sigma_N}(\boldsymbol{r}_N)\Big] \quad , \tag{14.23}$$

where $\mathcal{A}$ is the antisymmetrizer, and $\varphi_{i\sigma}(\boldsymbol{r})$ is a single particle wavefunction. In second-quantized notation, the Hamiltonian is

$$\hat{H} = \sum_{ij\sigma} T_{ij}^{\sigma}\,\psi_{i\sigma}^\dagger\,\psi_{j\sigma} + \sum_{\substack{ijkl\\ \sigma\sigma'}} V_{ijkl}^{\sigma\sigma'}\,\psi_{i\sigma}^\dagger\,\psi_{j\sigma'}^\dagger\,\psi_{k\sigma'}\,\psi_{l\sigma} \quad , \tag{14.24}$$

where

$$\begin{aligned} T_{ij}^{\sigma} &= \int d^3r\,\varphi_{i\sigma}^*(\boldsymbol{r})\Big\{-\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{|\boldsymbol{r}|}\Big\}\,\varphi_{j\sigma}(\boldsymbol{r}) \\ V_{ijkl}^{\sigma\sigma'} &= \tfrac{1}{2}\int d^3r\int d^3r'\,\varphi_{i\sigma}^*(\boldsymbol{r})\,\varphi_{j\sigma'}^*(\boldsymbol{r}')\,\frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\varphi_{k\sigma'}(\boldsymbol{r}')\,\varphi_{l\sigma}(\boldsymbol{r}) \quad . \end{aligned} \tag{14.25}$$

The Hartree-Fock energy is given by a sum over occupied orbitals:

$$E_{\rm HF} = \sum_{i\sigma} T_{ii}^{\sigma} + \sum_{ij\sigma\sigma'}\Big(V_{ijji}^{\sigma\sigma'} - V_{ijij}^{\sigma\sigma'}\,\delta_{\sigma\sigma'}\Big) \quad . \tag{14.26}$$

The term $V_{ijji}^{\sigma\sigma'}$ is called the direct Coulomb, or "Hartree" term, and $V_{ijij}^{\sigma\sigma'}\,\delta_{\sigma\sigma'}$ is the exchange term. Introducing Lagrange multipliers $\varepsilon_{i\sigma}$ to enforce normalization of the $\{\varphi_{i\sigma}(\boldsymbol{r})\}$ and subsequently varying with respect to the wavefunctions yields the Hartree-Fock equations:

$$0 = \left.\frac{\delta E_{\rm HF}}{\delta\varphi_{i\sigma}(\boldsymbol{r})}\right|_{\langle\Psi|\Psi\rangle=1} \qquad \Longrightarrow \tag{14.27}$$

$$\varepsilon_{i\sigma}\,\varphi_{i\sigma}(\boldsymbol{r}) = \Big\{-\frac{\hbar^2}{2m}\nabla^2 - \frac{Ze^2}{r}\Big\}\,\varphi_{i\sigma}(\boldsymbol{r}) + \sum_{j\neq i,\sigma'}^{\rm OCC}\int d^3r'\,\frac{\big|\varphi_{j\sigma'}(\boldsymbol{r}')\big|^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\varphi_{i\sigma}(\boldsymbol{r}) - \sum_{j\neq i}^{\rm OCC}\int d^3r'\,\frac{\varphi_{j\sigma}^*(\boldsymbol{r}')\,\varphi_{i\sigma}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}\,\varphi_{j\sigma}(\boldsymbol{r}) \quad ,$$

which is a set of N coupled integro-differential equations. Multiplying by $\varphi_i^*(\boldsymbol{r})$ and integrating, we find

$$\varepsilon_{i\sigma} = T_{ii}^{\sigma} + 2\sum_{j\sigma'}^{\rm OCC}\Big(V_{ijji}^{\sigma\sigma'} - V_{ijij}^{\sigma\sigma'}\,\delta_{\sigma\sigma'}\Big) \quad . \tag{14.28}$$

It is a good approximation to assume that the Hartree-Fock wavefunctions $\varphi_i(\boldsymbol{r})$ are spherically symmetric, *i.e.*

$$\varphi_{i\sigma}(\boldsymbol{r}) = R_{nl}(r)\,Y_{lm}(\theta,\phi) \quad , \tag{14.29}$$

independent of σ. We can then classify the single particle states by the quantum numbers $n \in \{1,2,\ldots\}$, $l \in \{0,1,\ldots,n-1\}$, $m_l \in \{-l,\ldots,+l\}$, and $m_s = \pm\frac{1}{2}$. The essential physics introduced by the Hartree-Fock method is that of *screening*. Close to the origin, a given electron senses a potential $-Ze^2/r$ due to the unscreened nucleus. Farther away, though, the nuclear charge is screened by the core electrons, and the potential decays faster than $1/r$. (Within the Thomas-Fermi approximation, the potential at long distances decays as $-\mathcal{C}e^2a_{\rm B}^3/r^4$, where $\mathcal{C} \simeq 100$ is a numerical factor, independent of Z.) Whereas states of different l and identical n are degenerate for the noninteracting hydrogenic atom, when the nuclear potential is screened, states of different l are no longer degenerate. Smaller l means smaller energy, since these states are localized closer to the nucleus, where the potential is large and negative and relatively unscreened. Hence, for a given n, the smaller l states fill up first. For a given l and n there are $(2s+1)\times(2l+1) = 4l+2$ states, labeled by the angular momentum and spin polarization quantum numbers m_l and m_s.

14.3 The Periodic Table

An excellent discussion is to be found in chapter 20 of G. Baym's *Lectures on Quantum Mechanics*. The eigenspectrum of single electron hydrogenic atoms is specified by quantum numbers $n \in \{1,2,\ldots\}$, $l \in \{0,1,\ldots,n-1\}$, $m_l \in \{-l,\ldots,+l\}$, and $m_s = \pm\frac{1}{2}$. The bound state energy eigenvalues $E_{nl} = -e^2/2na_{\rm B}$, where $a_{\rm B} = \hbar^2/me^2 = 0.529\,\text{Å}$ is the Bohr radius, depend only on the principal quantum number n. Accounting for electron-electron interactions using the Hartree-Fock method[3], the essential physics of *screening* is introduced, a result of which is that states of different l for a given n are no longer degenerate. Smaller l means lower energy since those states are localized closer to the nucleus, where the potential is less screened. Thus, for a given n, the smaller l states fill up first. For a given n and l, there are $(2s+1)\times(2l+1) = 4l+2$ states, labeled by m_l and m_s. This group of orbitals is called a *shell*.

[3]Hartree-Fock theory tends to overestimate ground state atomic energies by on the order of 1 eV per pair of electrons. The reason is that electron-electron correlations are not adequately represented in the Hartree-Fock many-body wavefunctions, which are single Slater determinants.



Figure 14.1: The *Aufbau* principle and the diagonal rule. Image credit: Wikipedia.

14.3.1 Aufbau principle

Based on the HF energy levels, the order in which the electron shells are filled throughout the periodic table is roughly given by that in Fig. 14.1. This is known as the *Aufbau principle* from the German *Aufbau* = "building up". The order in which the orbitals are filled follows the *diagonal rule*, which says that orbitals with lower values of $n + l$ are filled before those with higher values, and that in the case of equal $n + l$ values, the orbital with the lower n is filled first. There are hiccups here and there. For example, in filling the 3d shell of the transition metal series (row four of the PT) , ^{21}Sc, ^{22}Ti, and ^{23}V, are configured as $[\mathrm{Ar}]\,4s^2\,3d^1$, $[\mathrm{Ar}]\,4s^2\,3d^2$, and $[\mathrm{Ar}]\,4s^2\,3d^3$, respectively, but chromium's (dominant) configuration is $[\mathrm{Ar}]\,4s^1\,3d^5$. Similarly, copper is $[\mathrm{Ar}]\,4s^1\,3d^{10}$ rather than the expected $[\mathrm{Ar}]\,4s^2\,3d^9$. For palladium, the diagonal rule predicts an electronic configuration $[\mathrm{Kr}]\,5s^2\,4d^8$ whereas experiments say it is $[\mathrm{Kr}]\,5s^0\,4d^{10}$ Go figure. Again, don't take this shell configuration stuff too seriously, because the atomic ground states are really linear combinations of different shell configurations, so we should really think of these various configurations as being the dominant ones among a more general linear combination of states. Row five pretty much repeats row four, with the filling of the 5s, 4d, and 5p shells. In row six, the lanthanide (4f) series interpolates between the 6s and 5d shells, as the 5f actinide series interpolates in row seven between 7s and 6d.

Shell:	1s	2s	2p	3s	3p	4s	3d	4p	5s
Termination:	^{2}He	^{4}Be	^{10}Ne	^{12}Mg	^{18}Ar	^{20}Ca	^{30}Zn	^{36}Kr	^{38}Sr
Shell:	4d	5p	6s	4f	5d	6p	7s	5f/6d	
Termination:	^{48}Cd	^{54}Xe	^{56}Ba	^{70}Yb	^{80}Hg	^{86}Rn	^{88}Ra	^{102}No	

Table 14.1: Rough order in which shells of the Periodic Table are filled.

As we see from table 14.2, there are two anomalies in the otherwise orderly filling of the 3d shell. Chromium's configuration is $[\mathrm{Ar}]\,4s^1\,3d^5$ rather than the expected $[\mathrm{Ar}]\,4s^2\,3d^4$, and copper's is $[\mathrm{Ar}]\,4s^1\,3d^{10}$ and not $[\mathrm{Ar}]\,4s^2\,3d^9$. In reality, the ground state is not a single Slater determinant and involves linear combinations of different configurations. But the largest weights are for Cr and Cu configurations with only one $4s$ electron. Zinc terminates the 3d series, after which we get orderly filling of the 4p orbitals.

The 3d transition metal series ([Ar] core additions)					
Element (A^Z)	Sc^{21}	Ti^{22}	V^{23}	Cr^{24}	Mn^{25}
Configuration	$4s^2\,3d^1$	$4s^2\,3d^2$	$4s^2\,3d^3$	$4s^1\,3d^5$	$4s^2\,3d^5$
Element (A^Z)	Fe^{26}	Co^{27}	Ni^{28}	Cu^{29}	Zn^{30}
Configuration	$4s^2\,3d^6$	$4s^2\,3d^7$	$4s^2\,3d^8$	$4s^1\,3d^{10}$	$4s^2\,3d^{10}$

Table 14.2: Electronic configuration of 3d-series metals.

14.3.2 Splitting of configurations: Hund's rules

The electronic configuration does not uniquely specify a ground state. Consider, for example, carbon, whose configuration is $1s^2\,2s^2\,2p^2$. The filled 1s and 2s shells are inert. However, there are $\binom{6}{2} = 15$ possible ways to put two electrons in the 2p shell. It is convenient to label these states by total L, S, and J quantum numbers, where $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$ is the total angular momentum. It is standard to abbreviate each such multiplet with the label ${}^{2S+1}L_J$, called a *term*, where $L = \mathrm{S}$, P, D, F, G, H, *etc.*. For carbon, the largest L value we can get is $L = 2$, which requires $S = 0$ and hence $J = L = 2$. This 5-fold degenerate multiplet is then abbreviated ${}^1\mathrm{D}_2$. But we can also add together two $l = 1$ states to get total angular momentum $L = 1$ as well. The corresponding spatial wavefunction is antisymmetric, hence $S = 1$ in order to achieve a symmetric spin wavefunction. Since $|L - S| \leqslant J \leqslant |L + S|$ we have $J = 0$, $J = 1$, or $J = 2$ corresponding to multiplets ${}^3\mathrm{P}_0$, ${}^3\mathrm{P}_1$, and ${}^3\mathrm{P}_2$, with degeneracy 1, 3, and 5, respectively. The final state has $J = L = S = 0$: ${}^1\mathrm{S}_0$. The Hilbert space is then spanned by two $J = 0$ singlets, one $J = 1$ triplet, and two $J = 2$ quintuplets: $0 \oplus 0 \oplus 1 \oplus 2 \oplus 2$. That makes 15 states. Which of these is the ground state?

The ordering of the multiplets is determined by the famous *Hund's rules*:

1. The LS multiplet with the *largest* S has the lowest energy.
2. If the largest value of S is associated with several multiplets, the multiplet with the *largest* L has the lowest energy.

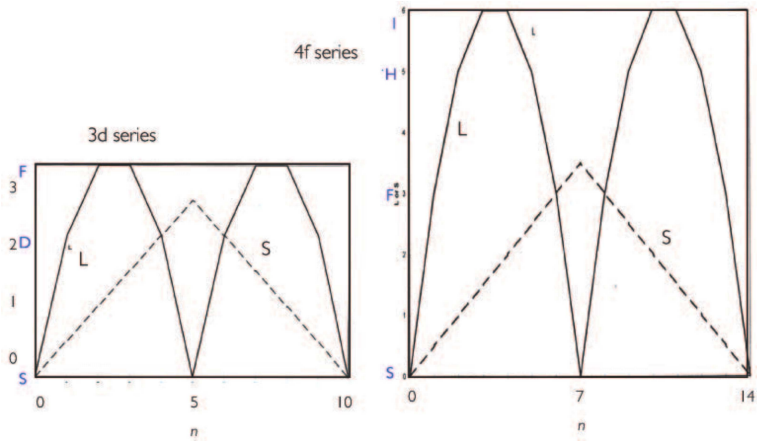


Figure 14.2: Variation of L, S, and J among the 3d and 4f series.

3. If an incomplete shell is not more than half-filled, then the lowest energy state has $J = |L - S|$. If the shell is more than half-filled, then $J = L + S$.

Hund's rules are largely empirical, but are supported by detailed atomic quantum many-body calculations. Basically, rule #1 prefers large S because this makes the spin part of the wavefunction maximally symmetric, which means that the spatial part is maximally antisymmetric. Electrons, which repel each other, prefer to exist in a spatially antisymmetric state. As for rule #2, large L expands the electron cloud somewhat, which also keeps the electrons away from each other. For neutral carbon, the ground state has $S = 1$, $L = 1$, and $J = |L - S| = 0$, hence the ground state term is ${}^3\mathrm{P}_0$.

Let's practice Hund's rules on a couple of ions:

- P: The electronic configuration for elemental phosphorus is $[\mathrm{Ne}]\,3\mathrm{s}^2\,3\mathrm{p}^3$. The unfilled 3d shell has three electrons. First maximize S by polarizing all spins parallel (up, say), yielding $S = \frac{3}{2}$. Next maximize L consistent with Pauli exclusion, which says $L = -1+0+1 = 0$. Finally, since the shell is exactly half-filled, and not more, $J = |L - S| = \frac{3}{2}$, and the ground state term is ${}^4\mathrm{S}_{3/2}$.

- Mn^{4+}: The electronic configuration $[\mathrm{Ar}]\,4\mathrm{s}^0\,3\mathrm{d}^3$ has an unfilled 3d shell with three electrons. First maximize S by polarizing all spins parallel, yielding $S = \frac{3}{2}$. Next maximize L consistent with Pauli exclusion, which says $L = 2 + 1 + 0 = 3$. Finally, since the shell is less than half-filled, $J = |L - S| = \frac{3}{2}$. The ground state term is ${}^4\mathrm{F}_{3/2}$.

- Fe^{2+}: The electronic configuration $[\mathrm{Ar}]\,4\mathrm{s}^0\,3\mathrm{d}^6$ has an unfilled 3d shell with six electrons, or four holes. First maximize S by making the spins of the holes parallel, yielding $S = 2$. Next, maximize L consistent with Pauli exclusion, which says $L = 2 + 1 + 0 + (-1) = 2$ (adding L_z for the four holes). Finally, the shell is more than half-filled, which means $J = L + S = 4$. The ground state term is ${}^5\mathrm{D}_4$.

n_p	0	1	2	3	4	5	6								
L	0	1	1	0	1	1	0								
S	0	$\frac{1}{2}$	1	$\frac{3}{2}$	1	$\frac{1}{2}$	0								
J	0	$\frac{1}{2}$	0	$\frac{3}{2}$	2	$\frac{3}{2}$	0								
n_d	0	1	2	3	4	5	6	7	8	9	10				
L	0	2	3	3	2	0	2	3	3	2	0				
S	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0				
J	0	$\frac{3}{2}$	2	$\frac{3}{2}$	0	$\frac{5}{2}$	4	$\frac{9}{2}$	4	$\frac{5}{2}$	0				
n_f	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
L	0	3	5	6	6	5	3	0	3	5	6	6	5	3	0
S	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	$\frac{7}{2}$	3	$\frac{5}{2}$	2	$\frac{3}{2}$	1	$\frac{1}{2}$	0
J	0	$\frac{5}{2}$	4	$\frac{9}{2}$	4	$\frac{5}{2}$	0	$\frac{7}{2}$	6	$\frac{15}{2}$	8	$\frac{15}{2}$	6	$\frac{7}{2}$	0

Table 14.3: Hund's rules applied to p, d, and f shells.

- Nd^{3+}: The electronic configuration $[Xe]\,6s^0\,4f^3$ has an unfilled 4f shell with three electrons. First maximize S by making the electron spins parallel, yielding $S = \frac{3}{2}$. Next, maximize L consistent with Pauli exclusion: $L = 3 + 2 + 1 = 6$. Finally, the shell is less than half-filled, we have $J = |L - S| = \frac{9}{2}$. The ground state term is ${}^4\mathrm{I}_{9/2}$.

14.3.3 Spin-orbit interaction

Hund's third rule derives from an analysis of the spin-orbit Hamiltonian,

$$\hat{H}_{\rm so} = \sum_{i=1}^{N} \zeta(r_i)\, \boldsymbol{l}_i \cdot \boldsymbol{s}_i \quad . \tag{14.30}$$

This commutes with $\boldsymbol{J}^2$, $\boldsymbol{L}^2$, and $\boldsymbol{S}^2$, so we can still classify eigenstates according to total J, L, and S. The Wigner-Eckart theorem then guarantees that within a given J multiplet, we can replace any tensor operator transforming as

$$\mathcal{R}\, T_{JM}\, \mathcal{R}^\dagger = \sum_{M'} \mathcal{D}^J_{MM'}(\alpha,\beta,\gamma)\, T_{JM'} \quad , \tag{14.31}$$

where $\mathcal{R}$ corresponds to a rotation through Euler angles α, β, and γ, by a product of a reduced matrix element and a Clebsch-Gordon coefficient:

$$\big\langle\, JM \,\big|\, T_{\boldsymbol{J}''M''} \,\big|\, J'M' \,\big\rangle = C\begin{pmatrix} J & J' & J'' \\ M & M' & M'' \end{pmatrix} \big\langle J \big|\big| T_{J''} \big|\big| J' \big\rangle \quad . \tag{14.32}$$

In other words, if two tensor operators have the same rank, their matrix elements are proportional. Both $\hat{H}_{\rm so}$ and $\boldsymbol{L}\cdot\boldsymbol{S}$ are products of rank $L=1$, $S=1$ tensor operators, hence we may replace

$$\hat{H}_{\rm so} \longrightarrow \hat{\tilde{H}}_{\rm so} = \Lambda\, \boldsymbol{L}\cdot\boldsymbol{S} \quad , \tag{14.33}$$

where $\Lambda = \Lambda(N,L,S)$ must be computed from, say, the expectation value of $\hat{H}_{\rm so}$ in the state $\big|\, JLSJ \,\big\rangle$. This requires detailed knowledge of the atomic many-body wavefunctions. However, once Λ is known, the multiplet splittings are easily obtained:

$$\begin{aligned} \hat{\tilde{H}}_{\rm so} &= \tfrac{1}{2}\Lambda\big(\boldsymbol{J}^2 - \boldsymbol{L}^2 - \boldsymbol{S}^2\big) \\ &= \tfrac{1}{2}\hbar^2\,\Lambda\,\big(J(J+1) - L(L+1) - S(S+1)\big) \quad . \end{aligned}$$

Thus,

$$E(N,L,S,J) - E(N,L,S,J-1) = \Lambda\, J\,\hbar^2 \quad . \tag{14.34}$$

If we replace $\zeta(r_i)$ by its average, then we can find Λ by the following argument. If the last shell is not more than half filled, then by Hund's first rule, the spins are all parallel. Thus we have $S = \frac{1}{2}N$ and $\boldsymbol{s}_i = \boldsymbol{S}/N$, whence $\Lambda = \langle\zeta\rangle/2S$. Finding $\langle\zeta\rangle$ is somewhat tricky. For $Z^{-1} \ll r/a_{\rm B} \ll 1$, one can use the WKB method to obtain $\psi(r = a_{\rm B}/Z) \sim \sqrt{Z}$, whence

$$\langle\zeta\rangle \sim \left(\frac{Ze^2}{\hbar c}\right)^2 \frac{me^4}{\hbar^4} \tag{14.35}$$

and $\Lambda \sim Z^2\,\alpha^2\,\hbar^{-2}\,{\rm Ry}$, where $\alpha = e^2/\hbar c \simeq 1/137$. For heavy atoms, $Z\alpha \sim 1$ and the energy is on the order of that for the outer electrons in the atom.

For shells which are more than half filled, we treat the problem in terms of the *holes* relative to the filled shell case. Since filled shells are inert,

$$\hat{H}_{\rm so} = -\sum_{j=1}^{N_{\rm h}} \zeta(r_i)\,\tilde{\boldsymbol{l}}_j\cdot\tilde{\boldsymbol{s}}_j \quad , \tag{14.36}$$

where $N_{\rm h} = 4l+2-N$. $\tilde{\boldsymbol{l}}_j$ and $\tilde{\boldsymbol{s}}_j$ are the orbital and spin angular momenta of the holes; $\boldsymbol{L} = -\sum_j \tilde{\boldsymbol{l}}_j$ and $\boldsymbol{S} = -\sum_j \tilde{\boldsymbol{s}}_j$. We then conclude $\Lambda = -\langle\zeta\rangle/2S$. Thus, we arrive at Hund's third rule, which says

$$\begin{aligned} &N \leqslant 2l+1 \quad (\leqslant \text{half-filled}) \qquad \Rightarrow \quad \Lambda > 0 \quad \Rightarrow \quad J = |L-S| \\ &N > 2l+1 \quad (> \text{half-filled}) \qquad \Rightarrow \quad \Lambda < 0 \quad \Rightarrow \quad J = |L+S| \quad . \end{aligned} \tag{14.37}$$

14.3.4 Crystal field splittings

Consider an ion with a single d electron (*e.g.* Cr^{3+}) or a single d hole (*e.g.* Cu^{2+}) in a cubic or octahedral environment. The 5-fold degeneracy of the d levels is lifted by the crystal electric field. Suppose the atomic environment is octahedral, with anions at the vertices of the octahedron (typically O^{2-} ions). In order to minimize the Coulomb repulsion between the d electron and the neighboring anions, the $d_{x^2-y^2}$ and $d_{3x^2-r^2}$ orbitals are energetically disfavored, and this doublet lies at higher energy than the $\{d_{xy}, d_{xz}, d_{yz}\}$ triplet.

The crystal field potential is crudely estimated as

$$V_{\rm CF} = \sum_{\boldsymbol{R}}^{\rm (nbrs)} V(\boldsymbol{r} - \boldsymbol{R}) \quad , \tag{14.38}$$

where the sum is over neighboring ions, and V is the atomic potential.

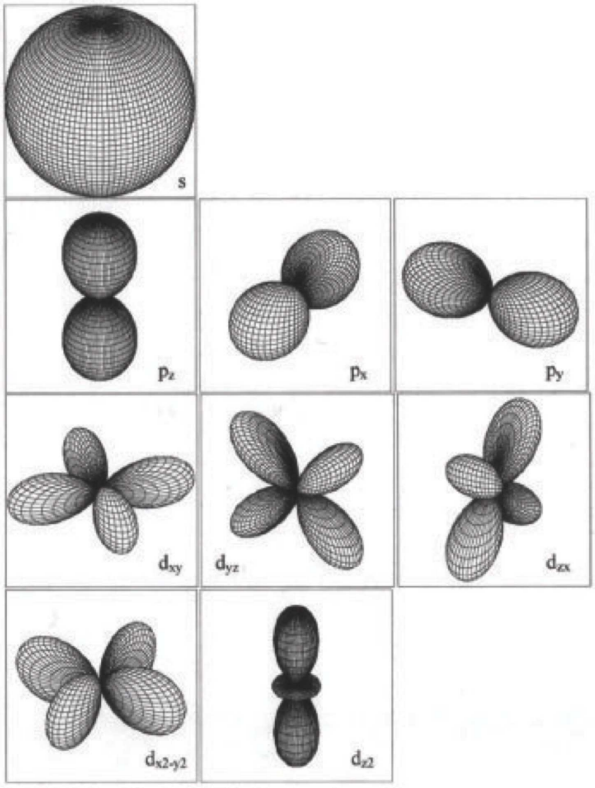


Figure 14.3: Effect on s, p, and d levels of a cubic crystal field.

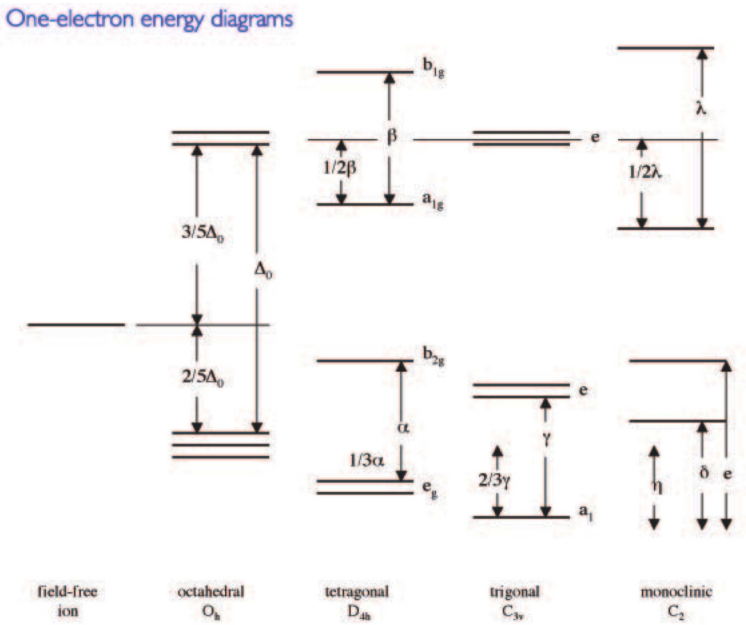


Figure 14.4: The splitting of one-electron states in different crystal field environments.

The angular dependence of the cubic crystal field states may be written as follows:

$$\begin{aligned} d_{x^2-y^2}(\hat{\boldsymbol{r}}) &= \tfrac{1}{\sqrt{2}}Y_{2,2}(\hat{\boldsymbol{r}}) + \tfrac{1}{\sqrt{2}}Y_{2,-2}(\hat{\boldsymbol{r}}) \\ d_{3z^2-r^2}(\hat{\boldsymbol{r}}) &= Y_{2,0}(\hat{\boldsymbol{r}}) \\ d_{xy}(\hat{\boldsymbol{r}}) &= \tfrac{i}{\sqrt{2}}Y_{2,-2}(\hat{\boldsymbol{r}}) - \tfrac{i}{\sqrt{2}}Y_{2,2}(\hat{\boldsymbol{r}}) \\ d_{xz}(\hat{\boldsymbol{r}}) &= \tfrac{1}{\sqrt{2}}Y_{2,1}(\hat{\boldsymbol{r}}) + \tfrac{1}{\sqrt{2}}Y_{2,-1}(\hat{\boldsymbol{r}}) \\ d_{yz}(\hat{\boldsymbol{r}}) &= \tfrac{i}{\sqrt{2}}Y_{2,-1}(\hat{\boldsymbol{r}}) - \tfrac{i}{\sqrt{2}}Y_{2,1}(\hat{\boldsymbol{r}}) \quad . \end{aligned} \tag{14.39}$$

Note that all of these wavefunctions are *real*. This means that the expectation value of L^z, and hence of general L^α, must vanish in any of these states. This is related to the phenomenon of *orbital quenching*, discussed below.

If the internal Hund's rule exchange energy $J_{\rm H}$ which enforces maximizing S is large compared with the ground state crystal field splitting Δ, then Hund's first rule is unaffected. However, there are examples of ions such as Co^{4+} for which $J_{\rm H} < V_{\rm CF}$. In such cases, the crystal field splitting wins and the ionic ground state is a *low spin state*. For Co^{4+} in an octahedral crystal field, the five 3d electrons all pile into the lower 3-fold degenerate t_{2g} manifold, and the spin is $S = \frac{1}{2}$. When the Hund's rule energy wins, the electrons all have parallel spin and $S = \frac{5}{2}$, which is the usual *high spin state*.

14.4 Magnetic Susceptibility of Atomic and Ionic Systems

To compute the susceptibility, we will need to know magnetic energies to order $\boldsymbol{H}^2$. This can be computed via perturbation theory. Treating the $\boldsymbol{H} = 0$ Hamiltonian as $\hat{H}_0$, we have

$$E_n(\boldsymbol{H}) = E_n(0) + \frac{1}{\hbar}\mu_{\rm B}\boldsymbol{H}\cdot\big\langle\, n\,\big|\,\boldsymbol{L}+2\boldsymbol{S}\,\big|\,n\,\big\rangle + \frac{e^2}{8mc^2}\big\langle\, n\,\big|\,\sum_{i=1}^{Z_{\rm ion}}(\boldsymbol{H}\times\boldsymbol{r}_i)^2\,\big|\,n\,\big\rangle + \frac{1}{\hbar^2}\,\mu_{\rm B}^2\,H^\alpha\,H^\beta\sum_{n'\neq n}\frac{\big\langle\, n\,\big|\,L^\alpha+2S^\alpha\,\big|\,n'\,\big\rangle\big\langle\, n'\,\big|\,L^\beta+2S^\beta\,\big|\,n\,\big\rangle}{E_n-E_{n'}} + \mathcal{O}(H^3) \quad , \tag{14.40}$$

where $Z_{\rm ion}$ is the number of electrons on the ion or atom in question. Since the $(\boldsymbol{H}\times\boldsymbol{r}_i)^2$ Larmor term is already second order in the field, its contribution can be evaluated in first order perturbation theory, *i.e.* by taking its expectation value in the state $\big|\,n\,\big\rangle$. The $(\boldsymbol{L}+2\boldsymbol{S})\cdot\boldsymbol{H}$ term, which is linear in the field, is treated in second order perturbation theory.

14.4.1 Filled shells: Larmor diamagnetism

If the ground state $\big|\,{\rm G}\,\big\rangle$ is a singlet with $\boldsymbol{J}\big|\,{\rm G}\,\big\rangle = \boldsymbol{L}\big|\,{\rm G}\,\big\rangle = \boldsymbol{S}\big|\,{\rm G}\,\big\rangle = 0$, corresponding to a filled shell configuration, then the only contribution to the ground state energy shift is from the Larmor term,

$$\Delta E_0(\boldsymbol{H}) = \frac{e^2\boldsymbol{H}^2}{12mc^2}\big\langle\,{\rm G}\,\big|\,\sum_{i=1}^{Z_{\rm ion}}\boldsymbol{r}_i^2\,\big|\,{\rm G}\,\big\rangle \quad , \tag{14.41}$$

and the susceptibility is

$$\chi = -\frac{N}{V}\,\frac{\partial^2\Delta E_0}{\partial H^2} = -\frac{ne^2}{6mc^2}\big\langle\,{\rm G}\,\big|\,\sum_{i=1}^{Z_{\rm ion}}\boldsymbol{r}_i^2\,\big|\,{\rm G}\,\big\rangle \quad , \tag{14.42}$$

where $n = N/V$ is the density of ions or atoms in question. The sum is over all the electrons in the ion or atom. Defining the mean square ionic radius as

$$\langle\boldsymbol{r}^2\rangle \equiv \frac{1}{Z_{\rm ion}}\big\langle\,{\rm G}\,\big|\,\sum_{i=1}^{Z_{\rm ion}}\boldsymbol{r}_i^2\,\big|\,{\rm G}\,\big\rangle \quad , \tag{14.43}$$

we obtain

$$\chi = -\frac{ne^2}{6mc^2}Z_{\rm ion}\langle\boldsymbol{r}^2\rangle = -\tfrac{1}{6}Z_{\rm ion}\,na_{\rm B}^3\left(\frac{e^2}{\hbar c}\right)^2\frac{\langle r^2\rangle}{a_{\rm B}^2} \quad . \tag{14.44}$$

Molar Susceptibilities of Noble Gas Atoms and Alkali and Halide Ions					
Atom or Ion	Molar Susceptibility	Atom or Atom or Ion	Molar Susceptibility	Atom or Atom or Ion	Molar Susceptibility
		He	-1.9	Li^+	-0.7
F^-	-9.4	Ne	-7.2	Na^+	-6.1
Cl^-	-24.2	Ar	-19.4	K^+	-14.6
Br^-	-34.5	Kr	-28	Rb^+	-22.0
I^-	-50.6	Xe	-43	Cs^+	-35.1

Table 14.4: Molar susceptibilities, in units of 10^{-6} cm^3/mol, of noble gas atoms and alkali and halide ions. (See R. Kubo and R. Nagamiya, eds., *Solid State Physics*, McGrow-Hill, 1969, p. 439.)

Note that χ is dimensionless. One defines the *molar susceptibility* as

$$\begin{aligned}\chi^{\rm molar} &\equiv N_{\rm A}\chi/n = -\tfrac{1}{6}Z_{\rm ion}N_{\rm A}a_{\rm B}^3\left(\frac{e^2}{\hbar c}\right)^2\Big\langle(\boldsymbol{r}/a_{\rm B})^2\Big\rangle \\ &= -7.91\times10^{-7}\,Z_{\rm ion}\Big\langle(\boldsymbol{r}/a_{\rm B})^2\Big\rangle\,{\rm cm}^3/{\rm mol}\quad .\end{aligned} \tag{14.45}$$

Typically, $\langle(\boldsymbol{r}/a_{\rm B})^2\rangle \sim 1$. Note that with $na_{\rm B}^3 \simeq 0.1$, we have $|\chi| \lesssim 10^{-5}$ and $M = \chi H$ is much smaller than H itself.

14.4.2 Partially filled shells: van Vleck paramagnetism

There are two cases to consider here. The first is when $J = 0$, which occurs whenever the last shell is one electron short of being half-fille. Examples include Eu^{3+} ($4f^6$), Cr^{2+} ($3d^4$), Mn^{3+} ($3d^4$), *etc.* In this case, the first order term vanishes in ΔE_0, and we have

$$\chi = -\frac{ne^2}{6mc^2}\,\langle\,{\rm G}\,|\,\sum_{i=1}^{Z_{\rm ion}}\boldsymbol{r}_i^2\,|\,{\rm G}\,\rangle + 2n\mu_{\rm B}^2\sum_{n\neq0}\frac{\big|\langle\,n\,|\,L^z+2S^z\,|\,{\rm G}\,\rangle\big|^2}{E_n-E_0}\ . \tag{14.46}$$

The second term is positive, favoring alignment of $\boldsymbol{M}$ with $\boldsymbol{H}$. This is called *van Vleck paramagnetism*, and competes with the Larmor diamagnetism.

The second possibility is $J > 0$, which occurs in all cases except filled shells and shells which are one electron short of being half-filled. In this case, the first order term is usually dominant. We label the states by the eigenvalues of the commuting observables $\{\boldsymbol{J}^2, J^z, \boldsymbol{L}^2, \boldsymbol{S}^2\}$. From the

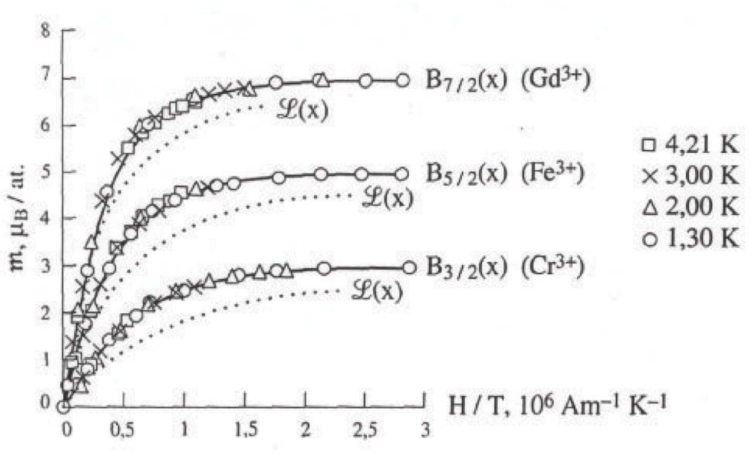


Figure 14.5: Reduced magnetization curves for three paramagnetic salts and comparison with Brillouin theory predictions. $\mathcal{L}(x) = B_{J\to\infty}(x) = \operatorname{ctnh}(x) - x^{-1}$ is the Langevin function.

Wigner-Eckart theorem, we know that

$$\left\langle\, JLSJ_z \,\middle|\, \boldsymbol{L} + 2\boldsymbol{S} \,\middle|\, JLSJ'_z \,\right\rangle = g_{\rm L}(J,L,S) \left\langle\, JLSJ_z \,\middle|\, \boldsymbol{J} \,\middle|\, JLSJ'_z \,\right\rangle \quad , \tag{14.47}$$

where

$$g_{\rm L}(J,L,S) = \tfrac{3}{2} + \frac{S(S+1) - L(L+1)}{2J(J+1)} \tag{14.48}$$

is known as the Landé g-factor. Thus, the effective Hamiltonian is

$$\hat{H}_{\rm eff} = g_{\rm L}\,\mu_{\rm B}\,\boldsymbol{J}\cdot\boldsymbol{H}/\hbar \quad . \tag{14.49}$$

The eigenvalues of $\hat{H}_{\rm eff}$ are $E_j = j\,\gamma\,H$, where $j \in \{-J, \ldots, +J\}$ and $\gamma = g_{\rm L}\,\mu_{\rm B}$. The problem is reduced to an elementary one in statistical mechanics. The partition function is

$$Z = e^{-F/k_{\rm B}T} = \sum_{j=-J}^{J} e^{-j\gamma H/k_{\rm B}T} = \frac{\sinh\left((J+\frac{1}{2})\gamma H/k_{\rm B}T\right)}{\sinh\left(\gamma H/2k_{\rm B}T\right)} \quad . \tag{14.50}$$

The magnetization density is

$$M = -\frac{N}{V}\frac{\partial F}{\partial H} = n\gamma J\, B_J(J\gamma H/k_{\rm B}T) \quad , \tag{14.51}$$

where $B_J(x)$ is the *Brillouin function*,

$$B_J(x) = \left(1+\tfrac{1}{2J}\right)\operatorname{ctnh}\left[\left(1+\tfrac{1}{2J}\right)x\right] - \tfrac{1}{2J}\operatorname{ctnh}\left(x/2J\right) \quad . \tag{14.52}$$

The magnetic susceptibility is thus

$$\begin{aligned}\chi(H,T) = \frac{\partial M}{\partial H} &= \frac{nJ^2\gamma^2}{k_{\rm B}T}\, B'_J(J\gamma H/k_{\rm B}T) \\ &= (Jg_{\rm L})^2\,(na_{\rm B}^3)\,(e^2/\hbar c)^2\left(\frac{e^2/a_{\rm B}}{k_{\rm B}T}\right) B'_J(g\mu_{\rm B}JH/k_{\rm B}T) \quad .\end{aligned} \tag{14.53}$$

Calculated and Measured Magneton Numbers of Rare Earth Ions				
Ion	Electronic Configuration	Ground State Term $^{(2S+1)}L_J$	magneton $p_{\rm theory}$	magneton $p_{\rm expt}$
La^{3+}	$[Xe]\,4f^0$	$^1S_0$	0.00	< 0
Ce^{3+}	$[Xe]\,4f^1$	$^2F_{5/2}$	2.54	2.4
Pr^{3+}	$[Xe]\,4f^2$	$^3H_4$	3.58	3.5
Nd^{3+}	$[Xe]\,4f^3$	$^4I_{9/2}$	3.62	3.5
Pm^{3+}	$[Xe]\,4f^4$	$^5I_4$	2.68	–
Sm^{3+}	$[Xe]\,4f^5$	$^6H_{5/2}$	0.84	1.5
Eu^{3+}	$[Xe]\,4f^6$	$^7F_0$	0.00	3.4
Gd^{3+}	$[Xe]\,4f^7$	$^8S_{7/2}$	7.94	8.0
Tb^{3+}	$[Xe]\,4f^8$	$^7F_6$	9.72	9.5
Dy^{3+}	$[Xe]\,4f^9$	$^6H_{15/2}$	10.63	10.6
Ho^{3+}	$[Xe]\,4f^{10}$	$^5I_8$	10.60	10.4
Er^{3+}	$[Xe]\,4f^{11}$	$^4I_{15/2}$	9.59	9.5
Tm^{3+}	$[Xe]\,4f^{12}$	$^3H_6$	7.57	7.3
Yb^{3+}	$[Xe]\,4f^{13}$	$^2F_{7/2}$	4.54	4.5
Lu^{3+}	$[Xe]\,4f^{14}$	$^1S_0$	0.00	< 0

Table 14.5: Calculated and measured effective magneton numbers p for rare earth ions. (From N. W. Ashcroft and N. D. Mermin, *Solid State Physics*.) The discrepancy in the cases of Sm and Eu is due to the existence of low-lying multiplets above the ground state.

At $H = 0$,

$$\chi(H=0,T) = \tfrac{1}{3}(g_{\rm L}\mu_{\rm B})^2\, n\, \frac{J(J+1)}{k_{\rm B}T} \quad . \tag{14.54}$$

The inverse temperature dependence is known as *Curie's law*.

Does Curie's law work in solids? The $1/T$ dependence is very accurately reflected in insulating crystals containing transition metal and rare earth ions. We can fit the coefficient of the $1/T$

Calculated and Measured Magneton Numbers of Transition Metal Ions					
Ion	Electronic Configuration	Ground State Term $^{(2S+1)}L_J$	magneton $p_{\text{theory}}^{J=\lvert L\pm S\rvert}$	magneton $p_{\text{theory}}^{J=S}$	magneton p_{expt}
Ti^{3+}	$[Ar]\,3d^1$	$^2D_{3/2}$	1.55	1.73	–
V^{4+}	$[Ar]\,3d^1$	$^2D_{3/2}$	1.55	1.73	1.8
V^{3+}	$[Ar]\,3d^2$	$^3F_2$	1.63	2.83	2.8
V^{2+}	$[Ar]\,3d^3$	$^4F_{3/2}$	0.77	3.87	3.8
Cr^{3+}	$[Ar]\,3d^3$	$^4F_{3/2}$	0.77	3.87	3.7
Mn^{4+}	$[Ar]\,3d^3$	$^4F_{3/2}$	0.77	3.87	4.0
Cr^{2+}	$[Ar]\,3d^4$	$^5D_0$	0.00	4.90	4.8
Mn^{3+}	$[Ar]\,3d^4$	$^5D_0$	0.00	4.90	5.0
Mn^{2+}	$[Ar]\,3d^5$	$^6S_{5/2}$	5.92	5.92	5.9
Fe^{3+}	$[Ar]\,3d^5$	$^6S_{5/2}$	5.92	5.92	5.9
Fe^{2+}	$[Ar]\,3d^6$	$^5D_4$	6.70	4.90	5.4
Co^{2+}	$[Ar]\,3d^7$	$^4F_{9/2}$	6.54	3.87	4.8
Ni^{2+}	$[Ar]\,3d^8$	$^3F_4$	5.59	2.83	3.2
Cu^{2+}	$[Ar]\,3d^9$	$^2D_{5/2}$	3.55	1.73	1.9

Table 14.6: Calculated and measured effective magneton numbers p for transition metal ions. (From N. W. Ashcroft and N. D. Mermin, *Solid State Physics*.) Due to the orbital quenching, the angular momentum is effectively $L = 0$.

behavior by defining the 'magneton number' p according to

$$\chi(T) = n\mu_{\rm B}^2 \, \frac{p^2}{3k_{\rm B}T} \quad . \tag{14.55}$$

The theory above predicts

$$p = g_{\rm L} \sqrt{J(J+1)} \quad . \tag{14.56}$$

One finds that the theory works well in the case of rare earth ions in solids. There, the 4f electrons of the rare earths are localized in the vicinity of the nucleus, and do not hybridize significantly with orbitals from neighboring ions.

In transition metal compounds, however, one finds poor agreement except in the case of S states ($L = 0$). This is because crystal field effects quench the orbital angular momentum, effectively rendering $L = 0$. Indeed, as shown in Table 14.6, the theory can be rescued if one ignores the ground state terms obtained by Hund's rules, and instead takes $L = 0$ and $J = S$, yielding $g_{\rm L} = 2$.

14.5 Moment Formation in Interacting Itinerant Systems

14.5.1 The Hubbard model

A noninteracting electron gas exhibits paramagnetism or diamagnetism, depending on the sign of χ, but never develops a spontaneous magnetic moment: $\boldsymbol{M}(\boldsymbol{H} = 0) = 0$. What gives rise to magnetism in solids? Overwhelmingly, the answer is that Coulomb repulsion between electrons is responsible for magnetism, in those instances in which magnetism arises. At first thought this might seem odd, since the Coulomb interaction is spin-independent. How then can it lead to a spontaneous magnetic moment?

To understand how Coulomb repulsion leads to magnetism, it is useful to consider a model interacting system, described by the Hamiltonian

$$\hat{H} = -t \sum_{\langle ij \rangle, \sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma} \right) + U \sum_i n_{i\uparrow}\, n_{i\downarrow} + \mu_{\rm B} \boldsymbol{H} \cdot \sum_{i,\alpha,\beta} c^\dagger_{i\alpha}\, \boldsymbol{\sigma}_{\alpha\beta}\, c_{i\beta} \quad . \tag{14.57}$$

This is none other than the famous *Hubbard model*, which has served as a kind of Rosetta stone for interacting electron systems. The first term describes hopping of electrons along the links of some regular lattice (the symbol $\langle ij \rangle$ denotes a link between sites i and j). The second term describes the local (on-site) repulsion of electrons. This is a single orbital model, so the repulsion exists when one tries to put two electrons in the orbital, with opposite spin polarization. Typically the Hubbard U parameter is on the order of electron volts. The last term is the Zeeman interaction of the electron spins with an external magnetic field. Orbital effects can be modeled by associating a phase $\exp(iA_{ij})$ to the hopping matrix element t between sites i and j, where the directed sum of A_{ij} around a plaquette yields the total magnetic flux through the plaquette in units of $\phi_0 = hc/e$. We will ignore orbital effects here. Note that the interaction term is short-ranged, whereas the Coulomb interaction falls off as $1/|\boldsymbol{R}_i - \boldsymbol{R}_j|$. The Hubbard model is thus unrealistic, although screening effects in metals do effectively render the interaction to be short-ranged.

Within the Hubbard model, the interaction term is local and written as $U n_\uparrow n_\downarrow$ on any given site. This term favors a local moment. This is because the chemical potential will fix the mean value of the total occupancy $n_\uparrow + n_\downarrow$, in which case it always pays to maximize the difference $|n_\uparrow - n_\downarrow|$.

14.5.2 Stoner mean field theory

There are no general methods available to solve for even the ground state of an interacting many-body Hamiltonian. We'll solve this problem using a *mean field* theory due to Stoner. The idea is to write the occupancy $n_{i\sigma}$ as a sum of average and fluctuating terms:

$$n_{i\sigma} = \langle n_{i\sigma} \rangle + \delta n_{i\sigma} \quad . \tag{14.58}$$

Here, $\langle n_{i\sigma} \rangle$ is the thermodynamic average; the above equation may then be taken as a definition of the fluctuating piece, $\delta n_{i\sigma}$. We assume that the average is site-independent. This is a significant assumption, for while we understand why each site should favor developing a moment, it is not clear that all these local moments should want to line up parallel to each other. Indeed, on a bipartite lattice, it is possible that the individual local moments on neighboring sites will be antiparallel, corresponding to an *antiferromagnetic* order of the pins. Our mean field theory will be one for *ferromagnetic* states.

We now write the interaction term as

$$\begin{aligned} n_{i\uparrow} n_{i\downarrow} &= \langle n_\uparrow \rangle \langle n_\downarrow \rangle + \langle n_\uparrow \rangle \, \delta n_{i\downarrow} + \langle n_\downarrow \rangle \, \delta n_{i\uparrow} + \overbrace{\delta n_{i\uparrow} \, \delta n_{i\downarrow}}^{(\text{fluct}^{\text{s}})^2} \\ &= -\langle n_\uparrow \rangle \langle n_\downarrow \rangle + \langle n_\uparrow \rangle \, n_{i\downarrow} + \langle n_\downarrow \rangle \, n_{i\uparrow} + \mathcal{O}\big((\delta n)^2\big) \\ &= \tfrac{1}{4}(m^2 - n^2) + \tfrac{1}{2} n \, (n_{i\uparrow} + n_{i\downarrow}) + \tfrac{1}{2} m \, (n_{i\uparrow} - n_{i\downarrow}) + \mathcal{O}\big((\delta n)^2\big) \quad , \end{aligned} \tag{14.59}$$

where n and m are the average occupancy per spin and average spin polarization, each per unit cell:

$$\begin{aligned} n &= \langle n_\downarrow \rangle + \langle n_\uparrow \rangle \\ m &= \langle n_\downarrow \rangle - \langle n_\uparrow \rangle \quad , \end{aligned} \tag{14.60}$$

i.e. $\langle n_\sigma \rangle = \frac{1}{2}(n - \sigma m)$. The mean field grand canonical Hamiltonian $\mathcal{K} = \hat{H} - \mu \mathcal{N}$, may then be written as

$$\begin{aligned} \mathcal{K}^{\text{MF}} = &-\tfrac{1}{2} \sum_{i,j,\sigma} t_{ij} \Big(c^\dagger_{i\sigma} c_{j\sigma} + c^\dagger_{j\sigma} c_{i\sigma} \Big) - \big(\mu - \tfrac{1}{2} U n\big) \sum_{i,\sigma} c^\dagger_{i\sigma} c_{i\sigma} \\ &+ \big(\mu_{\text{B}} H + \tfrac{1}{2} U m\big) \sum_{i,\sigma} \sigma \, c^\dagger_{i\sigma} c_{i\sigma} + \tfrac{1}{4} N_{\text{sites}} \, U (m^2 - n^2) \quad , \end{aligned} \tag{14.61}$$

where we've quantized spins along the direction of $\boldsymbol{H}$, defined as $\hat{\boldsymbol{z}}$. You should take note of two things here. First, the chemical potential is shifted *downward* (or the electron energies shifted *upward*) by an amount $\frac{1}{2}Un$, corresponding to the average energy of repulsion with the background. Second, the effective magnetic field has been shifted by an amount $\frac{1}{2}Um/\mu_{\text{B}}$, so the effective field is

$$H_{\text{eff}} = H + \frac{Um}{2\mu_{\text{B}}} \quad . \tag{14.62}$$

The *bare* single particle dispersions are given by $\varepsilon_\sigma(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k}) + \sigma\mu_{\rm B}H$, where

$$\hat{t}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} t(\boldsymbol{R})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \tag{14.63}$$

and $t_{ij} = t(\boldsymbol{R}_i - \boldsymbol{R}_j)$. For nearest neighbor hopping on a d-dimensional cubic lattice, we have $\hat{t}(\boldsymbol{k}) = t\sum_{\mu=1}^{d}\cos(k_\mu a)$, where a is the lattice constant. Including the mean field effects, the *effective* single particle dispersions become

$$\widetilde{\varepsilon}_\sigma(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k}) + \tfrac{1}{2}Un + \left(\mu_{\rm B}H + \tfrac{1}{2}Um\right)\sigma \quad . \tag{14.64}$$

We now solve the mean field theory, by obtaining the Gibbs free energy per site, $\varphi(n, T, H)$. First, note that $\varphi = \omega + \mu n$, where $\omega = \Omega/N_{\rm sites}$ is the Landau, or grand canonical, free energy per site. This follows from the general relation $\Omega = F - \mu N$; note that the total electron number is $N = nN_{\rm sites}$, since n is the electron number per unit cell (including both spin species). If $g(\varepsilon)$ is the density of states per unit cell (rather than per unit volume), then we have[4]

$$\varphi = \tfrac{1}{4}U(m^2+n^2) + \bar{\mu}n - \tfrac{1}{2}\,k_{\rm B}T\int_{-\infty}^{\infty}\!d\varepsilon\, g(\varepsilon)\left\{\ln\left(1+e^{(\bar{\mu}-\varepsilon-\Delta)/k_{\rm B}T}\right)+\ln\left(1+e^{(\bar{\mu}-\varepsilon+\Delta)/k_{\rm B}T}\right)\right\} \tag{14.65}$$

where $\bar{\mu} \equiv \mu - \frac{1}{2}Un$ and $\Delta \equiv \mu_{\rm B}H + \frac{1}{2}Um$. From this free energy we derive two self-consistent equations for μ and m. The first comes from demanding that φ be a function of n and not of μ, *i.e.* $\partial\varphi/\partial\mu = 0$, which leads to

$$n = \frac{1}{2}\int_{-\infty}^{\infty}\!d\varepsilon\, g(\varepsilon)\left\{f(\varepsilon-\Delta-\bar{\mu}) + f(\varepsilon+\Delta-\bar{\mu})\right\} \quad , \tag{14.66}$$

where $f(y) = \left[\exp(y/k_{\rm B}T)+1\right]^{-1}$ is the Fermi function. The second equation comes from minimizing f with respect to average moment m:

$$m = \frac{1}{2}\int_{-\infty}^{\infty}\!d\varepsilon\, g(\varepsilon)\left\{f(\varepsilon-\Delta-\bar{\mu}) - f(\varepsilon+\Delta-\bar{\mu})\right\} \quad . \tag{14.67}$$

Here, we will solve the first equation, eq. 14.66, and use the results to generate a Landau expansion of the free energy φ in powers of m^2. We assume that Δ is small, in which case we may write

$$n = \int_{-\infty}^{\infty}\!d\varepsilon\, g(\varepsilon)\left\{f(\varepsilon-\bar{\mu}) + \tfrac{1}{2}\Delta^2 f''(\varepsilon-\bar{\mu}) + \tfrac{1}{24}\Delta^4 f''''(\varepsilon-\bar{\mu}) + \ldots\right\} \quad . \tag{14.68}$$

[4]Note that we have written $\mu n = \bar{\mu}n + \frac{1}{2}Un^2$, which explains the sign of the coefficient of n^2.

We write $\bar{\mu}(\Delta) = \bar{\mu}_0 + \delta\bar{\mu}$ and expand in $\delta\bar{\mu}$. Since n is fixed in our (canonical) ensemble,

$$n = \int_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, f\big(\varepsilon - \bar{\mu}_0\big) \quad , \tag{14.69}$$

which defines $\bar{\mu}_0(n,T)$.[5] The remaining terms in the $\delta\bar{\mu}$ expansion of eqn. 14.68 must sum to zero. This yields

$$D(\bar{\mu}_0)\,\delta\bar{\mu} + \tfrac{1}{2}\Delta^2\, D'(\bar{\mu}_0) + \tfrac{1}{2}(\delta\bar{\mu})^2\, D'(\bar{\mu}_0) + \tfrac{1}{2}D''(\bar{\mu}_0)\,\Delta^2\,\delta\bar{\mu} + \tfrac{1}{24}\,D'''(\bar{\mu}_0)\,\Delta^4 + \mathcal{O}(\Delta^6) = 0 \quad , \tag{14.70}$$

where

$$D(\mu) = -\int_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\, f'(\varepsilon - \mu) \tag{14.71}$$

is the thermally averaged *bare* density of states at energy μ. Note that the $k^{\rm th}$ derivative is

$$D^{(k)}(\mu) = -\int_{-\infty}^{\infty} d\varepsilon\, g^{(k)}(\varepsilon)\, f'(\varepsilon - \mu) \quad . \tag{14.72}$$

Solving for $\delta\bar{\mu}$, we obtain

$$\delta\bar{\mu} = -\tfrac{1}{2}a_1\Delta^2 - \tfrac{1}{24}\big(3a_1^3 - 6a_1a_2 + a_3\big)\,\Delta^4 + \mathcal{O}(\Delta^6) \quad , \tag{14.73}$$

where $a_k \equiv D^{(k)}(\bar{\mu}_0)/D(\bar{\mu}_0)$. After integrating by parts and inserting this result for $\delta\bar{\mu}$ into our expression for the free energy φ, we obtain the expansion

$$\varphi(n,T,m) = \varphi_0(n,T) + \tfrac{1}{4}\,Um^2 - \tfrac{1}{2}\,D(\bar{\mu}_0)\,\Delta^2 + \frac{1}{24}\left(\frac{3\big[D'(\bar{\mu}_0)\big]^2}{D(\bar{\mu}_0)} - D''(\bar{\mu}_0)\right)\Delta^4 + \ldots \quad , \tag{14.74}$$

where prime denotes differentiation with respect to argument, at $m = 0$, and

$$\varphi_0(n,T) = \tfrac{1}{4}\,Un^2 + n\bar{\mu}_0 - \int_{-\infty}^{\infty} d\varepsilon\, \mathcal{N}(\varepsilon)\, f\big(\varepsilon - \bar{\mu}_0\big) \quad , \tag{14.75}$$

where $g(\varepsilon) = \mathcal{N}'(\varepsilon)$, so $\mathcal{N}(\varepsilon)$ is the integrated bare density of states per unit cell in the absence of any magnetic field (including both spin species).

We assume that H and m are small, in which case

$$\varphi = \varphi_0 + \tfrac{1}{2}am^2 + \tfrac{1}{4}bm^4 - \tfrac{1}{2}\chi_0\,H^2 - \frac{U\chi_0}{2\mu_{\rm B}}\,Hm + \ldots \quad , \tag{14.76}$$

[5] The Gibbs-Duhem relation guarantees that such an equation of state exists, relating any three intensive thermodynamic quantities.

where $\chi_0 = \mu_{\rm B}^2\, D(\bar\mu_0)$ is the Pauli susceptibility, and

$$a = \tfrac{1}{2}\,U\big(1 - \tfrac{1}{2}\,UD\big) \quad , \quad b = \frac{1}{96}\left(\frac{3(D')^2}{D} - D''\right) U^4 \quad , \tag{14.77}$$

where the argument of each $D^{(k)}$ above is $\bar\mu_0(n,T)$. The magnetization density (per unit cell, rather than per unit volume) is given by

$$M = -\frac{\partial\varphi}{\partial H} = \chi_0 H + \frac{U\chi_0}{2\mu_{\rm B}}\, m \quad . \tag{14.78}$$

Minimizing with respect to m yields

$$am + bm^3 - \frac{U\chi_0}{2\mu_{\rm B}}\, H = 0 \quad , \tag{14.79}$$

which gives, for small m,

$$m = \frac{\chi_0}{\mu_{\rm B}}\,\frac{H}{1 - \frac{1}{2}UD} \quad . \tag{14.80}$$

We therefore obtain $M = \chi H$ with

$$\chi = \frac{\chi_0}{1 - (U/U_{\rm c})} \quad , \tag{14.81}$$

where $U_{\rm c} = 2/D(\bar\mu_0)$. The denominator of χ increases the susceptibility above the bare Pauli value χ_0, and is referred to as – I kid you not – the *Stoner enhancement* (see Fig. 14.6).

It is worth emphasizing that the magnetization per unit cell is given by

$$M = -\frac{1}{N_{\rm sites}}\,\frac{\delta\hat H}{\delta H} = \mu_{\rm B} m \quad . \tag{14.82}$$



Figure 14.6: A graduate student experiences the Stoner enhancement.

This is an operator identity and is valid for any value of m, and not only small m.

When $H = 0$ we can still get a magnetic moment, provided $U > U_{\rm c}$. This is a consequence of the simple Landau theory we have derived. Solving for m when $H = 0$ gives $m = 0$ when $U < U_{\rm c}$ and

$$m(U) = \pm\left(\frac{U}{2b\,U_{\rm c}}\right)^{1/2}\sqrt{U - U_{\rm c}} \quad , \tag{14.83}$$

when $U > U_{\rm c}$, and assuming $b > 0$. Thus we have the usual mean field order parameter exponent of $\beta = \frac{1}{2}$.

14.5.3 Antiferromagnetic solution

In addition to ferromagnetism, there may be other ordered states which solve the mean field theory. One such example is antiferromagnetism. On a bipartite lattice, the antiferromagnetic mean field theory is obtained from

$$\langle n_{i\sigma}\rangle = \tfrac{1}{2}n + \tfrac{1}{2}\sigma\, e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i}\, m \quad , \tag{14.84}$$

where $\boldsymbol{Q} = (\pi/a, \pi/a, \ldots, \pi/a)$ is the antiferromagnetic ordering wavevector. The grand canonical Hamiltonian is then

$$\begin{aligned}\mathcal{K}^{\rm MF} &= -\tfrac{1}{2}\sum_{i,j,\sigma} t_{ij}\left(c^\dagger_{i\sigma}c_{j\sigma} + c^\dagger_{j\sigma}c_{i\sigma}\right) - \left(\mu - \tfrac{1}{2}Un\right)\sum_{i,\sigma} c^\dagger_{i\sigma}c_{i\sigma} \\ &\qquad + \tfrac{1}{2}Um\sum_{i,\sigma} e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i}\,\sigma\, c^\dagger_{i\sigma}c_{i\sigma} + \tfrac{1}{4}N_{\rm sites}\,U(m^2 - n^2) \qquad (14.85)\\ &= \tfrac{1}{2}\sum_{\boldsymbol{k},\sigma}\begin{pmatrix} c^\dagger_{\boldsymbol{k},\sigma} & c^\dagger_{\boldsymbol{k}+\boldsymbol{Q},\sigma}\end{pmatrix}\begin{pmatrix}\varepsilon(\boldsymbol{k}) - \mu + \frac{1}{2}Un & \frac{1}{2}\sigma\,Um \\ \frac{1}{2}\sigma\,Um & \varepsilon(\boldsymbol{k}+\boldsymbol{Q}) - \mu + \frac{1}{2}Un\end{pmatrix}\begin{pmatrix} c_{\boldsymbol{k},\sigma} \\ c_{\boldsymbol{k}+\boldsymbol{Q},\sigma}\end{pmatrix} + \tfrac{1}{4}N_{\rm sites}\,U(m^2 - n^2)\end{aligned}$$

where $\varepsilon(\boldsymbol{k}) = -\hat{t}(\boldsymbol{k})$, as before. On a bipartite lattice, with nearest neighbor hopping only, we have $\varepsilon(\boldsymbol{k}+\boldsymbol{Q}) = -\varepsilon(\boldsymbol{k})$. The above matrix is diagonalized by a unitary transformation, yielding the eigenvalues

$$\lambda_\pm = \pm\sqrt{\varepsilon^2(\boldsymbol{k}) + \Delta^2} - \bar{\mu} \tag{14.86}$$

with $\Delta = \frac{1}{2}Um$ and $\bar{\mu} = \mu - \frac{1}{2}Un$ as before. The free energy per site is then

$$\begin{aligned}\varphi = \tfrac{1}{4}U(m^2 + n^2) + \bar{\mu}n - \tfrac{1}{2}k_{\rm B}T\int_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\,\Big\{&\ln\left(1 + e^{\left(\bar{\mu} - \sqrt{\varepsilon^2+\Delta^2}\right)/k_{\rm B}T}\right) \\ &+ \ln\left(1 + e^{\left(\bar{\mu} + \sqrt{\varepsilon^2+\Delta^2}\right)/k_{\rm B}T}\right)\Big\} \quad .\end{aligned} \tag{14.87}$$

The mean field equations are then

$$\begin{aligned} n &= \frac{1}{2}\int_{-\infty}^{\infty} d\varepsilon\, g(\varepsilon)\left\{ f\left(-\sqrt{\varepsilon^2+\Delta^2}-\bar{\mu}\right) + f\left(\sqrt{\varepsilon^2+\Delta^2}-\bar{\mu}\right)\right\} \\ \frac{1}{U} &= \frac{1}{2}\int_{-\infty}^{\infty} d\varepsilon\, \frac{g(\varepsilon)}{\sqrt{\varepsilon^2+\Delta^2}}\left\{ f\left(-\sqrt{\varepsilon^2+\Delta^2}-\bar{\mu}\right) - f\left(\sqrt{\varepsilon^2+\Delta^2}-\bar{\mu}\right)\right\}. \end{aligned} \tag{14.88}$$

As in the case of the ferromagnet, a paramagnetic solution with $m = 0$ always exists, in which case the second of the above equations is no longer valid.

14.5.4 Mean field phase diagram of the Hubbard model

Let us compare the mean field theories for the ferromagnetic and antiferromagnetic states at $T = 0$ and $H = 0$. Due to particle-hole symmetry, we may assume $0 \leqslant n \leqslant 1$ without loss of generality. (The solutions repeat themselves under $n \to 2 - n$.) For the paramagnet, we have

$$\begin{aligned} n &= \int_{-\infty}^{\bar{\mu}} d\varepsilon\, g(\varepsilon) \\ \varphi &= \tfrac{1}{4}Un^2 + \int_{-\infty}^{\bar{\mu}} d\varepsilon\, g(\varepsilon)\,\varepsilon \quad , \end{aligned} \tag{14.89}$$

with $\bar{\mu} = \mu - \frac{1}{2}Un$ is the 'renormalized' Fermi energy and $g(\varepsilon)$ is the density of states per unit cell in the absence of any explicit (H) or implicit (m) symmetry breaking, including both spin polarizations.

For the ferromagnet,

$$\begin{aligned} n &= \tfrac{1}{2}\int_{-\infty}^{\bar{\mu}-\Delta} d\varepsilon\, g(\varepsilon) + \tfrac{1}{2}\int_{-\infty}^{\bar{\mu}+\Delta} d\varepsilon\, g(\varepsilon) \\ \frac{4\Delta}{U} &= \int_{\bar{\mu}-\Delta}^{\bar{\mu}+\Delta} d\varepsilon\, g(\varepsilon) \\ \varphi &= \tfrac{1}{4}Un^2 - \frac{\Delta^2}{U} + \int_{-\infty}^{\bar{\mu}-\Delta} d\varepsilon\, g(\varepsilon)\,\varepsilon + \int_{-\infty}^{\bar{\mu}+\Delta} d\varepsilon\, g(\varepsilon)\,\varepsilon \quad . \end{aligned} \tag{14.90}$$

Here, $\Delta = \frac{1}{2}Um$ is nonzero in the ordered phase.

Finally, the antiferromagnetic mean field equations are

$$n_{\bar{\mu}<0} = \int_{\varepsilon_0}^{\infty} d\varepsilon\, g(\varepsilon) \qquad ; \qquad n_{\bar{\mu}>0} = 2 - \int_{\varepsilon_0}^{\infty} d\varepsilon\, g(\varepsilon)$$
$$\frac{2}{U} = \int_{\varepsilon_0}^{\infty} d\varepsilon\, \frac{g(\varepsilon)}{\sqrt{\varepsilon^2 + \Delta^2}} \tag{14.91}$$
$$\varphi = \tfrac{1}{4} U n^2 + \frac{\Delta^2}{U} - \int_{\varepsilon_0}^{\infty} d\varepsilon\, g(\varepsilon)\, \sqrt{\varepsilon^2 + \Delta^2} \quad ,$$

where $\varepsilon_0 = \sqrt{\bar{\mu}^2 - \Delta^2}$ and $\Delta = \frac{1}{2} U m$ as before. Note that $|\bar{\mu}| \geqslant \Delta$ for these solutions. Exactly at half-filling, we have $n = 1$ and $\bar{\mu} = 0$. We then set $\varepsilon_0 = 0$.

The paramagnet to ferromagnet transition may be first or second order, depending on the details of $g(\varepsilon)$. If second order, it occurs at $U_{\rm c}^{\rm F} = 1\big/g(\bar{\mu}_{\rm P})$, where $\bar{\mu}_{\rm P}(n)$ is the paramagnetic solution for $\bar{\mu}$. The paramagnet to antiferromagnet transition is always second order in this mean field theory, since the RHS of eqn. (14.91) is a monotonic function of Δ. This transition occurs at $U_{\rm c}^{\rm A} = 2\Big/\int_{\bar{\mu}_{\rm P}}^{\infty} d\varepsilon\, g(\varepsilon)\, \varepsilon^{-1}$. Note that $U_{\rm c}^{\rm A} \to 0$ logarithmically for $n \to 1$, since $\bar{\mu}_{\rm P} = 0$ at half-filling.

For large U, the ferromagnetic solution always has the lowest energy, and therefore if $U_{\rm c}^{\rm A} < U_{\rm c}^{\rm F}$, there will be a first-order antiferromagnet to ferromagnet transition at some value $U^* > U_{\rm c}^{\rm F}$. In fig. 14.7, I plot the phase diagram obtained by solving the mean field equations assuming a semicircular density of states $g(\varepsilon) = \frac{2}{\pi} W^{-2} \sqrt{W^2 - \varepsilon^2}$. Also shown is the phase diagram for the $d = 2$ square lattice Hubbard model obtained by J. Hirsch (1985).

How well does Stoner theory describe the physics of the Hubbard model? Quantum Monte Carlo calculations by J. Hirsch (1985) found that the actual phase diagram of the $d = 2$ square lattice Hubbard Model exhibits no ferromagnetism for any n up to $U = 10$. Furthermore, he found the antiferromagnetic phase to be entirely confined to the vertical line $n = 1$. For $n \neq 1$ and $0 \leqslant U \leqslant 10$, the system is a paramagnet[6]. These results were state-of-the art at the time, but both computing power as well as numerical algorithms for interacting quantum systems have advanced considerably since 1985. Yet as of 2018, we *still* don't have a clear understanding of the $d = 2$ Hubbard model's $T = 0$ phase diagram! There is an emerging body of numerical evidence[7] that in the underdoped ($n < 1$) regime, there are portions of the phase diagram which exhibit a *stripe* ordering, in which antiferromagnetic order is interrupted by a parallel array of line defects containing excess holes (*i.e.* the absence of an electron)[8]. This problem

[6] A theorem due to Nagaoka establishes that the ground state is ferromagnetic for the case of a single hole in the $U = \infty$ system on bipartite lattices.

[7] See J. P. F. LeBlanc *et al.*, *Phys. Rev. X* **5**, 041041 (2015) and B. Zheng *et al.*, *Science* **358**, 1155 (2017).

[8] The best case for stripe order has been made at $T = 0$, $U/t = 8$, and hold doping $x = \frac{1}{8}$ (*i.e.* $n = \frac{7}{8}$).

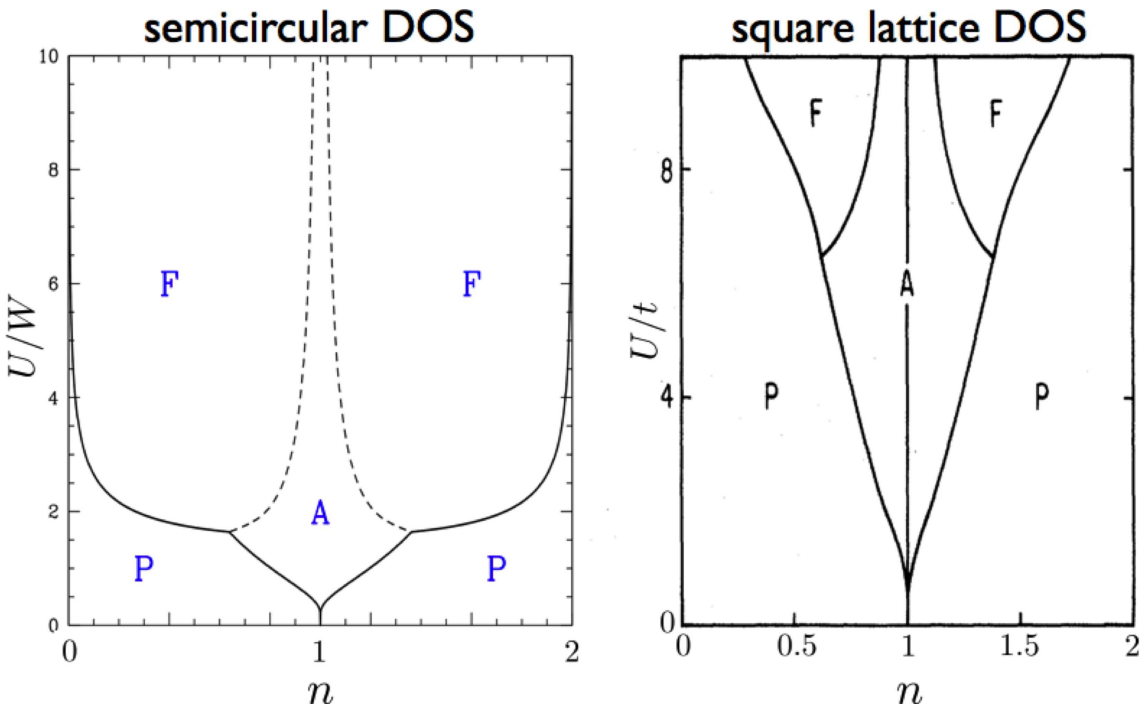


Figure 14.7: Mean field phase diagram of the Hubbard model, including paramagnetic (P), ferromagnetic (F), and antiferromagnetic (A) phases. Left panel: results using a semicircular density of states function of half-bandwidth W. Right panel: results using a two-dimensional square lattice density of states with nearest neighbor hopping t, from J. E. Hirsch, *Phys. Rev. B* **31**, 4403 (1985). The phase boundary between F and A phases is first order.

has turned out to be unexpectedly rich, complex, and numerically difficult to resolve due to the presence of *competing ordered states*, such as d-wave superconductivity and spiral magnetic phases, which lie nearby in energy with respect to the putative stripe ground state.

In order to achieve a ferromagnetic solution, it appears necessary to introduce geometric frustration, either by including a next-nearest-neighbor hopping amplitude t' or by defining the model on non-bipartite lattices. Numerical work by M. Ulmke (1997) showed the existence of a ferromagnetic phase at $T = 0$ on the FCC lattice Hubbard model for $U = 6$ and $n \in [0.15, 0.87]$ (approximately).

14.6 Interaction of Local Moments: the Heisenberg Model

While it is true that electrons have magnetic dipole moments, the corresponding dipole-dipole interactions in solids are usually negligible. This is easily seen by estimating the energy scale

of the dipole-dipole interaction:

$$E_{\rm d-d} = \frac{\boldsymbol{m}_1 \cdot \boldsymbol{m}_2 - 3(\boldsymbol{m}_1 \cdot \hat{\boldsymbol{n}})(\boldsymbol{m}_2 \cdot \hat{\boldsymbol{n}})}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|^3} \quad , \tag{14.92}$$

where $\hat{\boldsymbol{n}} = (\boldsymbol{r}_2 - \boldsymbol{r}_1)/|\boldsymbol{r}_2 - \boldsymbol{r}_1|$ is the direction vector pointing from $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$. Substituting $\boldsymbol{m} = -\mu_{\rm B}\,\boldsymbol{\sigma}$, we estimate $E_{\rm d-d}$ as

$$|E_{\rm d-d}| \simeq \frac{\mu_{\rm B}^2}{R^3} = \frac{e^2}{2a_{\rm B}} \left(\frac{e^2}{\hbar c}\right)^2 \left(\frac{a_{\rm B}}{R}\right)^3 \quad , \tag{14.93}$$

and with $R \simeq 2.5$Å we obtain $E_{\rm d-d} \simeq 1\,\mu$eV, which is tiny on the scale of electronic energies in solids. The dominant magnetic coupling comes from the Coulomb interaction.

14.6.1 Ferromagnetic exchange of orthogonal orbitals

In the Wannier basis, we may write the Coulomb interaction as

$$\hat{V} = \tfrac{1}{2} \sum_{\substack{\boldsymbol{R}_1, \boldsymbol{R}_2 \\ \boldsymbol{R}_3, \boldsymbol{R}_4}} \sum_{\sigma, \sigma'} \left\langle\, \boldsymbol{R}_1\, \boldsymbol{R}_2 \,\right| \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \left|\, \boldsymbol{R}_4\, \boldsymbol{R}_3 \,\right\rangle c^\dagger_{\boldsymbol{R}_1 \sigma}\, c^\dagger_{\boldsymbol{R}_2 \sigma'}\, c_{\boldsymbol{R}_3 \sigma'}\, c_{\boldsymbol{R}_4 \sigma} \quad , \tag{14.94}$$

where we have assumed a single energy band. The Coulomb matrix element is

$$\left\langle\, \boldsymbol{R}_1\, \boldsymbol{R}_2 \,\right| \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \left|\, \boldsymbol{R}_4\, \boldsymbol{R}_3 \,\right\rangle = \int d^3r \int d^3r'\, \varphi^*(\boldsymbol{r} - \boldsymbol{R}_1)\, \varphi^*(\boldsymbol{r}' - \boldsymbol{R}_2)\, \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|}\, \varphi(\boldsymbol{r}' - \boldsymbol{R}_3)\, \varphi(\boldsymbol{r} - \boldsymbol{R}_4) \quad . \tag{14.95}$$

Due to overlap factors, the matrix element will be small unless $\boldsymbol{R}_2 = \boldsymbol{R}_3$ and $\boldsymbol{R}_1 = \boldsymbol{R}_4$, in which case we obtain the *direct* Coulomb interaction,

$$\begin{aligned} V(\boldsymbol{R} - \boldsymbol{R}') &= \left\langle\, \boldsymbol{R}\, \boldsymbol{R}' \,\right| \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \left|\, \boldsymbol{R}\, \boldsymbol{R}' \,\right\rangle \\ &= \int d^3r \int d^3r'\, \left|\varphi(\boldsymbol{r} - \boldsymbol{R})\right|^2 \frac{e^2}{|\boldsymbol{r} - \boldsymbol{r}'|} \left|\varphi(\boldsymbol{r}' - \boldsymbol{R}')\right|^2 \quad . \end{aligned} \tag{14.96}$$

The direct interaction decays as $|\boldsymbol{R} - \boldsymbol{R}'|^{-1}$ at large separations. Were we to include only these matrix elements, the second quantized form of the Coulomb interaction would be

$$\begin{aligned} \hat{V}_{\rm direct} &= \tfrac{1}{2} \sum_{\substack{\boldsymbol{R}\boldsymbol{R}' \\ \sigma\sigma'}} V(\boldsymbol{R} - \boldsymbol{R}') \left(n_{\boldsymbol{R}\sigma}\, n_{\boldsymbol{R}'\sigma'} - \delta_{\boldsymbol{R}\boldsymbol{R}'}\, \delta_{\sigma\sigma'}\, n_{\boldsymbol{R}\sigma} \right) \\ &= \sum_{\boldsymbol{R}} V(0)\, n_{\boldsymbol{R}\uparrow}\, n_{\boldsymbol{R}\downarrow} + \tfrac{1}{2} \sum_{\boldsymbol{R} \neq \boldsymbol{R}'} V(\boldsymbol{R} - \boldsymbol{R}')\, n_{\boldsymbol{R}}\, n_{\boldsymbol{R}'} \quad , \end{aligned} \tag{14.97}$$

where $n_{\boldsymbol{R}} \equiv n_{\boldsymbol{R}\uparrow} + n_{\boldsymbol{R}\downarrow}$. The first term is the on-site Hubbard repulsion; one abbreviates $U \equiv V(0)$.

A second class of matrix elements can be identified: those with $\boldsymbol{R}_1 = \boldsymbol{R}_3 \equiv \boldsymbol{R}$ and $\boldsymbol{R}_2 = \boldsymbol{R}_4 \equiv \boldsymbol{R}'$, with $\boldsymbol{R} \neq \boldsymbol{R}'$. These are the so-called *exchange* integrals:

$$\begin{aligned} J(\boldsymbol{R}-\boldsymbol{R}') &= \left\langle\, \boldsymbol{R}\,\boldsymbol{R}' \,\middle|\, \frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|} \,\middle|\, \boldsymbol{R}'\,\boldsymbol{R} \,\right\rangle \\ &= \int\! d^3r \!\int\! d^3r'\, \varphi^*(\boldsymbol{r}-\boldsymbol{R})\, \varphi^*(\boldsymbol{r}'-\boldsymbol{R}')\, \frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\, \varphi(\boldsymbol{r}'-\boldsymbol{R})\, \varphi(\boldsymbol{r}-\boldsymbol{R}') \\ &= \int\! d^3r \!\int\! d^3r'\, \varphi^*(\boldsymbol{r})\, \varphi(\boldsymbol{r}+\boldsymbol{R}-\boldsymbol{R}')\, \frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|}\, \varphi^*(\boldsymbol{r}'+\boldsymbol{R}-\boldsymbol{R}')\, \varphi(\boldsymbol{r}') \quad . \end{aligned} \tag{14.98}$$

Note that $J(\boldsymbol{R}-\boldsymbol{R}')$ is real. The exchange part of $\hat{V}$ is then

$$\begin{aligned} \hat{V}_{\rm exchange} &= -\tfrac{1}{2} \sum_{\substack{\boldsymbol{R}\neq\boldsymbol{R}' \\ \sigma\sigma'}} J(\boldsymbol{R}-\boldsymbol{R}')\, c^\dagger_{\boldsymbol{R}\sigma}\, c_{\boldsymbol{R}\sigma'}\, c^\dagger_{\boldsymbol{R}'\sigma'}\, c_{\boldsymbol{R}'\sigma} \\ &= -\tfrac{1}{4} \sum_{\boldsymbol{R}\neq\boldsymbol{R}'} J(\boldsymbol{R}-\boldsymbol{R}') \left(n_{\boldsymbol{R}}\, n_{\boldsymbol{R}'} + \boldsymbol{\sigma}_{\boldsymbol{R}} \cdot \boldsymbol{\sigma}_{\boldsymbol{R}'} \right) \quad . \end{aligned} \tag{14.99}$$

The $n_{\boldsymbol{R}}\, n_{\boldsymbol{R}'}$ piece can be lumped with the direct density-density interaction. What is new is the Heisenberg interaction,

$$\hat{V}_{\rm Heis} = -\sum_{\boldsymbol{R}\neq\boldsymbol{R}'} J(\boldsymbol{R}-\boldsymbol{R}')\, \boldsymbol{S}_{\boldsymbol{R}} \cdot \boldsymbol{S}_{\boldsymbol{R}'} \quad . \tag{14.100}$$

$J(\boldsymbol{R}-\boldsymbol{R}')$ is usually positive, and this gives us an explanation of Hund's first rule, which says to maximize S. This raises an interesting point, because we know that the ground state spatial wavefunction for the general two-body Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2m}\big(\nabla_1^2 + \nabla_2^2\big) + V\big(|\boldsymbol{r}_1 - \boldsymbol{r}_2|\big) \tag{14.101}$$

is nodeless. Thus, for fermions, the ground state spin wavefunction is an antisymmetric singlet state, corresponding to $S = 0$. Yet the V^{3+} ion, with electronic configuration [Ar] $3d^2$, has a triplet $S = 1$ ground state, according to Hund's first rule. Why don't the two 3d electrons have a singlet ground state, as the 'no nodes theorem' would seem to imply? The answer must have to do with the presence of the core electrons. Two electrons in the 1s shell do have a singlet ground state – indeed that is the only possibility. But the two 3d electrons in V^{3+} are not independent, but must be orthogonalized to the core states. This in effect projects out certain parts of the wavefunction, rendering the no nodes theorem inapplicable.

14.6.2 Heitler-London theory of the H_2 molecule

The Hamiltonian for the H_2 molecule is

$$\begin{aligned}\hat{H} = \frac{\boldsymbol{p}_1^2}{2m} - \frac{e^2}{|\boldsymbol{r}_1 - \boldsymbol{R}_{\rm a}|} + \frac{\boldsymbol{p}_2^2}{2m} - \frac{e^2}{|\boldsymbol{r}_2 - \boldsymbol{R}_{\rm b}|} \\ + \frac{e^2}{|\boldsymbol{R}_{\rm a} - \boldsymbol{R}_{\rm b}|} - \frac{e^2}{|\boldsymbol{r}_1 - \boldsymbol{R}_{\rm b}|} - \frac{e^2}{|\boldsymbol{r}_2 - \boldsymbol{R}_{\rm a}|} + \frac{e^2}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|} \quad .\end{aligned} \tag{14.102}$$

The total wavefunction is antisymmetric: $\Psi(\boldsymbol{r}_1\,\sigma_1\,,\,\boldsymbol{r}_2\sigma_{\,2}) = -\Psi(\boldsymbol{r}_2\sigma_2\,,\,\boldsymbol{r}_1\sigma_1)$. The $N = 2$ electron case is special because the wavefunction factorizes into a product:

$$\Psi(\boldsymbol{r}_1\,\sigma_1, \boldsymbol{r}_2\,\sigma_2) = \Phi(\boldsymbol{r}_1, \boldsymbol{r}_2)\,\chi(\sigma_1, \sigma_2) \quad . \tag{14.103}$$

The spin wavefunction may either be symmetric (triplet, $S = 1$), or antisymmetric (singlet, $S = 0$):

$$\big|\,\chi\,\big\rangle = \begin{cases} \big|\uparrow\uparrow\big\rangle & S = 1 \\ \frac{1}{\sqrt{2}}\Big(\big|\uparrow\downarrow\big\rangle + \big|\downarrow\uparrow\big\rangle\Big) & S = 1 \\ \big|\downarrow\downarrow\big\rangle & S = 1 \\ \frac{1}{\sqrt{2}}\Big(\big|\uparrow\downarrow\big\rangle - \big|\downarrow\uparrow\big\rangle\Big) & S = 0 \quad . \end{cases} \tag{14.104}$$

A symmetric spin wavefunction requires an antisymmetric spatial one, and *vice versa*.

Despite the fact that $\hat{H}$ does not explicitly depend on spin, the effective low-energy Hamiltonian for this system is

$$\hat{H}_{\rm eff} = K + J\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 \quad . \tag{14.105}$$

The singlet-triplet splitting is $\Delta E = E_{S=0} - E_{S=1} = -J$, so if $J > 0$ the ground state is the singlet, and if $J < 0$ the ground state is the three-fold degenerate triplet.

The one-electron 1s eigenfunction $\psi(\boldsymbol{r})$ satisfies the following eigenvalue equation:

$$\left\{-\frac{\hbar^2}{2m}\,\nabla^2 - \frac{e^2}{r}\right\}\psi(\boldsymbol{r}) = \varepsilon_0(\boldsymbol{r})\,\psi(\boldsymbol{r}) \quad . \tag{14.106}$$

In the Heitler-London approach, we write the two-electron wavefunction as a linear combination

$$\Phi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \alpha\,\Phi_{\rm I}(\boldsymbol{r}_1, \boldsymbol{r}_2) + \beta\,\Phi_{\rm II}(\boldsymbol{r}_1, \boldsymbol{r}_2) \quad , \tag{14.107}$$

with

$$\begin{aligned}\Phi_{\rm I}(\boldsymbol{r}_1, \boldsymbol{r}_2) &= \psi(\boldsymbol{r}_1 - \boldsymbol{R}_{\rm a})\,\psi(\boldsymbol{r}_2 - \boldsymbol{R}_{\rm b}) \equiv \psi_{\rm a}(\boldsymbol{r}_1)\,\psi_{\rm b}(\boldsymbol{r}_2) \\ \Phi_{\rm II}(\boldsymbol{r}_1, \boldsymbol{r}_2) &= \psi(\boldsymbol{r}_1 - \boldsymbol{R}_{\rm b})\,\psi(\boldsymbol{r}_2 - \boldsymbol{R}_{\rm a}) \equiv \psi_{\rm b}(\boldsymbol{r}_1)\,\psi_{\rm a}(\boldsymbol{r}_2) \quad .\end{aligned} \tag{14.108}$$

Assuming the atomic orbital $\psi(\boldsymbol{r})$ to be normalized, we define the following integrals:

$$\Delta = \int d^3r\, \psi_{\rm a}^*(\boldsymbol{r})\, \psi_{\rm b}(\boldsymbol{r}) \tag{14.109}$$

$$X = \int d^3r_1 \int d^3r_2 \left|\Phi_{\rm I}(\boldsymbol{r}_1,\boldsymbol{r}_2)\right|^2 \left(\frac{e^2}{R_{\rm ab}} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{1b}} - \frac{e^2}{r_{2a}}\right) \tag{14.110}$$

$$= \int d^3r_1 \int d^3r_2 \left|\Phi_{\rm II}(\boldsymbol{r}_1,\boldsymbol{r}_2)\right|^2 \left(\frac{e^2}{R_{\rm ab}} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{1a}} - \frac{e^2}{r_{2b}}\right)$$

$$Y = \int d^3r_1 \int d^3r_2\, \Phi_{\rm I}^*(\boldsymbol{r}_1,\boldsymbol{r}_2)\, \Phi_{\rm II}(\boldsymbol{r}_1,\boldsymbol{r}_2) \left(\frac{e^2}{R_{\rm ab}} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{1b}} - \frac{e^2}{r_{2a}}\right) \quad , \tag{14.111}$$

with $\boldsymbol{r}_{1a} = \boldsymbol{r}_1 - \boldsymbol{R}_{\rm a}$, *etc.* The expectation value of $\hat{H}$ in the state Φ is

$$\left\langle\, \Phi \,\middle|\, \hat{H} \,\middle|\, \Phi \,\right\rangle = (|\alpha|^2 + |\beta|^2)\,(2\varepsilon_0 + X) + (\alpha^*\beta + \beta^*\alpha)\,(2\varepsilon|\Delta|^2 + Y) \quad , \tag{14.112}$$

and the self-overlap is

$$\left\langle\, \Phi \,\middle|\, \Phi \,\right\rangle = |\alpha|^2 + |\beta|^2 + |\Delta|^2\,(\alpha^*\beta + \beta^*\alpha) \quad . \tag{14.113}$$

We now minimize $\langle \hat{H} \rangle$ subject to the condition that Φ be normalized, using a Lagrange multiplier E to impose the normalization. Extremizing with respect to α^* and β^* yields

$$\begin{pmatrix} 2\varepsilon_0 + X & 2\varepsilon_0|\Delta|^2 + Y \\ 2\varepsilon_0|\Delta|^2 + Y & 2\varepsilon_0 + X \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = E \begin{pmatrix} 1 & |\Delta|^2 \\ |\Delta|^2 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad , \tag{14.114}$$

and extremizing with respect to E yields the normalization condition

$$|\alpha|^2 + |\beta|^2 + |\Delta|^2\,(\alpha^*\beta + \beta^*\alpha) = 1 \quad . \tag{14.115}$$

The solutions are symmetric and antisymmetric states, with $\beta/\alpha = \pm 1$, corresponding to the energies

$$E_\pm = 2\varepsilon_0 + \frac{X \pm Y}{1 \pm |\Delta|^2} \quad . \tag{14.116}$$

Note that E_+ is the energy of the spatially symmetric state, which means a spin singlet while E_- corresponds to the spatially antisymmetric spin triplet.

The singlet-triplet splitting is

$$J = E_- - E_+ = 2\,\frac{Y - X|\Delta|^2}{1 - |\Delta|^4} \quad . \tag{14.117}$$

If $J > 0$, the triplet lies higher than the singlet, which says the ground state is antiferromagnetic. If $J < 0$, the triplet lies lower, and the ground state is ferromagnetic. The energy difference is largely determined by the Y integral:

$$Y = \int d^3r_1 \int d^3r_2\, \Upsilon^*(\boldsymbol{r}_1)\, \Upsilon(\boldsymbol{r}_2) \left(\frac{e^2}{R_{\rm ab}} + \frac{e^2}{r_{12}}\right) - 2\,\Delta^* \int d^3r\, \psi_{\rm a}^*(\boldsymbol{r}) \frac{e^2}{|\boldsymbol{r} - \boldsymbol{R}_{\rm b}|}\, \psi_{\rm b}(\boldsymbol{r}) \quad , \tag{14.118}$$

with $\Upsilon(\boldsymbol{r}) = \psi_{\rm a}^*(\boldsymbol{r})\,\psi_{\rm b}(\boldsymbol{r})$. The first term is positive definite for the Coulomb interaction. The second term competes with the first if the overlap is considerable. The moral of the story now emerges:

$$\begin{aligned} \text{weak overlap} &\Longrightarrow \text{ferromagnetism } (J<0) \\ \text{strong overlap} &\Longrightarrow \text{antiferromagnetism } (J>0) \quad . \end{aligned} \tag{14.119}$$

One finds that the H_2 molecule is indeed bound in the singlet state – the total energy has a minimum as a function of the separation $|\boldsymbol{R}_{\rm a} - \boldsymbol{R}_{\rm b}|$. In the triplet state, the molecule is unbound.

14.6.3 Failure of Heitler-London theory

At large separations $R \equiv |\boldsymbol{R}_a - \boldsymbol{R}_b|$ the Heitler-London method describes two H atoms with tiny overlap of the electronic wavefunctions. But this tiny overlap is what determines whether the ground state is a total spin singlet or triplet (we ignore coupling to the nuclear spin). Sugiura obtained the following expression for the singlet-triplet splitting in the $R \to \infty$ limit:

$$J(R) \simeq \left\{ \tfrac{56}{45} - \tfrac{4}{15}\gamma - \tfrac{4}{15}\ln\left(\frac{R}{a_{\rm H}}\right) \right\} \left(\frac{R}{a_{\rm H}}\right)^3 \left(\frac{e^2}{a_{\rm H}}\right) e^{-2R/a_{\rm H}} \quad , \tag{14.120}$$

where $\gamma = 0.577\ldots$ is the Euler constant and where $\psi(r) = (\pi a_{\rm H}^3)^{-1/2}\exp(-r/a_{\rm H})$ is the hydrogenic wavefunction. This is negative for sufficiently large separations ($R > 50\,a_{\rm H}$). But this is a problem, since the eigenvalue problem is a Sturm-Liouville problem, hence the lowest energy eigenfunction must be spatially symmetric – *the singlet state must always lie at lower energy than the triplet*. The problem here is that Heitler-London theory does a good job on the wavefunction where it is large, *i.e.* in the vicinity of the protons, but a lousy job in the overlap region.

14.6.4 Herring's approach

Conyers Herring was the first to elucidate the failure of Heitler-London theory at large separations. He also showed how to properly derive a Heisenberg model for a lattice of hydrogenic orbitals. Herring started with the symmetric spatial wavefunction

$$\Psi(\boldsymbol{r}_1\,,\,\ldots\,,\,\boldsymbol{r}_N) = \prod_{i=1}^{N} \psi(\boldsymbol{r}_i - \boldsymbol{R}_i) \quad . \tag{14.121}$$

This wavefunction would be appropriate were the electrons distinguishable. If we permute the electron coordinates using a spatial permutation $P_r \in \mathcal{S}_N$, we obtain another wavefunction of the same energy, E_0. However, there will be an overlap between these states:

$$J_P \equiv \big\langle\, \Psi \,\big|\, \hat{H} - E_0 \,\big|\, P_r\,\Psi \,\big\rangle \quad . \tag{14.122}$$

The effective Hamiltonian is then

$$\hat{H}_{\rm eff} = E_0 + \sum_{P\in\mathcal{S}_N} J_P\, P_r \quad . \tag{14.123}$$

A complete permutation P is a product of spatial and spin permutations: $P = P_r\, P_\sigma$, and the product when acting on an electronic wavefunction is $(-1)^P$, which is $+1$ for an even permutation and (-1) for an odd one[9]. Thus,

$$\hat{H}_{\rm eff} = E_0 + \sum_{P\in\mathcal{S}_N} (-1)^P\, J_P\, P_\sigma \quad . \tag{14.124}$$

The spin permutation operators P_σ may be written in terms of the Pauli spin matrices, once we note that the two-cycle (ij) may be written

$$P_{(ij)} = \tfrac{1}{2} + \tfrac{1}{2}\,\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_j \quad . \tag{14.125}$$

Thus, accounting for only two-cycles, we have

$$\hat{H}_{\rm eff} = E_0 - \tfrac{1}{4}\sum_{i\neq j} J_{ij}\left(1+\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_j\right) \quad . \tag{14.126}$$

For three-cycles, we have

$$\begin{aligned} P_{(ijk)} &= P_{(ik)}\,P_{(jk)} \\ &= \tfrac{1}{4}\big(1+\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_k\big)\big(1+\boldsymbol{\sigma}_j\cdot\boldsymbol{\sigma}_k\big) \\ &= \tfrac{1}{4}\Big[1+\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_j+\boldsymbol{\sigma}_j\cdot\boldsymbol{\sigma}_k+\boldsymbol{\sigma}_i\cdot\boldsymbol{\sigma}_k+i\boldsymbol{\sigma}_i\times\boldsymbol{\sigma}_j\cdot\boldsymbol{\sigma}_k\Big] \quad . \end{aligned} \tag{14.127}$$

14.7 Mean Field Theory

We begin with the Heisenberg Hamiltonian

$$\hat{H} = -\tfrac{1}{2}\sum_{i,j} J_{ij}\,\boldsymbol{S}_i\cdot\boldsymbol{S}_j - \gamma\sum_i \boldsymbol{H}_i\cdot\boldsymbol{S}_i \quad , \tag{14.128}$$

and write

$$\boldsymbol{S}_i = \boldsymbol{m}_i + \delta\boldsymbol{S}_i \quad , \tag{14.129}$$

where $\boldsymbol{m}_i = \langle\boldsymbol{S}_i\rangle$ is the thermodynamic average of $\boldsymbol{S}_i$. We therefore have

$$\begin{aligned} \boldsymbol{S}_i\cdot\boldsymbol{S}_j &= \boldsymbol{m}_i\cdot\boldsymbol{m}_j + \boldsymbol{m}_i\cdot\delta\boldsymbol{S}_j + \boldsymbol{m}_j\cdot\delta\boldsymbol{S}_i + \delta\boldsymbol{S}_i\cdot\delta\boldsymbol{S}_j \\ &= -\boldsymbol{m}_i\cdot\boldsymbol{m}_j + \boldsymbol{m}_i\cdot\boldsymbol{S}_j + \boldsymbol{m}_j\cdot\boldsymbol{S}_i + \delta\boldsymbol{S}_i\cdot\delta\boldsymbol{S}_j \quad . \end{aligned} \tag{14.130}$$

[9]Here, 'even' and 'odd' refer to the number of 2-cycles into which a given permutation is decomposed.

The last term is quadratic in the fluctuations, and as an approximation we ignore it. This results in the following *mean field Hamiltonian*,

$$\begin{aligned}\hat{H}^{\text{MF}} &= +\tfrac{1}{2}\sum_{i,j} J_{ij}\,\boldsymbol{m}_i\cdot\boldsymbol{m}_j - \sum_i\Big(\gamma\boldsymbol{H}_i + \sum_j J_{ij}\,\boldsymbol{m}_j\Big)\cdot\boldsymbol{S}_i \\ &= E_0 - \gamma\sum_i \boldsymbol{H}_i^{\text{eff}}\cdot\boldsymbol{S}_i \quad ,\end{aligned} \tag{14.131}$$

where

$$\begin{aligned}E_0 &= \tfrac{1}{2}\sum_{i,j} J_{ij}\,\boldsymbol{m}_i\cdot\boldsymbol{m}_j \\ \boldsymbol{H}_i^{\text{eff}} &= \boldsymbol{H}_i + \gamma^{-1}\sum_j J_{ij}\,\boldsymbol{m}_j \quad .\end{aligned} \tag{14.132}$$

Note how the *effective field* $\boldsymbol{H}_i^{\text{eff}}$ is a sum of the external field $\boldsymbol{H}_i$ and the *internal field* $\boldsymbol{H}_i^{\text{int}} = \gamma^{-1}\sum_j J_{ij}\,\boldsymbol{m}_j$. Self-consistency now requires that

$$\boldsymbol{m}_i = \frac{\mathsf{Tr}\,\boldsymbol{S}_i\exp\big(\gamma\boldsymbol{H}_i^{\text{eff}}\cdot\boldsymbol{S}_i/k_{\text{B}}T\big)}{\mathsf{Tr}\exp\big(\gamma\boldsymbol{H}_i^{\text{eff}}\cdot\boldsymbol{S}_i/k_{\text{B}}T\big)} \quad , \tag{14.133}$$

where Tr means to sum or integrate over all local degrees of freedom (for site i). The free energy is then

$$F\big(\{\boldsymbol{m}_i\}\big) = \tfrac{1}{2}\sum_{i,j} J_{ij}\,\boldsymbol{m}_i\cdot\boldsymbol{m}_j - k_{\text{B}}T\sum_i \ln\mathsf{Tr}\exp\big(\gamma\boldsymbol{H}_i^{\text{eff}}\cdot\boldsymbol{S}_i/k_{\text{B}}T\big) \quad . \tag{14.134}$$

For classical systems, there are several common models:

- Ising Model with $S = \pm 1$:

$$\begin{aligned}m_i &= \tanh(\gamma H_i^{\text{eff}}/k_{\text{B}}T) \\ &= \tanh\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\,m_j\Big) \quad .\end{aligned} \tag{14.135}$$

  The free energy is

$$F = \tfrac{1}{2}\sum_{i,j} J_{ij}\,m_i\,m_j - k_{\text{B}}T\sum_i \ln 2\cosh\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\,m_j\Big) \quad . \tag{14.136}$$

- Ising Model with $S = -1, 0, +1$:

$$m_i = \frac{2\sinh\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\,m_j\Big)}{1 + 2\cosh\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\,m_j\Big)} \tag{14.137}$$

and

$$F = \tfrac{1}{2}\sum_{i,j} J_{ij}\, m_i\, m_j - k_{\rm B}T \sum_i \ln\left\{1 + 2\cosh\left(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\right)\right\} \quad . \tag{14.138}$$

- XY model with $\boldsymbol{S}_i = (\cos\theta_i, \sin\theta_i)$, $\boldsymbol{H} = H\,\hat{\boldsymbol{x}}$

$$\begin{aligned} m_i = \big\langle \cos\theta_i \big\rangle &= \frac{\int\limits_0^{2\pi} d\theta_i\ \cos\theta_i\ \exp\left(\gamma H_i^{\rm eff}\cos\theta_i/k_{\rm B}T\right)}{\int\limits_0^{2\pi} d\theta_i\ \exp\left(\gamma H_i^{\rm eff}\cos\theta_i/k_{\rm B}T\right)} \\ &= \frac{I_1\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\Big)}{I_0\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\Big)} \quad , \end{aligned} \tag{14.139}$$

where $I_n(z)$ is a modified Bessel function. The free energy is

$$F = \tfrac{1}{2}\sum_{i,j} J_{ij}\, m_i\, m_j - k_{\rm B}T \sum_i \ln 2\pi I_0\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\Big) \quad . \tag{14.140}$$

- O(3) model with $\boldsymbol{S}_i = (\sin\theta_i\cos\phi_i, \sin\theta_i\sin\phi_i, \cos\theta_i)$. Suppose that $\boldsymbol{m}_i$ points in the direction of $\boldsymbol{H}_i^{\rm eff}$. Then

$$\begin{aligned} m_i = \big\langle \cos\theta_i \big\rangle &= \frac{2\pi\int\limits_0^{2\pi} d\theta_i\ \sin\theta_i\ \cos\theta_i\ \exp\left(\gamma H_i^{\rm eff}\cos\theta_i/k_{\rm B}T\right)}{2\pi\int\limits_0^{2\pi} d\theta_i\ \sin\theta_i\ \exp\left(\gamma H_i^{\rm eff}\cos\theta_i/k_{\rm B}T\right)} \\ &= {\rm ctnh}\big(\gamma H_i^{\rm eff}/k_{\rm B}T\big) - \frac{k_{\rm B}T}{\gamma H_i^{\rm eff}} \\ &= {\rm ctnh}\Big(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\Big) - \frac{k_{\rm B}T}{\gamma H_i + \sum_j J_{ij}\, m_j} \quad . \end{aligned} \tag{14.141}$$

The free energy is

$$F = \tfrac{1}{2}\sum_{i,j} J_{ij}\, m_i\, m_j - k_{\rm B}T \sum_i \ln\left\{\frac{4\pi\sinh\left(\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j\right)}{\beta\gamma H_i + \beta\sum_j J_{ij}\, m_j}\right\} \quad . \tag{14.142}$$

EXERCISE: Show that the self-consistency is equivalent to $\partial F/\partial m_i = 0$.

14.7.1 Ferromagnets

Ising model – Let us assume that the system orders ferromagnetically, with $m_i = m$ on all sites. Then, defining

$$\hat{J}(\boldsymbol{q}) = \sum_{\boldsymbol{R}} J(\boldsymbol{R})\, e^{-i\boldsymbol{q}\cdot\boldsymbol{R}} \quad , \tag{14.143}$$

we have that the free energy per site, $f = F/N$, is

$$f(m) = \tfrac{1}{2}\hat{J}(0)\,\boldsymbol{m}^2 - k_{\rm B}T \ln \mathsf{Tr} \exp\Big\{ \big(\gamma \boldsymbol{H} + \hat{J}(0)\,\boldsymbol{m}\big)\cdot \boldsymbol{S}/k_{\rm B}T \Big\} \quad . \tag{14.144}$$

For the $\mathbb{Z}_2$ (Ising) model, we have

$$m = \tanh\big(\beta\gamma H + \beta \hat{J}(0) m\big) \quad , \tag{14.145}$$

a transcendental equation for m. For $H = 0$, we find $m = \tanh(\hat{J}(0)m/k_{\rm B}T)$, which yields the Curie temperature $T_{\rm C} = \hat{J}(0)/k_{\rm B}$.

O(3) model – We have $\boldsymbol{m} = m\hat{\boldsymbol{H}}$ lies along $\boldsymbol{H}$. In the $\boldsymbol{H} \to 0$ limit, there is no preferred direction. The amplitude, however, satisfies

$$\frac{\partial f}{\partial m} = 0 \quad \Rightarrow \quad m = \,\mathrm{ctnh}\big(\hat{J}(0)\,m/k_{\rm B}T\big) - \frac{k_{\rm B}T}{\hat{J}(0)\,m} \quad . \tag{14.146}$$

With $x \equiv \hat{J}(0)\,m/k_{\rm B}T$, then,

$$\frac{k_{\rm B}T}{\hat{J}(0)}\,x = \,\mathrm{ctnh}\,x - \frac{1}{x} = \frac{x}{3} - \frac{x^3}{45} + \ldots \quad , \tag{14.147}$$

hence $T_{\rm c} = \hat{J}(0)/3k_{\rm B}$.

14.7.2 Antiferromagnets

If the lattice is bipartite, then we have *two* order parameters: $\boldsymbol{m}_{\rm A}$ and $\boldsymbol{m}_{\rm B}$. Suppose $J_{ij} = -J < 0$ if i and j are nearest neighbors, and zero otherwise. The effective fields on the A and B sublattices are given by

$$\boldsymbol{H}^{\rm eff}_{\rm A,B} \equiv \boldsymbol{H} - \gamma^{-1} z J \boldsymbol{m}_{\rm B,A} \quad , \tag{14.148}$$

Note that the internal field on the A sublattice is $-\gamma^{-1}zJ\boldsymbol{m}_{\rm B}$, while the internal field on the B sublattice is $-\gamma^{-1}zJ\boldsymbol{m}_{\rm A}$. For the spin-$S$ quantum Heisenberg model, where $S^z \in \{-S, \ldots, +S\}$, we have

$$\mathsf{Tr}\exp(\boldsymbol{\xi}\cdot\boldsymbol{S}) = \frac{\sinh\big(S+\frac{1}{2}\big)\xi}{\sinh\frac{1}{2}\xi} \quad , \tag{14.149}$$

hence, with $\boldsymbol{\xi} = \gamma \boldsymbol{H}^{\rm eff}_{\rm A,B}/k_{\rm B}T$, we have

$$\big\langle \boldsymbol{S} \big\rangle = \hat{\boldsymbol{\xi}}\, S\, B_S(S\xi) \tag{14.150}$$

where $B_S(x)$ is the Brillouin function,

$$B_S(x) = \Big(1 + \frac{1}{2S}\Big)\,\mathrm{ctnh}\,\Big(\big(1+\frac{1}{2S}\big)\,x\Big) - \frac{1}{2S}\,\mathrm{ctnh}\,\Big(\frac{x}{2S}\Big) \quad . \tag{14.151}$$

In order to best take advantage of the antiferromagnetic interaction and the external magnetic field, the ordered state is characterized by a *spin flop* in which $\boldsymbol{m}_{\rm A}$ and $\boldsymbol{m}_{\rm B}$ are, for weak fields, oriented in opposite directions in a plane perpendicular to $\boldsymbol{H}$, but each with a small component along $\boldsymbol{H}$.

When $H = 0$, the mean field equations take the form

$$\begin{aligned} m_{\rm A} &= S B_S\big(zJSm_{\rm B}/k_{\rm B}T\big) \\ m_{\rm B} &= S B_S\big(zJSm_{\rm A}/k_{\rm B}T\big) \end{aligned} \quad , \tag{14.152}$$

where we have assumed $\boldsymbol{m}_{\rm A}$ and $\boldsymbol{m}_{\rm B}$ are antiparallel, with $\boldsymbol{m}_{\rm A} = m_{\rm A}\,\hat{\boldsymbol{n}}$ and $\boldsymbol{m}_{\rm B} = -m_{\rm B}\,\hat{\boldsymbol{n}}$, where $\hat{\boldsymbol{n}}$ is a unit vector. From the expansion of the Brillouin function, we obtain the Néel temperature $T_{\rm N} = S(S+1)zJ/3k_{\rm B}$.

14.7.3 Susceptibility

For $T > T_{\rm c}$ the system is paramagnetic, and there is a linear response to an external field,

$$\begin{aligned} \chi_{ij}^{\mu\nu} &= \frac{\partial M_i^\mu}{\partial H_j^\nu} = \gamma\,\frac{\partial m_i^\mu}{\partial H_j^\nu} = -\frac{\partial^2 F}{\partial H_i^\mu\,\partial H_j^\nu} \\ &= \frac{\gamma^2}{k_{\rm B}T}\Big\{\big\langle S_i^\mu\, S_j^\nu\big\rangle - \big\langle S_i^\mu\big\rangle\,\big\langle S_j^\nu\big\rangle\Big\} \end{aligned} \tag{14.153}$$

where $\{i,j\}$ are site indices and $\{\mu,\nu\}$ are internal spin indices. The mean field Hamiltonian is, up to a constant,

$$\hat{H}^{\rm MF} = -\gamma \sum_i \boldsymbol{H}_i^{\rm eff}\cdot\boldsymbol{S}_i \quad , \tag{14.154}$$

which is a sum of single site terms. Hence, the response within $\hat{H}^{\rm MF}$ must be purely local as well as isotropic. That is, for weak effective fields, using $\boldsymbol{M}_i = \gamma\,\boldsymbol{m}_i$,

$$\boldsymbol{M}_i = \chi_0\,\boldsymbol{H}_i^{\rm eff} = \chi_0\,\boldsymbol{H}_i + \gamma^{-2}\,\chi_0\,J_{ij}\,\boldsymbol{M}_j \quad , \tag{14.155}$$

which is equivalent to

$$\big(\delta_{ij} - \gamma^{-2}\chi_0\,J_{ij}\big)\,\boldsymbol{M}_j = \chi_0\,\boldsymbol{H}_i \quad , \tag{14.156}$$

and the mean field susceptibility is

$$\chi_{ij}^{\mu\nu} = \left[\chi_0^{-1} - \gamma^{-2} J\right]_{ij}^{-1} \delta^{\mu\nu} \quad . \tag{14.157}$$

It is convenient to work in Fourier space, in which case the matrix structure is avoided and one has

$$\hat{\chi}(\boldsymbol{q}) = \frac{\chi_0}{1 - \gamma^{-2}\chi_0\, \hat{J}(\boldsymbol{q})} \quad . \tag{14.158}$$

The local susceptibility χ_0 is readily determined:

$$\begin{aligned} M^\mu = \gamma \left\langle S^\mu \right\rangle &= \gamma \, \frac{\mathsf{Tr}\, \boldsymbol{S}\, \exp(\gamma \boldsymbol{H} \cdot \boldsymbol{S}/k_{\rm B}T)}{\mathsf{Tr}\, \exp(\gamma \boldsymbol{H} \cdot \boldsymbol{S}/k_{\rm B}T)} \\ &= \gamma S B_S(S\gamma H/k_{\rm B}T)\, \hat{H}^\mu \quad , \end{aligned} \tag{14.159}$$

where $B_S(x)$ is the Brillouin function from eqn. 14.151. As $\boldsymbol{H} \to 0$ we have $\boldsymbol{M} = \chi_0 \boldsymbol{H}$, with

$$\chi_0^{\mu\nu} = \frac{\gamma^2}{k_{\rm B}T} \cdot \frac{\mathsf{Tr}(S^\mu S^\nu)}{\mathsf{Tr}\, 1} \equiv \chi_0\, \delta^{\mu\nu} \quad , \tag{14.160}$$

where $\chi_0 = \frac{1}{N}\, \mathsf{Tr}(\boldsymbol{S}^2)/\, \mathsf{Tr}\, 1$, where N is the number of components of S^μ. Thus, for the Ising model ($N = 1$) we have $\chi_0^{\rm Ising} = \gamma^2/k_{\rm B}T$, while for the spin-$S$ quantum Heisenberg model we have $\chi_0^{\rm Heis} = S(S+1)\gamma^2/3k_{\rm B}T$. Note that $\chi_0 \propto T^{-1}$; the splitting of the degenerate energy levels by the magnetic field is of little consequence at high temperatures.

In many cases one deals with 'single ion anisotropy' terms. For example, one can add to the Heisenberg Hamiltonian a term such as

$$\hat{H}_{\rm a} = D \sum_i \left(S_i^z\right)^2 \quad , \tag{14.161}$$

which for $D < 0$ results in an *easy axis* anisotropy (*i.e.* the spins prefer to align along the $\hat{\boldsymbol{z}}$-axis), and for $D > 0$ results in an *easy plane* anisotropy (*i.e.* the spins prefer to lie in the (x, y) plane). Since this term is already the sum of single site Hamiltonians, there is no need to subject it to a mean field treatment. One then obtains the mean field Hamiltonian

$$\hat{H}^{\rm MF} = D \sum_i \left(S_i^z\right)^2 - \gamma \sum_i \boldsymbol{H}_i^{\rm eff} \cdot \boldsymbol{S}_i \quad . \tag{14.162}$$

In this case, χ_0 is now *anisotropic* in spin space. The general formula for χ_0 is

$$\chi_0^{\mu\nu} = \frac{\gamma^2}{k_{\rm B}T} \left\langle S^\mu S^\nu \right\rangle \tag{14.163}$$

where the thermodynamic average is taken with respect to the single site Hamiltonian.[10] One then has

$$\hat{\chi}^{\mu\nu}(\boldsymbol{q}) = \chi_0^{\mu\lambda} \left[\mathbb{I} - \gamma^{-2} \hat{J}(\boldsymbol{q}) \overset{\leftrightarrow}{\chi}_0\right]_{\lambda\nu}^{-1} \quad , \tag{14.164}$$

where the matrix inverse is now in internal spin space.

[10] Note that in (14.160) the single site Hamiltonian is simply $\hat{H}_0 = 0$.

14.7.4 Variational probability distribution

Here's another way to derive mean field theory. Let Ω represent a configuration and let P_Ω be a probability distribution, normalized such that $\sum_\Omega P_\Omega = 1$. We define the entropy of the distribution as

$$S[P] = -k_{\rm B} \sum_\Omega P_\Omega \, \ln P_\Omega \quad . \tag{14.165}$$

We now ask what distribution P_Ω minimizes the free energy $F = \langle \hat{H} \rangle - TS$. Working in an eigenbasis of $\hat{H}$, we have

$$F = \sum_\Omega P_\Omega \, E_\Omega + k_{\rm B} T \sum_\Omega P_\Omega \, \ln P_\Omega \quad . \tag{14.166}$$

We extremize F subject to the normalization constraint, which is implemented with a Lagrange multiplier λ. This means we form the extended function

$$F^*\big(\{P_\Omega\}, \lambda\big) = \sum_\Omega P_\Omega \, E_\Omega + k_{\rm B} T \sum_\Omega P_\Omega \, \ln P_\Omega - \lambda \Big(\sum_\Omega P_\Omega - 1 \Big) \quad , \tag{14.167}$$

and demand $dF^*/dP_\Omega = 0$ for all Ω as well as $dF^*/d\lambda = 0$. This results in the Boltzmann distribution,

$$P_\Omega^{\rm eq} = \frac{1}{Z} \, e^{-E_\Omega / k_{\rm B} T} \qquad , \qquad Z = \sum_l e^{-E_l / k_{\rm B} T} \quad . \tag{14.168}$$

Thus, *any distribution other than $P_\Omega^{\rm eq}$ results in a larger free energy*.

Mean field theory may be formulated in terms of a variational probability distribution. Thus, rather than working with the Boltzmann distribution $P_\Omega^{\rm eq}$, which is usually intractable, we invoke a trial distribution $P_\Omega(x_1, x_2, \ldots)$, parameterized by $\{x_1, x_2, \ldots\}$, and minimize the resultant $F = \langle \hat{H} \rangle - TS$ with respect to those parameters.

As an example, consider the Ising model with spins $\sigma_i = \pm 1$. Each configuration is given by the set of spin polarizations: $\Omega = \{\sigma_1, \ldots, \sigma_N\}$. The full equilibrium probability distribution,

$$P_\Omega^{\rm eq} = Z^{-1} \, \exp\Big(\beta J \sum_{\langle ij \rangle} \sigma_i \sigma_j \Big) \quad , \tag{14.169}$$

with $\beta = 1/k_{\rm B} T$, is too cumbersome to work with. We replace this with a variational single-site distribution,

$$\begin{aligned} P_\Omega &= \prod_{j=1}^N P_i(\sigma_i) \\ P_i(\sigma_i) &= \tfrac{1}{2}(1 + m_i) \, \delta_{\sigma_i, +1} + \tfrac{1}{2}(1 - m_i) \, \delta_{\sigma_i, -1} \quad . \end{aligned} \tag{14.170}$$

The variational parameters are $\{m_1, \ldots, m_N\}$. Note that P_Ω is properly normalized, by construction.

The entropy of our trial distribution is decomposed into a sum over single site terms:

$$\begin{aligned} S[P] &= \sum_i s(m_i) \\ s(m) &= -k_{\rm B}\left\{ \frac{1+m}{2}\,\ln\Big(\frac{1+m}{2}\Big) + \frac{1-m}{2}\,\ln\Big(\frac{1-m}{2}\Big)\right\} \end{aligned} \tag{14.171}$$

The thermodynamic average $\langle\sigma_i\rangle$ is simply

$$\langle\sigma_i\rangle = {\rm Tr}\, P_i(\sigma_i)\,\sigma_i = m_i \quad , \tag{14.172}$$

hence from

$$\hat{H} = -\tfrac{1}{2}\sum_{i,j} J_{ij}\,\sigma_i\,\sigma_j - \gamma\sum_i H_i\,\sigma_i \quad , \tag{14.173}$$

we derive the free energy

$$\begin{aligned} F\big(\{m_i\}\big) &= -\tfrac{1}{2}\sum_{i,j} J_{ij}\,m_i\,m_j - \gamma\sum_i H_i\,m_i \\ &\quad + k_{\rm B}T\sum_i\left\{ \frac{1+m_i}{2}\,\ln\Big(\frac{1+m_i}{2}\Big) + \frac{1-m_i}{2}\,\ln\Big(\frac{1-m_i}{2}\Big)\right\} \end{aligned} \tag{14.174}$$

Varying with respect to each m_i, we obtain the coupled nonlinear mean field equations,

$$m_i = \tanh\Big[\Big(\sum_j J_{ij}\,m_j + \gamma H_i\Big)/k_{\rm B}T\Big] \quad . \tag{14.175}$$

For uniform magnetization ($m_i = m\;\forall\, i$), the free energy per site is

$$\begin{aligned} \frac{F}{N} &= -\tfrac{1}{2}\hat{J}(0)\,m^2 - \gamma H m + k_{\rm B}T\left\{ \frac{1+m}{2}\,\ln\Big(\frac{1+m}{2}\Big) + \frac{1-m}{2}\,\ln\Big(\frac{1-m}{2}\Big)\right\} \\ &= \tfrac{1}{2}\big(k_{\rm B}T - \hat{J}(0)\big)\,m^2 - \gamma H m + \tfrac{1}{12}\,k_{\rm B}T\,m^4 + \tfrac{1}{30}k_{\rm B}T m^6 + \ldots \end{aligned} \tag{14.176}$$

To compute the correlations, we may use the expression

$$\chi_{ij}(T) = \frac{\gamma^2}{k_{\rm B}T}\Big\{\langle\sigma_i\,\sigma_j\rangle - \langle\sigma_i\rangle\,\langle\sigma_j\rangle\Big\} \tag{14.177}$$

$$= \frac{\partial M_i}{\partial H_j} = \gamma\,\frac{\partial m_i}{\partial H_j} = -\frac{\partial^2 F}{\partial H_i\,\partial H_j} \quad . \tag{14.178}$$

Thus, there are two ways to compute the susceptibility. One is to evaluate the spin-spin *correlation function*, as in (14.177). The other is to differentiate the magnetization to obtain the *response function*, as in (14.178). The equality between the two – called the "fluctuation-dissipation theorem" – is in fact only valid for the equilibrium Boltzmann distribution $P^{\rm eq}_\Omega$. Which side of the equation should we use in our variational mean field theory? It is more accurate to use the response function. To roughly see this, let us write $P = P^{\rm eq} + \delta P$, with δP small in some sense. The free energy is given by

$$F[P] = F[P^{\rm eq}] + \delta P \cdot \frac{\delta F}{\delta P}\bigg|_{P=P^{\rm eq}} + \mathcal{O}\big((\delta P)^2\big) \quad . \tag{14.179}$$

Our variational treatment guarantees that the second term vanishes, since we extremize F with respect to P. Thus, in some sense, the error in F is only of order $(\delta P)^2$. If we compute the correlation function using $\langle A\rangle = {\rm Tr}\,(P\,A)$, where A is any operator, then the error will be linear in δP. So it is better to use the response function than the correlation function.

EXERCISE: Articulate the correspondence between this variational version of mean field theory and the 'neglect of fluctuations' approach we derived earlier.

14.8 Magnetic Ordering

The $\boldsymbol{q}$-dependent susceptibility in (14.158) diverges when $\gamma^{-2}\chi_0\hat{J}(\boldsymbol{q}) = 1$. As we know, such a divergence heralds the onset of a *phase transition* where there is a *spontaneous magnetization* in the ordered (*i.e.* low temperature) phase. Typically this happens at a particular wavevector $\boldsymbol{Q}$, or a set of symmetry related wavevectors $\{\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots\}$. The ordering wavevector is that value of $\boldsymbol{q}$ which results in a *maximum* of $\hat{J}(\boldsymbol{q})$: $\max_{\boldsymbol{q}}\big\{\hat{J}(\boldsymbol{q})\big\} = \hat{J}(\boldsymbol{Q})$. The susceptibility, for isotropic systems, can be written

$$\hat{\chi}(\boldsymbol{q}) = \frac{\chi_0}{\big[1 - \gamma^{-2}\chi_0\,\hat{J}(\boldsymbol{Q})\big] + \gamma^{-2}\chi_0\big[\hat{J}(\boldsymbol{Q}) - \hat{J}(\boldsymbol{q})\big]} \quad . \tag{14.180}$$

The critical temperature $T_{\rm c}$ is determined by the relation

$$\gamma^{-2}\chi_0(T_{\rm c})\,\hat{J}(\boldsymbol{Q}) = 1. \tag{14.181}$$

Expanding about $T = T_{\rm c}$, and about $\boldsymbol{q} = \boldsymbol{Q}$, where

$$\hat{J}(\boldsymbol{q}) = \hat{J}(\boldsymbol{Q})\Big\{1 - (\boldsymbol{q}-\boldsymbol{Q})^2 R_*^2 + \ldots\Big\} \quad , \tag{14.182}$$

we have

$$\hat{\chi}(\boldsymbol{q}) \approx \frac{\chi_0/R_*^2}{\xi^{-2}(T) + (\boldsymbol{q}-\boldsymbol{Q})^2} \quad , \tag{14.183}$$

where

$$\xi^{-2}(T) = -\frac{\chi_0'(T_{\rm c})}{\chi_0(T_{\rm c})} \cdot R_*^{-2} \cdot (T - T_{\rm c}) \quad . \tag{14.184}$$

Thus, $\xi(T) \propto (T - T_{\rm c})^{-1/2}$. The real space susceptibility $\chi(\boldsymbol{R}_i - \boldsymbol{R}_j)$ oscillates with wavevector $\boldsymbol{Q}$ and decays on the scale of the correlation length $\xi(T)$.

- *Ferromagnet:* $J_{ij} = +J > 0$ if i and j are nearest neighbors; otherwise $J_{ij} = 0$. On a hypercubic lattice (d dimensions, $2d$ nearest neighbors), we then have

$$\hat{J}(\boldsymbol{q}) = J \sum_{\boldsymbol{\delta}} e^{-i\boldsymbol{q}\cdot\boldsymbol{\delta}} = 2J\big\{\cos(q_1 a) + \cos(q_2 a) + \ldots + \cos(q_d a)\big\} \quad . \tag{14.185}$$

  The ordering wavevector is $\boldsymbol{Q} = 0$, and $\hat{J}(\boldsymbol{Q}) = 2dJ$. For the spin-$S$ Heisenberg model, then, $T_{\rm C} = \frac{2}{3} d\, S(S+1)\, J/k_{\rm B}$, and the susceptibility is

$$\hat{\chi}(\boldsymbol{q}) = \frac{\frac{1}{3}\gamma^2 S(S+1)/k_{\rm B}}{(T - T_{\rm C}) + T_{\rm C}\, d^{-1} \sum_{\nu=1}^{d} \big[1 - \cos(q_\nu a)\big]} \quad . \tag{14.186}$$

  The uniform susceptibility $\chi = \hat{\chi}(\boldsymbol{q} = 0)$ is then

$$\chi(T) = \frac{\gamma^2 S(S+1)}{3k_{\rm B}(T - T_{\rm C})} \quad . \tag{14.187}$$

  Ferromagnetic insulators: ferrites, EuO, TDAE-C_{60}.

- *Antiferromagnet:* $J_{ij} = -J < 0$ if i and j are nearest neighbors; otherwise $J_{ij} = 0$. On a hypercubic lattice (d dimensions, $2d$ nearest neighbors), we then have

$$\hat{J}(\boldsymbol{q}) = -J \sum_{\boldsymbol{\delta}} e^{-i\boldsymbol{q}\cdot\boldsymbol{\delta}} = -2J\big\{\cos(q_1 a) + \cos(q_2 a) + \ldots + \cos(q_d a)\big\} \quad . \tag{14.188}$$

  The ordering wavevector is $\boldsymbol{Q} = (\pi/a, \ldots, \pi/a)$, at the zone corner, where $\hat{J}(\boldsymbol{Q}) = 2dJ$. For the spin-$S$ Heisenberg model, then, $T_{\rm N} = \frac{2}{3} d\, S(S+1)\, J/k_{\rm B}$, and the susceptibility is

$$\hat{\chi}(\boldsymbol{q}) = \frac{\gamma^2 S(S+1)/3k_{\rm B}}{(T - T_{\rm N}) + T_{\rm N}\, d^{-1} \sum_{\nu=1}^{d} \big[1 + \cos(q_\nu a)\big]} \quad . \tag{14.189}$$

  The uniform susceptibility $\chi = \hat{\chi}(\boldsymbol{q} = 0)$ is then

$$\chi(T) = \frac{\gamma^2 S(S+1)}{3k_{\rm B}(T + T_{\rm N})} \quad , \tag{14.190}$$

  which *does not diverge*. Indeed, plotting $\chi^{-1}(T)$ *versus* T, one obtains an intercept along the T-axis at $T = -T_{\rm N}$. This is one crude way of estimating the Néel temperature. What *does* diverge is the *staggered susceptibility* $\chi_{\rm stag} \equiv \hat{\chi}(\boldsymbol{Q}, T)$, *i.e.* the susceptibility at the ordering wavevector:

$$\chi_{\rm stag}(T) = \frac{\gamma^2 S(S+1)}{3k_{\rm B}(T - T_{\rm N})} \quad . \tag{14.191}$$

- *Frustrated Antiferromagnet:* On the triangular lattice, the antiferromagnetic state is frustrated. What does mean field theory predict? We begin by writing primitive direct lattice vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$ and primitive reciprocal lattice vectors $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$:

$$\begin{aligned} \boldsymbol{a}_1 &= a\,(1,0) & \boldsymbol{b}_1 &= \frac{4\pi}{a\sqrt{3}}\left(\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}\right) \\ \boldsymbol{a}_2 &= a\left(\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}\right) & \boldsymbol{b}_2 &= \frac{4\pi}{a\sqrt{3}}\,(0,1) \quad , \end{aligned} \tag{14.192}$$

where a is the lattice constant. The six nearest neighbor vectors are then

$$\delta \in \Big\{ \boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_2 - \boldsymbol{a}_1, -\boldsymbol{a}_1, -\boldsymbol{a}_2, \boldsymbol{a}_1 - \boldsymbol{a}_2 \Big\} \quad , \tag{14.193}$$

and writing $\boldsymbol{q} \equiv x_1 \boldsymbol{b}_1 + x_2 \boldsymbol{b}_2$, we find

$$\hat{J}(\boldsymbol{q}) = -2J\big\{\cos(2\pi x_1) + \cos(2\pi x_2) + \cos(2\pi x_1 - 2\pi x_2)\big\} \quad . \tag{14.194}$$

We suspect that this should be maximized somewhere along the perimeter of the Brillouin zone. The face center lies at $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$, where $\hat{J}(\boldsymbol{q}) = +2J$. However, an even greater value is obtained either of the two inequivalent zone corners, $(x_1, x_2) = (\frac{2}{3}, \frac{1}{3})$ and $(x_1, x_2) = (\frac{1}{3}, \frac{2}{3})$, where $\hat{J}(\boldsymbol{q}) = +3J$. Each of these corresponds to a *tripartite* division of the triangular lattice in to three $\sqrt{3} \times \sqrt{3}$ triangular sublattices.

Antiferromagnetic insulators: MnO, CoO, FeO, NiO, La_2CuO_4.

- *Helimagnet:* Consider a cubic lattice system with mixed ferromagnetic and antiferromagnetic interactions:

$$J_{ij} = \begin{cases} +J_1 > 0 & \text{6 nearest neighbors} \\ -J_2 < 0 & \text{12 next-nearest neighbors} \\ 0 & \text{otherwise} \quad . \end{cases} \tag{14.195}$$

Then

$$\begin{aligned} \hat{J}(\boldsymbol{q}) = {} & 2J_1\big[\cos(q_x a) + \cos(q_y a) + \cos(q_z a)\big] \\ & - 4J_2\big[\cos(q_x a)\cos(q_y a) + \cos(q_x a)\cos(q_z a) + \cos(q_y a)\cos(q_z a)\big] \quad . \end{aligned} \tag{14.196}$$

The ordering wavevector is then

$$\boldsymbol{Q} = \begin{cases} a^{-1}\cos^{-1}\left(\frac{J_1}{4J_2}\right)(\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}} + \hat{\boldsymbol{z}}) & \text{if } J_1 < 4J_2 \\ 0 & \text{if } J_1 \geqslant 4J_2 \quad . \end{cases} \tag{14.197}$$

Thus, for $J_1 < 4J_2$ the order is *incommensurate* with the lattice. The maximum value of $\hat{J}(\boldsymbol{q})$ is

$$\hat{J}(\boldsymbol{Q}) = \begin{cases} \frac{3J_1^2}{4J_2} & \text{if } J_1 < 4J_2 \\ 6(J_1 - 2J_2) & \text{if } J_1 \geqslant 4J_2 \quad , \end{cases} \tag{14.198}$$

hence incommensurate order sets in at $T_{\rm I} = S(S+1)J_1^2/4k_{\rm B}J_2$. The uniform susceptibility is

$$\hat{\chi}(0) = \frac{\gamma^2 S(S+1)/3k_{\rm B}}{T - 8\,T_{\rm I}\,\frac{J_2}{J_1}\Big(1 - \frac{2J_2}{J_1}\Big)} \quad . \tag{14.199}$$

Thus,

$$\chi(T) \simeq \begin{cases} \frac{\mathcal{C}}{T+T^*} & 0 < J_1 < 2J_2 \text{ (like AFM)} \\ \\ \frac{\mathcal{C}}{T-T^*} & 2J_2 < J_1 < 4J_2 \text{ (like FM)} \quad . \end{cases} \tag{14.200}$$

14.8.1 Mean field theory of anisotropic magnetic systems

Consider the anisotropic Heisenberg model,

$$\hat{H} = -\overbrace{\sum_{i<j} J_{ij}^{\parallel}\,\boldsymbol{S}_i \cdot \boldsymbol{S}_j}^{\textit{intra}} - \overbrace{\sum_{i<j} J_{ij}^{\perp}\,\boldsymbol{S}_i \cdot \boldsymbol{S}_j}^{\textit{inter}} - \gamma \sum_i \boldsymbol{H}_i \cdot \boldsymbol{S}_i \quad . \tag{14.201}$$

Here, $J_{ij}^{\parallel}$ only connects sites within the same plane (quasi-2d) or chain (quasi-1d), while $J_{ij}^{\perp}$ only connects sites in different planes/chains. We assume that we have an adequate theory for isolated plains/chains, and we effect a mean field decomposition on the interplane/interchain term:

$$\boldsymbol{S}_i \cdot \boldsymbol{S}_j = -\langle \boldsymbol{S}_i \rangle \cdot \langle \boldsymbol{S}_j \rangle + \langle \boldsymbol{S}_i \rangle \cdot \boldsymbol{S}_j + \langle \boldsymbol{S}_j \rangle \cdot \boldsymbol{S}_i + \overbrace{\delta\boldsymbol{S}_i \cdot \delta\boldsymbol{S}_j}^{\text{(fluct)}^2} \quad , \tag{14.202}$$

resulting in the effective field

$$\boldsymbol{H}^{\rm eff}(\boldsymbol{q},\omega) = \boldsymbol{H}(\boldsymbol{q},\omega) + \gamma^{-2}\hat{J}^{\perp}(\boldsymbol{q}_\perp)\,\boldsymbol{M}(\boldsymbol{q},\omega) \quad , \tag{14.203}$$

where $\boldsymbol{M}(\boldsymbol{q},\omega) = \gamma\langle \boldsymbol{S}(\boldsymbol{q},\omega)\rangle$. Thus,

$$\hat{\chi}(\boldsymbol{q},\omega) = \frac{\hat{\chi}^{\parallel}(\boldsymbol{q}_\parallel,\omega)}{1 - \gamma^{-2}\hat{J}^{\perp}(\boldsymbol{q}_\perp)\,\hat{\chi}^{\parallel}(\boldsymbol{q}_\parallel,\omega)} \quad , \tag{14.204}$$

where $\hat{\chi}^{\parallel}(\boldsymbol{q}_\parallel,\omega)$ is assumed known.

14.8.2 Quasi-1D chains

Consider a ferromagnet on a cubic lattice where the exchange interaction along the $\hat{\boldsymbol{z}}$-direction ($\parallel$) is much larger than that in the (x,y) plane ($\perp$). Treating the in-plane interactions via mean

field theory, we have

$$\hat{\chi}(\boldsymbol{q}_\perp, q_z) = \frac{\hat{\chi}_{\rm 1D}(q_z)}{1 - \gamma^{-2}\hat{J}^\perp(\boldsymbol{q}_\perp)\,\hat{\chi}_{\rm 1D}(q_z)} \quad , \tag{14.205}$$

with

$$\hat{J}^\perp(\boldsymbol{q}_\perp) = 2J_\perp\Big\{\cos(q_x a) + \cos(q_y a)\Big\} \quad . \tag{14.206}$$

For the Ising model we can compute $\hat{\chi}_{\rm 1D}(q_z)$ exactly using the high temperature expansion:

$$\begin{aligned}\langle\sigma_n\,\sigma_{n'}\rangle &= \frac{{\rm Tr}\left\{\sigma_n\,\sigma_{n'}\,\prod_j\left(1+\tanh(J_\parallel/k_{\rm B}T)\,\sigma_j\,\sigma_{j+1}\right)\right\}}{{\rm Tr}\prod_j\left(1+\tanh(J_\parallel/k_{\rm B}T)\,\sigma_j\,\sigma_{j+1}\right)} \\ &= \tanh^{|n-n'|}(J_\parallel/k_{\rm B}T) \quad .\end{aligned} \tag{14.207}$$

Thus,

$$\begin{aligned}\hat{\chi}_{\rm 1D}(q_z) &= \frac{\gamma^2}{k_{\rm B}T}\sum_{n=-\infty}^{\infty}\tanh^{|n|}(J_\parallel/k_{\rm B}T)\,e^{inq_zc} \\ &= \frac{\gamma^2}{k_{\rm B}T}\,\frac{1}{\cosh(2J_\parallel/k_{\rm B}T) - \sinh(2J_\parallel/k_{\rm B}T)\,\cos(q_zc)} \\ &\approx \frac{2\pi\gamma^2}{ck_{\rm B}T}\cdot\frac{1}{\pi}\,\frac{\xi^{-1}}{\xi^{-2}+q_z^2} \quad ,\end{aligned} \tag{14.208}$$

where c is the lattice spacing along the chains, and where the last approximation is valid for $q \to 0$ and $\xi \to \infty$. The correlation length in this limit is given by

$$\xi(T) \simeq \frac{c}{2}\,\exp(2J_\parallel/k_{\rm B}T) \quad . \tag{14.209}$$

Note that $\xi(T)$ diverges only at $T = 0$. This is consistent with the well-known fact that the *lower critical dimension* for systems with *discrete global symmetries* and *short-ranged interactions* is $d = 1$. That is to say that there is no spontaneous breaking of any discrete symmetry in one-dimension (with the proviso of sufficiently short-ranged interactions). For *continuous symmetries* the lower critical dimension is $d = 2$, which is the content of the Hohenberg-Mermin-Wagner (HMW) theorem.

Accounting for the residual interchain interactions via mean field theory, we obtain the anisotropic (in space) susceptibility

$$\hat{\chi}(\boldsymbol{q}_\perp, q_z) = \frac{\hat{\chi}_{\rm 1D}(q_z)}{1 - \gamma^{-2}\cdot 2J_\perp\big\{\cos(q_x a) + \cos(q_y a)\big\}\cdot\hat{\chi}_{\rm 1D}(q_z)} \quad . \tag{14.210}$$

Three-dimensional ordering at $\boldsymbol{Q} = 0$ sets in at $T = T_{\rm c}$, which occurs when $\hat{\chi}(\boldsymbol{Q})$ has a pole. The equation for this pole is

$$4\gamma^{-2}J_\perp\,\chi_{\rm 1D} = 1 \qquad \Rightarrow \qquad \frac{4J_\perp}{k_{\rm B}T_{\rm c}} = \exp(-2J_\parallel/k_{\rm B}T_{\rm c}) \quad . \tag{14.211}$$

This transcendental equation is equivalent to

$$x\, e^x = \frac{1}{\epsilon} \tag{14.212}$$

where $x = 2J_\parallel/k_{\rm B}T_{\rm c}$ and $\epsilon = 2J_\perp/J_\parallel$. The solution, for small ϵ, is

$$k_{\rm B}T_{\rm c} = \frac{2J_\parallel}{\ln\left(J_\parallel/2J_\perp\right)} + \ldots \quad . \tag{14.213}$$

Thus, $T_{\rm c} > 0$ for all finite $J_\perp$, with $T_{\rm c}$ going to zero rather slowly as $J_\perp \to 0$.

Similar physics is present in the antiferromagnetic insulator phase of the cuprate superconductors. The antiferromagnetic (staggered) susceptibility of the two-dimensional Heisenberg model diverges as $T \to 0$ as $\chi^{\rm stag}_{\rm 2D} \sim J^{-1}\exp(\rho J/k_{\rm B}T)$, where ρ is a dimensionless measure of quantum fluctuations. As in the $d = 1$ Ising case, there is no phase transition at any finite temperature, in this case owing to the HMW theorem. However, when the quasi-2D layers are weakly coupled with antiferromagnetic coupling J' (the base structure is a cubic perovskite), three-dimensional Néel ordering sets in at the antiferromagnetic wavevector $\boldsymbol{Q} = (\pi/a, \pi/a, \pi/c)$ at a critical temperature $T_{\rm N} \approx J/k_{\rm B}\ln(J/J')$.

14.9 Spin Wave Theory

Recall the SU(2) algebra of quantum spin: $[S^\alpha, S^\beta] = i\epsilon_{\alpha\beta\gamma}S^\gamma$ (set $\hbar = 1$ for convenience). Defining $S^\pm = S^x \pm iS^y$, we have, equivalently,

$$[S^z, S^\pm] = \pm\, S^\pm \quad , \quad [S^+, S^-] = 2\, S^z \quad . \tag{14.214}$$

The Holstein-Primakoff transformation (1940) maps the spin algebra onto that of a single bosonic oscillator:

$$\begin{aligned} S^+ &= a^\dagger\,(2S - a^\dagger a)^{1/2} \\ S^- &= (2S - a^\dagger a)^{1/2}\, a \\ S^z &= a^\dagger a - S \quad . \end{aligned} \tag{14.215}$$

The state $|\, S^z = -S\,\rangle$ is the vacuum $|\, 0\,\rangle$ in the boson picture. The highest weight state, $|\, S^z = +S\,\rangle$ corresponds to the state $|\, 2S\,\rangle$ in the boson picture, *i.e.* an occupancy of $n = 2S$ bosons.

<u>*EXERCISE:*</u> Verify that the bosonic representation of the spin operators in (15.36) satisfies the SU(2) commutation relations of quantum spin.

What does it mean to take the square root of an operator like $2S - a^\dagger a$? Simple! Just evaluate it in a basis diagonal in $a^\dagger a$, *i.e.* the number basis:

$$a^\dagger a\,|\, n\,\rangle = n\,|\, n\,\rangle \quad \Rightarrow \quad (2S - a^\dagger a)^{1/2}\,|\, n\,\rangle = (2S - n)^{1/2}\,|\, n\,\rangle \quad . \tag{14.216}$$

Note that physical boson states are restricted to $n \in \{0, 1, \ldots, 2S\}$. What about states with $n > 2S$? The nice thing here is that we needn't worry about them at all, because S^+, S^-, and S^z do not connect states with $0 \leqslant n \leqslant 2S$ to states with $n > 2S$. For example, when applying the spin raising operator S^+ to the highest weight state $|\, S^z = +S \,\rangle$, in boson language we have

$$S^+ \,\big|\, S^z = +S \,\big\rangle = a^\dagger \, (2S - a^\dagger a)^{1/2} \,\big|\, n = 2S \,\big\rangle = 0 \quad , \tag{14.217}$$

as required.

While the HP transformation is exact, it really doesn't buy us anything unless we start making some approximations and derive a systematic expansion in 'spin wave' interactions.

14.9.1 Ferromagnetic spin waves

Consider the classical ground state $|\, \mathrm{F} \,\rangle = |\!\downarrow\downarrow \cdots \downarrow\rangle$ in which all spins are pointing 'down', with $S^z = -S$. In the boson language, the occupancy at each site is zero. This is in fact an eigenstate of the Heisenberg Hamiltonian

$$\mathcal{H} = -\sum_{i<j} J_{ij} \, \boldsymbol{S}_i \cdot \boldsymbol{S}_j \tag{14.218}$$

with eigenvalue $E_0 = -S^2 \sum_{i<j} J_{ij}$. If all the interactions are ferromagnetic, *i.e.* $J_{ij} > 0 \; \forall \; (i,j)$, then this state clearly is the ground state. We now express the Heisenberg interaction $\boldsymbol{S}_i \cdot \boldsymbol{S}_j$ in terms of the boson creation and annihilation operators. To this end, we perform a Taylor expansion of the radical,

$$(2S - a^\dagger a)^{1/2} = \sqrt{2S} \left\{ 1 - \frac{1}{2}\left(\frac{a^\dagger a}{2S}\right) - \frac{1}{8}\left(\frac{a^\dagger a}{2S}\right)^2 + \ldots \right\} \quad , \tag{14.219}$$

so that

$$\boldsymbol{S}_i \cdot \boldsymbol{S}_j = \tfrac{1}{2} \, S_i^+ S_j^- + \tfrac{1}{2} \, S_i^- S_j^+ + S_i^z S_j^z \tag{14.220}$$

$$\begin{aligned} &= S \, a_i^\dagger \left(1 - \frac{a_i^\dagger a_i}{4S} + \ldots\right) \left(1 - \frac{a_j^\dagger a_j}{4S} + \ldots\right) a_j \\ &\quad + S \left(1 - \frac{a_i^\dagger a_i}{4S} + \ldots\right) a_i \, a_j^\dagger \left(1 - \frac{a_j^\dagger a_j}{4S} + \ldots\right) + (a_i^\dagger a_i - S)\,(a_j^\dagger a_j - S) \\ &= S^2 + S\Big(a_i^\dagger a_j + a_j^\dagger a_i - a_i^\dagger a_i - a_j^\dagger a_j\Big) + \Big\{ a_i^\dagger a_i a_j^\dagger a_j - \tfrac{1}{4} \, a_i^\dagger a_i^\dagger a_i a_j \\ &\qquad\qquad - \tfrac{1}{4} a_i^\dagger a_j^\dagger a_j a_j - \tfrac{1}{4} \, a_j^\dagger a_i^\dagger a_i a_i - \tfrac{1}{4} a_j^\dagger a_j^\dagger a_j a_i \Big\} + \mathcal{O}(1/S) \quad . \end{aligned} \tag{14.221}$$

Note that a systematic expansion in powers of $1/S$ can be performed. The Heisenberg Hamiltonian now becomes

$$\mathcal{H} = \overbrace{-S^2 \sum_{i<j} J_{ij}}^{\substack{\text{classical ground} \\ \text{state energy } \mathcal{O}(S^2)}} + \overbrace{S \sum_{i<j} J_{ij} \left(a_i^\dagger a_i + a_j^\dagger a_j - a_i^\dagger a_j - a_j^\dagger a_i \right)}^{\text{spin-wave Hamiltonian } \mathcal{H}_{\rm sw}} + \overbrace{\mathcal{O}(S^0)}^{\substack{\text{spin-wave} \\ \text{interactions}}} \quad . \tag{14.222}$$

We assume our sites are elements of a Bravais lattice, and we Fourier transform:

$$a_i = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{q}} e^{+i\boldsymbol{q}\cdot\boldsymbol{R}_i} \, a_{\boldsymbol{q}} \qquad\qquad a_i^\dagger = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{q}} e^{-i\boldsymbol{q}\cdot\boldsymbol{R}_i} \, a_{\boldsymbol{q}}^\dagger \tag{14.223}$$

$$a_{\boldsymbol{q}} = \frac{1}{\sqrt{N}} \sum_{i} e^{-i\boldsymbol{q}\cdot\boldsymbol{R}_i} \, a_i \qquad\qquad a_{\boldsymbol{q}}^\dagger = \frac{1}{\sqrt{N}} \sum_{i} e^{+i\boldsymbol{q}\cdot\boldsymbol{R}_i} \, a_i^\dagger \quad . \tag{14.224}$$

Note that the canonical commutation relations are preserved by this transformation:

$$[a_i, a_j^\dagger] = \delta_{ij} \quad \Longleftrightarrow \quad [a_{\boldsymbol{q}}, a_{\boldsymbol{q}'}^\dagger] = \delta_{\boldsymbol{q}\boldsymbol{q}'} \quad . \tag{14.225}$$

Using the result

$$\frac{1}{N} \sum_i e^{i(\boldsymbol{q}-\boldsymbol{q}')\cdot\boldsymbol{R}_i} = \delta_{\boldsymbol{q}\boldsymbol{q}'} \quad , \tag{14.226}$$

we obtain the spin-wave Hamiltonian

$$\mathcal{H}_{\rm sw} = S \sum_{\boldsymbol{q}} \left[\hat{J}(0) - \hat{J}(\boldsymbol{q}) \right] a_{\boldsymbol{q}}^\dagger a_{\boldsymbol{q}} \quad , \tag{14.227}$$

from which we read off the spin-wave dispersion

$$\begin{aligned} \hbar\omega_{\boldsymbol{q}} &= S\big[\hat{J}(0) - \hat{J}(\boldsymbol{q})\big] \\ &= \tfrac{1}{6} S \Big[\sum_{\boldsymbol{R}} J(\boldsymbol{R}) \, \boldsymbol{R}^2 \Big] \boldsymbol{q}^2 + \mathcal{O}(q^4) \quad . \end{aligned} \tag{14.228}$$

The above sum on $\boldsymbol{R}$ converges if $J(R \to \infty) \sim R^{-(d+2+\epsilon)}$ with $\epsilon > 0$.

14.9.2 Static correlations in the ferromagnet

The transverse spin-spin correlation function is

$$\begin{aligned} \langle S_i^+ S_j^- \rangle &= \left\langle a_i^\dagger \left(2S - a_i^\dagger a_i\right)^{1/2} \left(2S - a_j^\dagger a_j\right)^{1/2} a_j \right\rangle \\ &= 2S \left\langle a_i^\dagger a_j \right\rangle + \mathcal{O}(S^0) \\ &= 2S\Omega \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} \, \frac{e^{i\boldsymbol{k}\cdot(\boldsymbol{R}_j - \boldsymbol{R}_i)}}{e^{\hbar\omega_{\boldsymbol{k}}/k_{\rm B}T} - 1} \quad . \end{aligned} \tag{14.229}$$

The longitudinal spin-spin correlation function is

$$\langle S_i^z S_j^z \rangle - \langle S_i^z \rangle \langle S_j^z \rangle = \langle a_i^\dagger a_i a_j^\dagger a_j \rangle - \langle a_i^\dagger a_i \rangle \langle a_j^\dagger a_j \rangle = \mathcal{O}(S^0) \quad . \tag{14.230}$$

Note that the average spin polarization per site is

$$\begin{aligned} \langle S_i^z \rangle &= -S + \langle a_i^\dagger a_i \rangle \\ &= -S + \Omega \int\limits_{\hat{\Omega}} \frac{d^d\!k}{(2\pi)^d} \, \frac{1}{e^{\hbar\omega_{\boldsymbol{k}}/k_{\rm B}T} - 1} \quad . \end{aligned} \tag{14.231}$$

Now as $k \to 0$ the denominator above vanishes as k^2, hence the average spin polarization per site *diverges* when $d \leqslant 2$. This establishes a "poor man's version" of the HMW theorem: as infinite spin polarization is clearly absurd, there must have been something wrong with our implicit assumption that long-ranged order persists to finite T. In $d = 3$ dimensions, one finds $\langle S_i^z \rangle = -S + \mathcal{O}(T^{3/2})$.

14.9.3 Antiferromagnetic spin waves

The case of the ferromagnet is special because the classical ground state $|\,\mathrm{F}\,\rangle$ is in fact a quantum eigenstate – indeed the ground state – of the ferromagnetic Heisenberg Hamiltonian.[11] In the case of the Heisenberg antiferromagnet, this is no longer the case. The ground state itself is a linear combination of classical states. What is the classical ground state? For an antiferromagnet on a bipartite lattice,[12] the classical ground state has each sublattice maximally polarized, with the magnetization on the two sublattices oppositely oriented. Choosing the axis of polarization as $\hat{\boldsymbol{z}}$, this means $S_i^z = -S$ is $i \in A$ and $S_i^z = +S$ if $i \in B$. We'll call this state $|\,\mathrm{N}\,\rangle$, since it is a classical Néel state.

Let is assume that the lattice is a Bravais lattice with a two-element basis described by basis vectors 0 and $\boldsymbol{\delta}$. Thus, if $\boldsymbol{R}$ is any direct lattice vector, an A sublattice site lies at $\boldsymbol{R}$ and a B site at $\boldsymbol{R} + \boldsymbol{\delta}$. The Heisenberg Hamiltonian is written

$$\begin{aligned} \mathcal{H} = -\sum_{\boldsymbol{R},\boldsymbol{R}'} \Big\{ &\tfrac{1}{2}\, J_{\rm AA}(\boldsymbol{R} - \boldsymbol{R}')\, \boldsymbol{S}_{\rm A}(\boldsymbol{R}) \cdot \boldsymbol{S}_{\rm A}(\boldsymbol{R}') + \tfrac{1}{2}\, J_{\rm BB}(\boldsymbol{R} - \boldsymbol{R}')\, \boldsymbol{S}_{\rm B}(\boldsymbol{R}) \cdot \boldsymbol{S}_{\rm B}(\boldsymbol{R}') \\ &+ J_{\rm AB}(\boldsymbol{R} - \boldsymbol{R}' - \boldsymbol{\delta})\, \boldsymbol{S}_{\rm A}(\boldsymbol{R}) \cdot \boldsymbol{S}_{\rm B}(\boldsymbol{R}') \Big\} \quad . \end{aligned} \tag{14.232}$$

[11] Of course, $|\mathrm{F}\rangle$ is also an eigenstate – the highest lying excited state – of the antiferromagnetic Heisenberg Hamiltonian.

[12] A bipartite lattice is one which may be decomposed into two sublattices A and B, such that all the neighbors of any site in A lie in B, and all the neighbors of any site in B lie in A. Examples of bipartite lattices: square, honeycomb, simple cubic, body-centered cubic, hexagonal. Examples of lattices which are not bipartite: triangular, Kagomé, face-centered cubic.

Here $\boldsymbol{S}_{\rm A}(\boldsymbol{R})$ represents the spin on the A sublattice located at position $\boldsymbol{R}$, while $\boldsymbol{S}_{\rm B}(\boldsymbol{R})$ represents the B sublattice spin located at $\boldsymbol{R}+\boldsymbol{\delta}$. The factor of $\frac{1}{2}$ multiplying the $J_{\rm AA}$ and $J_{\rm BB}$ terms avoids double-counting the AA and BB interactions. The Néel state will be the classical ground state if $J_{\rm AA}>0$ and $J_{\rm BB}>0$ and $J_{\rm AB}<0$. It may remain the ground state even if some of the interactions are frustrating, *i.e.* $J_{\rm AA}<0$, $J_{\rm BB}<0$, and/or $J_{\rm AB}>0$ between certain sites.

We'd like the Néel state $|\,{\rm N}\,\rangle = |\uparrow\downarrow\uparrow\downarrow\uparrow\ldots\rangle$ to be the vacuum for the Holstein-Primakoff bosons. To accomplish this, we rotate the spin operators on the B sublattice by π about the $\hat{\boldsymbol{y}}$-axis in the internal SU(2) space, sending $S^x\to -S^x$, $S^y\to S^y$, and $S^z\to -S^z$. In the language of HP bosons, we have the following:

$$\begin{array}{cc}\underline{\text{A Sublattice}} & \underline{\text{B Sublattice}} \\ S^+ = a^\dagger(2S-a^\dagger a)^{1/2} & S^+ = -(2S-b^\dagger b)^{1/2}b \\ S^- = (2S-a^\dagger a)^{1/2}a & S^- = -b^\dagger(2S-b^\dagger b)^{1/2} \\ S^z = a^\dagger a - S & S^z = S - b^\dagger b\end{array} \tag{14.233}$$

We may now write the Heisenberg interaction as an expansion in powers of $1/S$:

$$\begin{aligned}\boldsymbol{S}_{\rm A}(\boldsymbol{R})\cdot\boldsymbol{S}_{\rm A}(\boldsymbol{R}') &= S^2 + S\left(a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}'} + a^\dagger_{\boldsymbol{R}'}\,a_{\boldsymbol{R}} - a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}} - a^\dagger_{\boldsymbol{R}'}\,a_{\boldsymbol{R}'}\right) + \mathcal{O}(S^0) \\ \boldsymbol{S}_{\rm B}(\boldsymbol{R})\cdot\boldsymbol{S}_{\rm B}(\boldsymbol{R}') &= S^2 + S\left(b^\dagger_{\boldsymbol{R}}\,b_{\boldsymbol{R}'} + b^\dagger_{\boldsymbol{R}'}\,b_{\boldsymbol{R}} - b^\dagger_{\boldsymbol{R}}\,b_{\boldsymbol{R}} - b^\dagger_{\boldsymbol{R}'}\,b_{\boldsymbol{R}'}\right) + \mathcal{O}(S^0) \\ \boldsymbol{S}_{\rm A}(\boldsymbol{R})\cdot\boldsymbol{S}_{\rm B}(\boldsymbol{R}') &= -S^2 + S\left(a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}} + b^\dagger_{\boldsymbol{R}'}\,b_{\boldsymbol{R}'} - a^\dagger_{\boldsymbol{R}}\,b^\dagger_{\boldsymbol{R}'} - a_{\boldsymbol{R}}\,b_{\boldsymbol{R}'}\right) + \mathcal{O}(S^0)\quad .\end{aligned} \tag{14.234}$$

Thus, the classical ground state energy is the $\mathcal{O}(S^2)$ term,

$$E_{\rm cl} = S^2\sum_{\boldsymbol{R},\boldsymbol{R}'}\Big\{-\tfrac{1}{2}\,J_{\rm AA}(\boldsymbol{R}-\boldsymbol{R}') - \tfrac{1}{2}\,J_{\rm BB}(\boldsymbol{R}-\boldsymbol{R}') + J_{\rm AB}(\boldsymbol{R}-\boldsymbol{R}'-\boldsymbol{\delta})\Big\}\quad . \tag{14.235}$$

The spin-wave Hamiltonian is the $\mathcal{O}(S^1)$ piece,

$$\begin{aligned}\mathcal{H}_{\rm sw} = -S\sum_{\boldsymbol{R},\boldsymbol{R}'}\Big\{&J_{\rm AA}(\boldsymbol{R}-\boldsymbol{R}')\Big(a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}'} - a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}}\Big) + J_{\rm BB}(\boldsymbol{R}-\boldsymbol{R}')\Big(b^\dagger_{\boldsymbol{R}}\,b_{\boldsymbol{R}'} - b^\dagger_{\boldsymbol{R}}\,b_{\boldsymbol{R}}\Big) \\ &+ J_{\rm AB}(\boldsymbol{R}-\boldsymbol{R}'-\boldsymbol{\delta})\Big(a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}} + b^\dagger_{\boldsymbol{R}}\,b_{\boldsymbol{R}} - a^\dagger_{\boldsymbol{R}}\,b^\dagger_{\boldsymbol{R}'} - a_{\boldsymbol{R}}\,b_{\boldsymbol{R}'}\Big)\Big\}\quad .\end{aligned} \tag{14.236}$$

We now Fourier transform:

$$a_{\boldsymbol{R}} = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{+i\boldsymbol{k}\cdot\boldsymbol{R}}\,a_{\boldsymbol{k}} \qquad\qquad a^\dagger_{\boldsymbol{R}} = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\,a^\dagger_{\boldsymbol{k}} \tag{14.237}$$

$$b_{\boldsymbol{R}} = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{+i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\delta})}\,b_{\boldsymbol{k}} \qquad\qquad b^\dagger_{\boldsymbol{R}} = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\delta})}\,b^\dagger_{\boldsymbol{k}}\quad , \tag{14.238}$$

which leads to

$$\sum_{\boldsymbol{R},\boldsymbol{R}'} J_{\rm AA}(\boldsymbol{R}-\boldsymbol{R}')\,a^\dagger_{\boldsymbol{R}}\,a_{\boldsymbol{R}'} = \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'}\sum_{\boldsymbol{R},\boldsymbol{R}'} J_{\rm AA}(\boldsymbol{R}-\boldsymbol{R}')e^{i(\boldsymbol{k}'\cdot\boldsymbol{R}'-\boldsymbol{k}\cdot\boldsymbol{R})}\,a^\dagger_{\boldsymbol{k}}a_{\boldsymbol{k}'}$$
$$= \sum_{\boldsymbol{k}} \hat{J}_{\rm AA}(\boldsymbol{k})\,a^\dagger_{\boldsymbol{k}}\,a_{\boldsymbol{k}} \tag{14.239}$$

$$\sum_{\boldsymbol{R},\boldsymbol{R}'} J_{\rm AB}(\boldsymbol{R}-\boldsymbol{R}'-\boldsymbol{\delta})\,a^\dagger_{\boldsymbol{R}}\,b^\dagger_{\boldsymbol{R}'} = \frac{1}{N}\sum_{\boldsymbol{k},\boldsymbol{k}'}\sum_{\boldsymbol{R},\boldsymbol{R}'} J_{\rm AB}(\boldsymbol{R}-\boldsymbol{R}'-\boldsymbol{\delta})e^{i\left(\boldsymbol{k}'\cdot(\boldsymbol{R}'+\boldsymbol{\delta})-\boldsymbol{k}\cdot\boldsymbol{R}\right)}\,a^\dagger_{\boldsymbol{k}}b^\dagger_{-\boldsymbol{k}'}$$
$$= \sum_{\boldsymbol{k}} \hat{J}_{\rm AB}(\boldsymbol{k})\,a^\dagger_{\boldsymbol{k}}\,b^\dagger_{-\boldsymbol{k}} \quad , \tag{14.240}$$

where, assuming $J_{\rm AA}$, $J_{\rm BB}$ and $J_{\rm AB}$ are functions only of the magnitude of their arguments,

$$\begin{aligned}
\hat{J}_{\rm AA}(\boldsymbol{k}) &\equiv \sum_{\boldsymbol{R}} J_{\rm AA}\big(|\boldsymbol{R}|\big)\,e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \\
\hat{J}_{\rm BB}(\boldsymbol{k}) &\equiv \sum_{\boldsymbol{R}} J_{\rm BB}\big(|\boldsymbol{R}|\big)\,e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \\
\hat{J}_{\rm AB}(\boldsymbol{k}) &\equiv \sum_{\boldsymbol{R}} J_{\rm AB}\big(|\boldsymbol{R}+\delta|\big)\,e^{i\boldsymbol{k}\cdot(\boldsymbol{R}+\boldsymbol{\delta})} \quad .
\end{aligned} \tag{14.241}$$

Note that $\hat{J}_{\rm AA}(\boldsymbol{k}) = \hat{J}_{\rm AA}(-\boldsymbol{k}) = \big[\hat{J}_{\rm AA}(\boldsymbol{k})\big]^*$ (similarly for $J_{\rm BB}$), and $\hat{J}_{\rm AB}(\boldsymbol{k}) = \big[\hat{J}_{\rm AB}(-\boldsymbol{k})\big]^*$.
The spin-wave Hamiltonian may now be written as

$$\mathcal{H}_{\rm sw} = S\sum_{\boldsymbol{k}}\Big\{\Big(\hat{J}_{\rm AA}(0)-\hat{J}_{\rm AA}(\boldsymbol{k})-\hat{J}_{\rm AB}(0)\Big)\,a^\dagger_{\boldsymbol{k}}\,a_{\boldsymbol{k}} + \Big(\hat{J}_{\rm BB}(0)-\hat{J}_{\rm BB}(\boldsymbol{k})-\hat{J}_{\rm AB}(0)\Big)\,b^\dagger_{\boldsymbol{k}}\,b_{\boldsymbol{k}}$$
$$+\,\hat{J}_{\rm AB}(\boldsymbol{k})\,a^\dagger_{\boldsymbol{k}}\,b^\dagger_{-\boldsymbol{k}} + J^*_{\rm AB}(\boldsymbol{k})\,a_{\boldsymbol{k}}\,b_{-\boldsymbol{k}}\Big\} \quad . \tag{14.242}$$

In other words,

$$\boxed{\mathcal{H}_{\rm sw} = \sum_{\boldsymbol{k}}\Big\{\Omega^{\rm AA}_{\boldsymbol{k}}\,a^\dagger_{\boldsymbol{k}}\,a_{\boldsymbol{k}} + \Omega^{\rm BB}_{\boldsymbol{k}}\,b^\dagger_{\boldsymbol{k}}\,b_{\boldsymbol{k}} + \Delta_{\boldsymbol{k}}\,a^\dagger_{\boldsymbol{k}}\,b^\dagger_{-\boldsymbol{k}} + \Delta^*_{\boldsymbol{k}}\,a_{\boldsymbol{k}}\,b_{-\boldsymbol{k}}\Big\}} \tag{14.243}$$

with

$$\begin{aligned}
\Omega^{\rm AA}_{\boldsymbol{k}} &= S\Big(\hat{J}_{\rm AA}(0)-\hat{J}_{\rm AA}(\boldsymbol{k})-\hat{J}_{\rm AB}(0)\Big) \\
\Omega^{\rm BB}_{\boldsymbol{k}} &= S\Big(\hat{J}_{\rm BB}(0)-\hat{J}_{\rm BB}(\boldsymbol{k})-\hat{J}_{\rm AB}(0)\Big)
\end{aligned} \tag{14.244}$$

and

$$\Delta_{\boldsymbol{k}} = S\,\hat{J}_{\rm AB}(\boldsymbol{k}) \quad . \tag{14.245}$$

Henceforth we shall assume $J_{\rm AA}(\boldsymbol{R}) = J_{\rm BB}(\boldsymbol{R})$, so $\Omega_{\boldsymbol{k}}^{\rm AA} = \Omega_{\boldsymbol{k}}^{\rm BB} \equiv \Omega_{\boldsymbol{k}}$.

Note that the vacuum $\big|\,0\,\big\rangle$ for the a and b bosons is not an eigenstate of $\mathcal{H}_{\rm sw}$, owing to the spin-wave pair creation term $\Delta_{\boldsymbol{k}}^*\, a_{\boldsymbol{k}}\, b_{-\boldsymbol{k}}$. This can be traced back to the effect on the Néel state of the Heisenberg interaction,

$$\boldsymbol{S}_i \cdot \boldsymbol{S}_j = \tfrac{1}{2}\, S_i^+\, S_j^- + \tfrac{1}{2}\, S_i^-\, S_j^+ + S_i^z\, S_j^z \quad . \tag{14.246}$$

If $i \in {\rm A}$ and $j \in {\rm B}$, then the term $S_i^+\, S_j^-$ acts on the configuration $|\,-S\,,\,+S\,\rangle$ and converts it to $2S\,|\,-S+1\,,\,S-1\,\rangle$. Nevertheless, we can diagonalize $\mathcal{H}_{\rm sw}$ by means of a canonical (but not unitary!) transformation, known as the *Bogoliubov transformation*. Note that for each $\boldsymbol{k} \in \hat{\Omega}$, the spin-wave Hamiltonian couples only four operators: $a_{\boldsymbol{k}}^\dagger$, $a_{\boldsymbol{k}}$, $b_{-\boldsymbol{k}}^\dagger$, and $b_{-\boldsymbol{k}}$. We write the Bogoliubov transformation as

$$a_{\boldsymbol{k}} = u_{\boldsymbol{k}}\, \alpha_{\boldsymbol{k}} - v_{\boldsymbol{k}}^*\, \beta_{-\boldsymbol{k}}^\dagger \qquad\qquad b_{-\boldsymbol{k}} = u_{\boldsymbol{k}}\, \beta_{-\boldsymbol{k}} - v_{\boldsymbol{k}}^*\, \alpha_{\boldsymbol{k}}^\dagger \tag{14.247}$$

$$a_{\boldsymbol{k}}^\dagger = u_{\boldsymbol{k}}^*\, \alpha_{\boldsymbol{k}}^\dagger - v_{\boldsymbol{k}}\, \beta_{-\boldsymbol{k}} \qquad\qquad b_{-\boldsymbol{k}}^\dagger = u_{\boldsymbol{k}}^*\, \beta_{-\boldsymbol{k}}^\dagger - v_{\boldsymbol{k}}\, \alpha_{\boldsymbol{k}} \tag{14.248}$$

One can readily verify that this transformation preserves the canonical bosonic commutation relations,

$$\big[a_{\boldsymbol{k}}, a_{\boldsymbol{k}'}^\dagger\big] = \big[b_{\boldsymbol{k}}, b_{\boldsymbol{k}'}^\dagger\big] = \big[\alpha_{\boldsymbol{k}}, \alpha_{\boldsymbol{k}'}^\dagger\big] = \big[\beta_{\boldsymbol{k}}, \beta_{\boldsymbol{k}'}^\dagger\big] = \delta_{\boldsymbol{k}\boldsymbol{k}'} \tag{14.249}$$

provided that

$$u_{\boldsymbol{k}}^*\, u_{\boldsymbol{k}} - v_{\boldsymbol{k}}^*\, v_{\boldsymbol{k}} = 1 \quad . \tag{14.250}$$

The inverse transformation is

$$\alpha_{\boldsymbol{k}} = u_{\boldsymbol{k}}^*\, a_{\boldsymbol{k}} + v_{\boldsymbol{k}}^*\, b_{-\boldsymbol{k}}^\dagger \qquad\qquad \beta_{-\boldsymbol{k}} = u_{\boldsymbol{k}}^*\, b_{-\boldsymbol{k}} + v_{\boldsymbol{k}}^*\, a_{\boldsymbol{k}}^\dagger \tag{14.251}$$

$$\alpha_{\boldsymbol{k}}^\dagger = u_{\boldsymbol{k}}\, a_{\boldsymbol{k}}^\dagger + v_{\boldsymbol{k}}\, b_{-\boldsymbol{k}} \qquad\qquad \beta_{-\boldsymbol{k}}^\dagger = u_{\boldsymbol{k}}\, b_{-\boldsymbol{k}}^\dagger + v_{\boldsymbol{k}}\, a_{\boldsymbol{k}} \quad . \tag{14.252}$$

We'll write

$$u_{\boldsymbol{k}} = \exp(i\eta_{\boldsymbol{k}})\, \cosh(\theta_{\boldsymbol{k}}) \qquad , \qquad v_{\boldsymbol{k}} = \exp(-i\eta_{\boldsymbol{k}})\, \sinh(\theta_{\boldsymbol{k}}) \quad . \tag{14.253}$$

We may then write

$$\begin{aligned} a_{\boldsymbol{k}} &= \exp(i\eta_{\boldsymbol{k}})\, \cosh(\theta_{\boldsymbol{k}})\, \alpha_{\boldsymbol{k}} - \exp(i\eta_{\boldsymbol{k}})\, \sinh(\theta_{\boldsymbol{k}})\, \beta_{-\boldsymbol{k}}^\dagger \\ b_{-\boldsymbol{k}} &= \exp(i\eta_{\boldsymbol{k}})\, \cosh(\theta_{\boldsymbol{k}})\, \beta_{-\boldsymbol{k}} - \exp(i\eta_{\boldsymbol{k}})\, \sinh(\theta_{\boldsymbol{k}})\, \alpha_{\boldsymbol{k}}^\dagger \end{aligned} \tag{14.254}$$

as well as the inverse

$$\begin{aligned} \alpha_{\boldsymbol{k}} &= \exp(-i\eta_{\boldsymbol{k}})\, \cosh(\theta_{\boldsymbol{k}})\, a_{\boldsymbol{k}} + \exp(i\eta_{\boldsymbol{k}})\, \sinh(\theta_{\boldsymbol{k}})\, b_{-\boldsymbol{k}}^\dagger \\ \beta_{-\boldsymbol{k}} &= \exp(-i\eta_{\boldsymbol{k}})\, \cosh(\theta_{\boldsymbol{k}})\, b_{-\boldsymbol{k}} + \exp(i\eta_{\boldsymbol{k}})\, \sinh(\theta_{\boldsymbol{k}})\, a_{\boldsymbol{k}}^\dagger \quad . \end{aligned} \tag{14.255}$$

Substituting into the expressions from $\mathcal{H}_{\rm sw}$, we find

$$\begin{aligned} \Omega_{\boldsymbol{k}}\, \big(a_{\boldsymbol{k}}^\dagger\, a_{\boldsymbol{k}} + b_{\boldsymbol{k}}^\dagger\, b_{\boldsymbol{k}}\big) = \Omega_{\boldsymbol{k}}\, \cosh(2\theta_{\boldsymbol{k}})\, \big(\alpha_{\boldsymbol{k}}^\dagger\, \alpha_{\boldsymbol{k}} + \beta_{-\boldsymbol{k}}^\dagger\, \beta_{-\boldsymbol{k}} + 1\big) - \Omega_{\boldsymbol{k}} \\ - \Omega_{\boldsymbol{k}}\, \sinh(2\theta_{\boldsymbol{k}})\, \big(\alpha_{\boldsymbol{k}}^\dagger\, \beta_{-\boldsymbol{k}}^\dagger + \alpha_{\boldsymbol{k}}\, \beta_{-\boldsymbol{k}}\big) \end{aligned} \tag{14.256}$$

and

$$\begin{aligned}\Delta_{\boldsymbol{k}}\,a_{\boldsymbol{k}}^\dagger\,a_{-\boldsymbol{k}}^\dagger + \Delta_{\boldsymbol{k}}^*\,a_{\boldsymbol{k}}\,b_{-\boldsymbol{k}} = &-\big|\Delta_{\boldsymbol{k}}\big|\,\sinh(2\theta_{\boldsymbol{k}})\,\big(\alpha_{\boldsymbol{k}}^\dagger\,\alpha_{\boldsymbol{k}} + \beta_{-\boldsymbol{k}}^\dagger\,\beta_{-\boldsymbol{k}} + 1\big) \\ &+ \big|\Delta_{\boldsymbol{k}}\big|\,\cosh(2\theta_{\boldsymbol{k}})\,\big(\alpha_{\boldsymbol{k}}^\dagger\,\beta_{-\boldsymbol{k}}^\dagger + \alpha_{\boldsymbol{k}}\,\beta_{-\boldsymbol{k}}\big) \quad ,\end{aligned} \tag{14.257}$$

where we have taken $\eta_{\boldsymbol{k}} = \frac{1}{2}\arg(\Delta_{\boldsymbol{k}})$. Up until now, $\theta_{\boldsymbol{k}}$ has been arbitrary. We now use this freedom to specify $\theta_{\boldsymbol{k}}$ such that the $\big(\alpha_{\boldsymbol{k}}^\dagger\,\beta_{-\boldsymbol{k}}^\dagger + \alpha_{\boldsymbol{k}}\,\beta_{-\boldsymbol{k}}\big)$ terms vanish from $\mathcal{H}_{\rm sw}$. This requires

$$\big|\Delta_{\boldsymbol{k}}\big|\,\cosh(2\theta_{\boldsymbol{k}}) - \Omega_{\boldsymbol{k}}\,\sinh(2\theta_{\boldsymbol{k}}) = 0 \quad\Longrightarrow\quad \tanh(2\theta_{\boldsymbol{k}}) = \frac{\big|\Delta_{\boldsymbol{k}}\big|}{\Omega_{\boldsymbol{k}}} \quad , \tag{14.258}$$

which means

$$\cosh(2\theta_{\boldsymbol{k}}) = \frac{\Omega_{\boldsymbol{k}}}{E_{\boldsymbol{k}}} \quad , \quad \sinh(2\theta_{\boldsymbol{k}}) = \frac{\big|\Delta_{\boldsymbol{k}}\big|}{E_{\boldsymbol{k}}} \tag{14.259}$$

along with the dispersion relation

$$E_{\boldsymbol{k}} = \sqrt{\Omega_{\boldsymbol{k}}^2 - \big|\Delta_{\boldsymbol{k}}\big|^2} \quad . \tag{14.260}$$

Finally, we may write the diagonalized spin-wave Hamiltonian as

$$\boxed{\mathcal{H}_{\rm sw} = \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}\,\big(\alpha_{\boldsymbol{k}}^\dagger\,\alpha_{\boldsymbol{k}} + \beta_{\boldsymbol{k}}^\dagger\,\beta_{\boldsymbol{k}}\big) + \sum_{\boldsymbol{k}}\big(E_{\boldsymbol{k}} - \Omega_{\boldsymbol{k}}\big) \quad .} \tag{14.261}$$

Note that $E_{\boldsymbol{k}} = E_{-\boldsymbol{k}}$ since $\hat{J}_{\rm AB}(\boldsymbol{k}) = \hat{J}_{\rm AB}^*(-\boldsymbol{k})$. The two terms above represent, respectively, the spin-wave excitation Hamiltonian, and the $\mathcal{O}(S^1)$ quantum correction to the ground state energy. Since $E_{\boldsymbol{k}} < \Omega_{\boldsymbol{k}}$, this correction is always negative.

As $\boldsymbol{k} \to 0$, we have, assuming cubic or higher symmetry,

$$\begin{aligned}\Omega_{\boldsymbol{k}} &= -S\sum_{\boldsymbol{R}} J_{\rm AB}\big(|\boldsymbol{R}+\boldsymbol{\delta}|\big) + \tfrac{1}{6}\,S\,\boldsymbol{k}^2\sum_{\boldsymbol{R}} J_{\rm AA}\big(|\boldsymbol{R}|\big)\,\boldsymbol{R}^2 + \ldots \\ &\equiv SW + SX\boldsymbol{k}^2 + \ldots\end{aligned} \tag{14.262}$$

and

$$\begin{aligned}\Delta_{\boldsymbol{k}} &= +S\sum_{\boldsymbol{R}} J_{\rm AB}\big(|\boldsymbol{R}+\boldsymbol{\delta}|\big) - \tfrac{1}{6}\,S\,\boldsymbol{k}^2\sum_{\boldsymbol{R}} J_{\rm AB}\big(|\boldsymbol{R}+\boldsymbol{\delta}|\big)\,\big|\boldsymbol{R}+\boldsymbol{\delta}\big|^2 + \ldots \\ &\equiv -SW + SY\boldsymbol{k}^2 + \ldots \quad .\end{aligned} \tag{14.263}$$

The energy dispersion is linear: $E_{\boldsymbol{k}} = \hbar c|\boldsymbol{k}|$, where $c = S\sqrt{2W(X+Y)}$. Antiferromagnetic spin waves are Goldstone bosons corresponding to the broken continuous symmetry of global spin rotation. The dispersion vanishes linearly as $\boldsymbol{k} \to 0$, in contrast to the case of ferromagnetic spin waves, where $E_{\boldsymbol{k}}$ vanishes quadratically.

Reduction in Sublattice Magnetization

Let's compute the average of S^z for a spin on the A sublattice:

$$\begin{aligned}\langle S^z(\boldsymbol{R})\rangle &= -S + \langle a^\dagger_{\boldsymbol{R}}\, a_{\boldsymbol{R}}\rangle \\ &= -S + \frac{1}{N}\sum_{\boldsymbol{k}}\langle a^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}}\rangle \\ &= -S + \frac{1}{N}\sum_{\boldsymbol{k}}\left\langle (u^*_{\boldsymbol{k}}\,\alpha^\dagger_{\boldsymbol{k}} - v_{\boldsymbol{k}}\,\beta_{-\boldsymbol{k}})\,(u_{\boldsymbol{k}}\,\alpha_{\boldsymbol{k}} - v^*_{\boldsymbol{k}}\,\beta^\dagger_{-\boldsymbol{k}})\right\rangle \\ &= -S + v_0\int\limits_{\rm BZ}\frac{d^dk}{(2\pi)^d}\left\{\frac{\Omega_{\boldsymbol{k}}}{E_{\boldsymbol{k}}}\,\frac{1}{\exp(E_{\boldsymbol{k}}/k_{\rm B}T)-1} + \frac{1}{2}\left(\frac{\Omega_{\boldsymbol{k}}}{E_{\boldsymbol{k}}}-1\right)\right\}\quad ,\end{aligned} \tag{14.264}$$

where v_0 is the Wigner-Seitz cell volume, and the integral is over the first Brillouin zone. The deviation $\delta S^z = \langle a^\dagger\, a\rangle$ from the classical value $\langle S^z\rangle = -S$ is due to thermal and quantum fluctuations. Note that even at $T = 0$, when the thermal fluctuations vanish, there is still a reduction in sublattice magnetization due to quantum fluctuations. The Néel state satisfies the $S^z_i S^z_j$ part of the Heisenberg interaction, but the full interaction prefers neighboring spins to be arranged in *singlets*, which involves fluctuations about local Néel order.

We've seen that $\Omega_{\boldsymbol{k}} \simeq SW$ and $E_{\boldsymbol{k}} \simeq \hbar c\,|\boldsymbol{k}|$ as $\boldsymbol{k}\to 0$. Thus, the integrand behaves as $T/\boldsymbol{k}^2$ for the first term and as $1/|\boldsymbol{k}|$ for the second term. The integral therefore diverges in $d \leqslant 2$ at finite T and in $d = 1$ even at $T = 0$. Thermal and quantum fluctuations melt the classical ordered state.

14.9.4 Specific heat due to spin waves

The long wavelength dispersion $\omega_{\boldsymbol{q}} = Aq^2$ has thermodynamic consequences. Consider a general case of a bosonic dispersion $\omega_{\boldsymbol{q}} = A|\boldsymbol{q}|^\sigma$. The internal energy for a system in d space dimensions is then

$$\begin{aligned}E(T) &= V\int\frac{d^dk}{(2\pi)^d}\,\frac{Ak^\sigma}{e^{\beta Aq^\sigma}-1} \\ &= \frac{AV\Omega_d}{(2\pi)^d}\Big(\frac{k_{\rm B}T}{A}\Big)^{1+\frac{d}{\sigma}}\int\limits_0^\infty du\,\frac{u^{d/\sigma}}{e^u-1}\end{aligned} \tag{14.265}$$

where $\Omega_d = 2\pi^{d/2}/\Gamma(d/2)$ is the area of the unit sphere in d dimensions. Thus, $E(T)\propto T^{1+\frac{d}{\sigma}}$, leading to a low-temperature heat capacity of

$$C_V = \Gamma(2+\tfrac{1}{2}d)\,\zeta(1+\tfrac{1}{2}d)\,\frac{k_{\rm B}V\Omega_d}{(2\pi)^d}\left(\frac{k_{\rm B}T}{A}\right)^{d/\sigma}\quad . \tag{14.266}$$

At high T, one must impose a cutoff at the edge of the Brillouin zone, where $k \sim \pi/a$, in order not to overcount the modes. One finds

$$E(T) = k_{\rm B} T \, V \int\limits_{\hat{\Omega}} \frac{d^d k}{(2\pi)^d} = N k_{\rm B} T \quad , \tag{14.267}$$

where N is the number of unit cells. This simply is the Dulong-Petit result of $k_{\rm B}T$ per mode.

For ferromagnetic spin waves, we found $\sigma = 2$, hence $C_V \propto T^{d/2}$ at low temperatures. As we shall see, for antiferromagnetic spin waves, one has $\sigma = 1$, as in the case of acoustic phonons, hence $C_V \propto T^d$.

Suppose we write the long-wavelength ferromagnetic spin-wave dispersion as $\hbar\omega_{\boldsymbol{q}} = CJ(\boldsymbol{q}a)^2$, where a is the lattice spacing, J is the nearest neighbor exchange, and C is a dimensionless constant. The ferromagnetic low-temperature specific heat is then

$$C_V^{\rm F} = \Gamma(2 + \tfrac{1}{2}d)\, \zeta(1 + \tfrac{1}{2}d)\, \frac{k_{\rm B} V \Omega_d}{(2\pi a)^d} \left(\frac{k_{\rm B} T}{CJ} \right)^{d/2} \quad , \tag{14.268}$$

hence $C_V^{\rm F} \propto (T/\Theta_J)^{d/2}$, with $\Theta_J \equiv CJ/k_{\rm B}$. Acoustic phonons with a $\omega_{\boldsymbol{k}} = \hbar c |\boldsymbol{k}|$ dispersion lead to a Debye heat capacity

$$C_V^{\rm D} = \Gamma(2 + d)\, \zeta(1 + d)\, \frac{k_{\rm B} V \Omega_d}{(2\pi a)^d} \left(\frac{k_{\rm B} T}{\hbar c/a} \right)^{d} \quad , \tag{14.269}$$

hence $C^{\rm D} \propto (T/\Theta_{\rm D})^d$, with $\Theta_{\rm D} \equiv \hbar c/a k_{\rm B}$. Thus, at the lowest temperatures, the specific heat due to spin waves dominates, but at intermediate temperatures it is the phonon specific heat which dominates. The temperature scale T^* at which the two contributions are roughly equal is given by

$$(T^*/\Theta_J)^{d/2} \simeq (T^*/\Theta_{\rm D})^d \quad \Longrightarrow \quad T^* \simeq \Theta_{\rm D}^2/\Theta_J \quad . \tag{14.270}$$

14.10 Appendix : Generalized Spin Wave Theory for Isotropic Systems

14.10.1 General form of Heisenberg Hamiltonian

Consider an isotropic Heisenberg Hamiltonian,

$$H = \sum_{i<j} J_{ij} \boldsymbol{S}_i \cdot \boldsymbol{S}_j \quad , \tag{14.271}$$

defined on an arbitrary lattice structure. On each site i, we may rotate the spin operators, writing

$$S^\mu = \mathsf{e}^\alpha_\mu \, \mathsf{S}^\alpha \quad , \tag{14.272}$$

where the unit vectors $\{\hat{\mathsf{e}}^1, \hat{\mathsf{e}}^2, \hat{\mathsf{e}}^3\}$ satisfy

$$\hat{\mathsf{e}}^\alpha \times \hat{\mathsf{e}}^\beta = \epsilon^{\alpha\beta\gamma} \, \hat{\mathsf{e}}^\gamma \quad . \tag{14.273}$$

I.e. they form an orthonormal triad. The Heisenberg interaction between spins on sites i and j may then be written

$$\boldsymbol{S}_i \cdot \boldsymbol{S}_j = \hat{\mathsf{e}}^\alpha_i \cdot \hat{\mathsf{e}}^\beta_j \, \mathsf{S}^\alpha_i \, \mathsf{S}^\beta_j \quad . \tag{14.274}$$

We now represent the spin operators S^α in terms of Holstein-Primakoff bosons:

$$\mathsf{S}^+ = \psi^\dagger \left(2S - \psi^\dagger\psi\right)^{1/2} \quad , \quad \mathsf{S}^- = \left(2S - \psi^\dagger\psi\right)^{1/2} \psi \quad , \quad \mathsf{S}^z = \psi^\dagger\psi - S \quad . \tag{14.275}$$

We now expand the Heisenberg interaction in powers of S. Including terms of orders S^2, $S^{3/2}$, and S, and neglecting terms of $\mathcal{O}(S^{1/2})$, we have

$$\mathsf{S}^\alpha_i \, \mathsf{S}^\beta_j = \begin{pmatrix} \frac{1}{2}S\big(\psi_i + \psi_i^\dagger\big)\big(\psi_j + \psi_j^\dagger\big) & \frac{i}{2}S\big(\psi_i + \psi_i^\dagger\big)\big(\psi_j - \psi_j^\dagger\big) & -\frac{1}{\sqrt{2}}S^{3/2}\big(\psi_i + \psi_i^\dagger\big) \\ \frac{i}{2}S\big(\psi_i - \psi_i^\dagger\big)\big(\psi_j + \psi_j^\dagger\big) & -\frac{1}{2}S\big(\psi_i - \psi_i^\dagger\big)\big(\psi_j - \psi_j^\dagger\big) & -\frac{i}{\sqrt{2}}S^{3/2}\big(\psi_i - \psi_i^\dagger\big) \\ -\frac{1}{\sqrt{2}}S^{3/2}\big(\psi_j + \psi_j^\dagger\big) & -\frac{i}{\sqrt{2}}S^{3/2}\big(\psi_j - \psi_j^\dagger\big) & S^2 - S\big(\psi_i^\dagger\psi_i + \psi_j^\dagger\psi_j\big) \end{pmatrix}$$

$$= S^2 \, \delta_{\alpha 3} \, \delta_{\beta 3} + S^{3/2} \, C^{\alpha\beta}_{ij} + S \, Q^{\alpha\beta}_{ij} + \mathcal{O}\big(S^{1/2}\big) \quad , \tag{14.276}$$

where

$$C^{\alpha\beta}_{ij} = \begin{pmatrix} 0 & 0 & -\frac{1}{\sqrt{2}}\big(\psi_i + \psi_i^\dagger\big) \\ 0 & 0 & -\frac{i}{\sqrt{2}}\big(\psi_i - \psi_i^\dagger\big) \\ -\frac{1}{\sqrt{2}}\big(\psi_j + \psi_j^\dagger\big) & -\frac{i}{\sqrt{2}}\big(\psi_j - \psi_j^\dagger\big) & 0 \end{pmatrix} \tag{14.277}$$

and

$$Q^{\alpha\beta}_{ij} = \begin{pmatrix} \frac{1}{2}\big(\psi_i + \psi_i^\dagger\big)\big(\psi_j + \psi_j^\dagger\big) & \frac{i}{2}\big(\psi_i + \psi_i^\dagger\big)\big(\psi_j - \psi_j^\dagger\big) & 0 \\ \frac{i}{2}\big(\psi_i - \psi_i^\dagger\big)\big(\psi_j + \psi_j^\dagger\big) & -\frac{1}{2}\big(\psi_i - \psi_i^\dagger\big)\big(\psi_j - \psi_j^\dagger\big) & 0 \\ 0 & 0 & -\psi_i^\dagger\psi_i - \psi_j^\dagger\psi_j \end{pmatrix} \quad . \tag{14.278}$$

The classical energy is

$$E_0^{\rm CL} = S^2 \sum_{i<j} J_{ij} \, \hat{\mathsf{e}}^3_i \cdot \hat{\mathsf{e}}^3_j \quad . \tag{14.279}$$

The $\mathcal{O}\big(S^{3/2}\big)$ term is

$$H_1 \equiv S^{3/2} \sum_{i<j} J_{ij} \, \hat{\mathsf{e}}^\alpha_i \cdot \hat{\mathsf{e}}^\beta_j \, C^{\alpha\beta}_{ij} \tag{14.280}$$

$$= -\tfrac{1}{\sqrt{2}} S^{3/2} \sum_i \left(\hat{\mathsf{e}}^+_i \cdot \sum_j J_{ij} \, \hat{\mathsf{e}}^3_j \right) \psi_i - \tfrac{1}{\sqrt{2}} S^{3/2} \sum_i \left(\hat{\mathsf{e}}^-_i \cdot \sum_j J_{ij} \, \hat{\mathsf{e}}^3_j \right) \psi_i^\dagger \tag{14.281}$$

where

$$\hat{\mathsf{e}}_i^{\pm} \equiv \hat{\mathsf{e}}_i^1 \pm i\,\hat{\mathsf{e}}_i^2 \quad . \tag{14.282}$$

Therefore, if for each i we have

$$\sum_j J_{ij}\,\hat{\mathsf{e}}_j^3 \propto \hat{\mathsf{e}}_i^3 \qquad \text{or} \qquad \sum_j J_{ij}\,\hat{\mathsf{e}}_j^3 = 0 \quad , \tag{14.283}$$

then

$$\hat{\mathsf{e}}_i^{\pm} \cdot \sum_j J_{ij}\,\hat{\mathsf{e}}_j^3 = 0 \quad , \tag{14.284}$$

and H_1 vanishes. This is the condition that the classical ground state configuration lie at a local extremum of the energy. If the condition in eqn. (14.283) did not hold, then there would be a finite mean field from the neighbors of site i whose direction was not completely aligned with the moment on that site. The spin on site i would then be able to lower its energy by canting to antialign with this mean field.

The $\mathcal{O}(S)$ piece of the Hamiltonian is the spin-wave contribution:

$$\begin{aligned} H_{\rm sw} = S \sum_{i<j} J_{ij} \Big\{ &\tfrac{1}{2}\,\hat{\mathsf{e}}_i^- \cdot \hat{\mathsf{e}}_j^+\,\psi_i^\dagger\,\psi_j + \tfrac{1}{2}\,\hat{\mathsf{e}}_i^+ \cdot \hat{\mathsf{e}}_j^-\,\psi_i\,\psi_j^\dagger \\ &+ \tfrac{1}{2}\,\hat{\mathsf{e}}_i^- \cdot \hat{\mathsf{e}}_j^-\,\psi_i^\dagger\,\psi_j^\dagger + \tfrac{1}{2}\,\hat{\mathsf{e}}_i^+ \cdot \hat{\mathsf{e}}_j^+\,\psi_i\,\psi_j - \hat{\mathsf{e}}_i^3 \cdot \hat{\mathsf{e}}_j^3\,\left(\psi_i^\dagger\psi_i + \psi_j^\dagger\psi_j\right) \Big\} \quad . \end{aligned} \tag{14.285}$$

Note that rotation of the basis on site i by an angle θ_i about $\hat{\mathsf{e}}_i^3$ entails $\hat{\mathsf{e}}_i^{\pm} \to e^{\pm i\theta_i}\,\hat{\mathsf{e}}_i^{\pm}$, which is then cancelled by the unitary transformation $\psi_i \to e^{i\theta_i}\,\psi_i$ and $\psi_i^\dagger \to e^{-i\theta_i}\,\psi_i^\dagger$.

14.10.2 Planar spiral phases

As a general parameterization of the classical state, take

$$\hat{\mathsf{e}}_i^1 = \cos\theta_i\,\cos\phi_i\,\hat{\boldsymbol{x}} + \cos\theta_i\,\sin\phi_i\,\hat{\boldsymbol{y}} - \sin\theta_i\,\hat{\boldsymbol{z}} \tag{14.286}$$

$$\hat{\mathsf{e}}_i^2 = -\sin\phi_i\,\hat{\boldsymbol{x}} + \cos\phi_i\,\hat{\boldsymbol{y}} \tag{14.287}$$

$$\hat{\mathsf{e}}_i^3 = \sin\theta_i\,\cos\phi_i\,\hat{\boldsymbol{x}} + \sin\theta_i\,\sin\phi_i\,\hat{\boldsymbol{y}} + \cos\theta_i\,\hat{\boldsymbol{z}} \quad . \tag{14.288}$$

Now consider a planar spiral on a Bravais lattice, where

$$\theta_i = \boldsymbol{Q}\cdot\boldsymbol{R}_i \qquad , \qquad \phi_i = 0 \quad . \tag{14.289}$$

We then have

$$\hat{\mathsf{e}}_i^- \cdot \hat{\mathsf{e}}_j^+ = 1 + \cos(\theta_i - \theta_j) = 1 + \cos(\boldsymbol{Q}\cdot\boldsymbol{R}_{ij}) \tag{14.290}$$

$$\hat{\mathsf{e}}_i^+ \cdot \hat{\mathsf{e}}_j^- = -1 + \cos(\theta_i - \theta_j) = -1 + \cos(\boldsymbol{Q}\cdot\boldsymbol{R}_{ij}) \tag{14.291}$$

$$\hat{\mathsf{e}}_i^3 \cdot \hat{\mathsf{e}}_j^3 = \cos(\theta_i - \theta_j) = \cos(\boldsymbol{Q}\cdot\boldsymbol{R}_{ij}) \quad , \tag{14.292}$$

where $\boldsymbol{R}_{ij} = \boldsymbol{R}_i - \boldsymbol{R}_j$. Fourier transforming, we arrive at the spin wave Hamiltonian

$$H = \tfrac{1}{2}\sum_{\boldsymbol{k}} \Big\{ \omega_{\boldsymbol{k}} \big(\psi^\dagger_{\boldsymbol{k}} \psi_{\boldsymbol{k}} + \psi_{-\boldsymbol{k}} \psi^\dagger_{-\boldsymbol{k}} \big) + \Delta_{\boldsymbol{k}} \big(\psi^\dagger_{\boldsymbol{k}} \psi^\dagger_{-\boldsymbol{k}} + \psi_{\boldsymbol{k}} \psi_{-\boldsymbol{k}} \big) \Big\} + \tfrac{1}{2} NS\hat{J}(\boldsymbol{Q}) \quad . \tag{14.293}$$

where

$$\omega_{\boldsymbol{k}} = \tfrac{1}{2} S \Big[\hat{J}(\boldsymbol{k}) - 2\hat{J}(\boldsymbol{Q}) + \tfrac{1}{2}\hat{J}(\boldsymbol{k}+\boldsymbol{Q}) + \tfrac{1}{2}\hat{J}(\boldsymbol{k}-\boldsymbol{Q}) \Big] \tag{14.294}$$

$$\Delta_{\boldsymbol{k}} = \tfrac{1}{2} S \Big[-\hat{J}(\boldsymbol{k}) + \tfrac{1}{2}\hat{J}(\boldsymbol{k}+\boldsymbol{Q}) + \tfrac{1}{2}\hat{J}(\boldsymbol{k}-\boldsymbol{Q}) \Big] \quad , \tag{14.295}$$

where

$$\hat{J}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} J(\boldsymbol{R})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{14.296}$$

The Bogoliubov dispersion is then

$$E_{\boldsymbol{k}} = \sqrt{\omega_{\boldsymbol{k}}^2 - \Delta_{\boldsymbol{k}}^2} \tag{14.297}$$

$$= S\sqrt{\big[\hat{J}(\boldsymbol{k}) - \hat{J}(\boldsymbol{Q})\big]\big[\tfrac{1}{2}\hat{J}(\boldsymbol{k}+\boldsymbol{Q}) + \tfrac{1}{2}\hat{J}(\boldsymbol{k}-\boldsymbol{Q}) - \hat{J}(\boldsymbol{Q})\big]} \quad . \tag{14.298}$$

This result agrees with that of P. Locher, *Phys. Rev. B* **41**, 2537 (1990). There are then two possible conditions for zero modes at wavevector $\boldsymbol{k} = \boldsymbol{\kappa}$:

$$\hat{J}(\boldsymbol{\kappa}) = \hat{J}(\boldsymbol{Q}) \qquad \text{or} \qquad \hat{J}(\boldsymbol{\kappa}+\boldsymbol{Q}) + \hat{J}(\boldsymbol{\kappa}-\boldsymbol{Q}) = 2\hat{J}(\boldsymbol{Q}) \quad . \tag{14.299}$$

If one condition is met, then the spin wave dispersion vanishes linearly in $\boldsymbol{k} - \boldsymbol{\kappa}$. If both conditions are met, the spin wave dispersion has a quadratic minimum.

14.10.3 Sublattices

We presume that there is an underlying Bravais lattice, and that the classical ground state is periodic, with a q sublattice structure. The site index i can then be partitioned into a Bravais lattice site $\boldsymbol{R}$ plus a sublattice index $a \in \{1, \ldots, q\}$. We assume that

$$J_{ij} \longrightarrow J_{\boldsymbol{R}a,\boldsymbol{R}'b} = J_{ab}(\boldsymbol{R} - \boldsymbol{R}') = J_{ba}(\boldsymbol{R}' - \boldsymbol{R}) \tag{14.300}$$

depends only on the difference $\boldsymbol{R} - \boldsymbol{R}'$, for each (ab) pair. The lattice Fourier transforms are defined as

$$\psi_a(\boldsymbol{R}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, \psi_{a,\boldsymbol{k}} \qquad , \qquad \psi^\dagger_a(\boldsymbol{R}) = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, \psi^\dagger_{a,\boldsymbol{k}} \tag{14.301}$$

and

$$\hat{J}_{ab}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}\, J_{ab}(\boldsymbol{R}) \quad . \tag{14.302}$$

Note that

$$\hat{J}_{ab}(\boldsymbol{k}) = \hat{J}^*_{ab}(-\boldsymbol{k}) = \hat{J}_{ba}(-\boldsymbol{k}) \quad . \tag{14.303}$$

The spin wave Hamiltonian is then given by

$$\begin{aligned} H_{\rm SW} = \tfrac{1}{4} S \sum_{\boldsymbol{k}} \sum_{a,b=1}^{q} \Big\{ \hat{J}_{ab}(\boldsymbol{k}) \Big[\hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^+_b \, \psi^\dagger_{a,\boldsymbol{k}} \, \psi_{b,\boldsymbol{k}} + \hat{\mathsf{e}}^+_a \cdot \hat{\mathsf{e}}^-_b \, \psi_{a,-\boldsymbol{k}} \, \psi^\dagger_{b,-\boldsymbol{k}} \\ + \hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^-_b \, \psi^\dagger_{a,\boldsymbol{k}} \, \psi^\dagger_{b,-\boldsymbol{k}} + \hat{\mathsf{e}}^+_a \cdot \hat{\mathsf{e}}^+_b \, \psi_{a,-\boldsymbol{k}} \, \psi_{b,\boldsymbol{k}} \Big] \\ - 2 \, \hat{\mathsf{e}}^3_a \cdot \hat{\mathsf{e}}^3_b \, \hat{J}_{ab}(0) \Big[\psi^\dagger_{a,\boldsymbol{k}} \, \psi_{a,\boldsymbol{k}} + \psi_{b,-\boldsymbol{k}} \, \psi^\dagger_{b,-\boldsymbol{k}} - 1 \Big] \Big\} \end{aligned} \tag{14.304}$$

$$\begin{aligned} = \tfrac{1}{4} S \sum_{\boldsymbol{k},a,b} \begin{pmatrix} \psi^\dagger_{a,\boldsymbol{k}} & \psi_{a,-\boldsymbol{k}} \end{pmatrix} \overbrace{\begin{pmatrix} \Omega_{ab}(\boldsymbol{k}) & \Delta_{ab}(\boldsymbol{k}) \\ \Delta^*_{ab}(-\boldsymbol{k}) & \Omega_{ba}(-\boldsymbol{k}) \end{pmatrix}}^{H(\boldsymbol{k})} \begin{pmatrix} \psi_{b,\boldsymbol{k}} \\ \psi^\dagger_{b,-\boldsymbol{k}} \end{pmatrix} \\ + \tfrac{1}{2} S \sum_{\boldsymbol{k},a,b} \hat{J}_{ab}(0) \, \hat{\mathsf{e}}^3_a \cdot \hat{\mathsf{e}}^3_b \quad . \end{aligned} \tag{14.305}$$

Here, the matrices $\Omega(\boldsymbol{k})$ and $\Delta(\boldsymbol{k})$ are given by

$$\Omega_{ab}(\boldsymbol{k}) = \hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^+_b \, \hat{J}_{ab}(\boldsymbol{k}) - 2 \, \delta_{ab} \sum_c \hat{J}_{ac}(0) \, \hat{\mathsf{e}}^3_a \cdot \hat{\mathsf{e}}^3_c \tag{14.306}$$

$$\Delta_{ab}(\boldsymbol{k}) = \hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^-_b \, \hat{J}_{ab}(\boldsymbol{k}) \quad . \tag{14.307}$$

Note that

$$\Omega_{ba}(-\boldsymbol{k}) = \hat{\mathsf{e}}^+_a \cdot \hat{\mathsf{e}}^-_b \, \hat{J}_{ab}(\boldsymbol{k}) - 2 \, \delta_{ab} \sum_c \hat{J}_{ac}(0) \, \hat{\mathsf{e}}^3_a \cdot \hat{\mathsf{e}}^3_c \tag{14.308}$$

$$\Delta^*_{ab}(-\boldsymbol{k}) = \hat{\mathsf{e}}^+_a \cdot \hat{\mathsf{e}}^+_b \, \hat{J}_{ab}(\boldsymbol{k}) \quad . \tag{14.309}$$

We will find it notationally convenient to define separately the dimensionless matrices

$$M_{ab} \equiv \hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^+_b \qquad , \qquad N_{ab} \equiv \hat{\mathsf{e}}^-_a \cdot \hat{\mathsf{e}}^-_b \tag{14.310}$$

and the vector

$$\Lambda_a \equiv -2 \sum_c \hat{J}_{ac}(0) \, \hat{\mathsf{e}}^3_a \cdot \hat{\mathsf{e}}^3_c \quad , \tag{14.311}$$

which has dimensions of energy. Then

$$H(\boldsymbol{k}) = \begin{pmatrix} M_{ab} \, \hat{J}_{ab}(\boldsymbol{k}) + \Lambda_a \, \delta_{ab} & N_{ab} \, \hat{J}_{ab}(\boldsymbol{k}) \\ \\ N^*_{ab} \, \hat{J}_{ab}(\boldsymbol{k}) & M^*_{ab} \, \hat{J}_{ab}(\boldsymbol{k}) + \Lambda_a \, \delta_{ab} \end{pmatrix} \quad . \tag{14.312}$$

14.10.4 Diagonalization

We diagonalize via a generalized Bogoliubov transformation, writing

$$\psi_{a,\boldsymbol{k}} = \sum_l \Big[U_{a,l}(\boldsymbol{k})\,\beta_{l,\boldsymbol{k}} + \widetilde{V}^*_{a,l}(\boldsymbol{k})\,\beta^\dagger_{l,-\boldsymbol{k}} \Big] \tag{14.313}$$

$$\psi^\dagger_{a,-\boldsymbol{k}} = \sum_l \Big[V_{a,l}(\boldsymbol{k})\,\beta_{l,\boldsymbol{k}} + \widetilde{U}^*_{a,l}(\boldsymbol{k})\,\beta^\dagger_{l,-\boldsymbol{k}} \Big] \tag{14.314}$$

Thus, we may write

$$\overbrace{\begin{pmatrix} \psi_{a,\boldsymbol{k}} \\ \psi^\dagger_{a,-\boldsymbol{k}} \end{pmatrix}}^{\Psi(\boldsymbol{k})} = \overbrace{\begin{pmatrix} U & \widetilde{V}^* \\ V & \widetilde{U}^* \end{pmatrix}}^{\mathcal{S}(\boldsymbol{k})} \overbrace{\begin{pmatrix} \beta_{l,\boldsymbol{k}} \\ \beta^\dagger_{l,-\boldsymbol{k}} \end{pmatrix}}^{B(\boldsymbol{k})} \quad . \tag{14.315}$$

The Hamiltonian is

$$H = \tfrac{1}{2} S {\sum_{\boldsymbol{k}}}' \Psi^\dagger_i(\boldsymbol{k})\, H_{ij}(\boldsymbol{k})\, \Psi_j(\boldsymbol{k}) - \tfrac{1}{4} N S \sum_{a=1}^{q} \Lambda_a \quad , \tag{14.316}$$

where the prime on the sum indicates that only one of $(\boldsymbol{k}, -\boldsymbol{k})$ is included, *i.e.* the sum is over precisely one half of the Brillouin zone.

In order to preserve the commutation relations, we must have

$$\Sigma_{ij} = \big[\, \Psi_i \,,\, \Psi^\dagger_k \,\big] = \mathcal{S}_{ik} \big[\, B_k \,,\, B^\dagger_l \,\big]\, \mathcal{S}^\dagger_{lj} = \big(\mathcal{S}\, \Sigma\, \mathcal{S}^\dagger \big)_{ij} \quad , \tag{14.317}$$

where

$$\Sigma = \begin{pmatrix} 1_{q\times q} & 0 \\ 0 & -1_{q\times q} \end{pmatrix} \quad , \tag{14.318}$$

and where we have suppressed the $\boldsymbol{k}$ labels. Thus,

$$\mathcal{S}^\dagger = \Sigma\, \mathcal{S}^{-1}\, \Sigma \quad . \tag{14.319}$$

This pseudounitarity condition on $\mathcal{S}$ requires

$$U^\dagger U - V^\dagger V = 1 \qquad\qquad U U^\dagger - \widetilde{V}^* \widetilde{V}^{\rm t} = 1 \tag{14.320}$$

$$\widetilde{U}^{\rm t} \widetilde{U}^* - \widetilde{V}^{\rm t} \widetilde{V}^* = 1 \qquad\qquad \widetilde{U}^* \widetilde{U}^{\rm t} - V V^\dagger = 1 \tag{14.321}$$

and

$$U^\dagger \widetilde{V}^* - V^\dagger \widetilde{U}^* = 0 \qquad\qquad U V^\dagger - \widetilde{V}^* \widetilde{U}^{\rm t} = 0 \tag{14.322}$$

$$\widetilde{V}^{\rm t} U - \widetilde{U}^{\rm t} V = 0 \qquad\qquad V U^\dagger - \widetilde{U}^* \widetilde{V}^{\rm t} = 0 \quad . \tag{14.323}$$

The matrix $\mathcal{S}$ is chosen so as to diagonalize H:

$$\mathcal{S}_{\boldsymbol{k}}^\dagger \, H_{\boldsymbol{k}} \, \mathcal{S}_{\boldsymbol{k}} = \Sigma \, \mathcal{S}_{\boldsymbol{k}}^{-1} \, \Sigma \, H_{\boldsymbol{k}} \, \mathcal{S}_{\boldsymbol{k}} = \begin{pmatrix} E(\boldsymbol{k}) & 0 \\ 0 & \widetilde{E}(\boldsymbol{k}) \end{pmatrix} \quad , \tag{14.324}$$

where both E and $\widetilde{E}$ are diagonal $q \times q$ matrices. Suppressing the $\boldsymbol{k}$ label, we then have the eigenvalue equations

$$\sum_{b=1}^{q} \Big[M_{ab} \, \hat{J}_{ab} \, U_{bl} + N_{ab} \, \hat{J}_{ab} \, V_{bl} \Big] + \Lambda_a \, U_{al} = +U_{al} \, E_l \tag{14.325}$$

$$\sum_{b=1}^{q} \Big[N_{ab}^* \, \hat{J}_{ab} \, U_{bl} + M_{ab}^* \, \hat{J}_{ab} \, V_{bl} \Big] + \Lambda_a \, V_{al} = -V_{al} \, E_l \tag{14.326}$$

$$\sum_{b=1}^{q} \Big[M_{ab} \, \hat{J}_{ab} \, \widetilde{V}_{bl}^* + N_{ab} \, \hat{J}_{ab} \, \widetilde{U}_{bl}^* \Big] + \Lambda_a \, \widetilde{V}_{al}^* = -\widetilde{V}_{al}^* \, \widetilde{E}_l \tag{14.327}$$

$$\sum_{b=1}^{q} \Big[N_{ab}^* \, \hat{J}_{ab} \, \widetilde{V}_{bl}^* + M_{ab}^* \, \hat{J}_{ab} \, \widetilde{U}_{bl}^* \Big] + \Lambda_a \, \widetilde{U}_{al}^* = +\widetilde{U}_{al}^* \, \widetilde{E}_l \tag{14.328}$$

We next multiply eqn. (14.325) by $U_{al'}^*$ and eqn. (14.326) by $V_{al'}^*$ and sum over a. Then take the complex conjugate of this equation and exchange the indices l and l'. The result is

$$\big(E_l - E_{l'}^*\big) \, \big(U^\dagger U - V^\dagger V\big)_{l'l} = 0 \quad . \tag{14.329}$$

Corresponding manipulations with eqns. (14.327) and (14.328) yield

$$\big(\widetilde{E}_l - \widetilde{E}_{l'}^*\big) \, \big(\widetilde{U}^{\rm t} \widetilde{U}^* - \widetilde{V}^{\rm t} \widetilde{V}^*\big)_{l'l} = 0 \quad . \tag{14.330}$$

Thus, provided the norm $\big(U^\dagger U - V^\dagger V\big)_{ll}$ is finite, the corresponding eigenvalue E_l is real (and similarly for $\widetilde{E}_l$). Also, eigenvectors corresponding to different eigenvalues are orthogonal.

Multiplying eqn. (14.325) by $\widetilde{V}_{al'}$, eqn. (14.326) by $\widetilde{U}_{al'}$, eqn. (14.327) by $U_{al'}^*$, and eqn. (14.328) by $V_{al'}^*$, conjugating, and exchanging l and l' where necessary, we obtain

$$\big(E_l + \widetilde{E}_{l'}^*\big) \, \big(\widetilde{V}^{\rm t} U - \widetilde{U}^{\rm t} V\big)_{l'l} = 0 \quad , \tag{14.331}$$

which is consistent with eqn. (14.322).

Finally, sending $\boldsymbol{k} \to -\boldsymbol{k}$ in eqns. (14.327) and (14.328), followed by conjugation, establishes

$$\widetilde{U}_{bl}(\boldsymbol{k}) = U_{bl}(-\boldsymbol{k}) \quad , \quad \widetilde{V}_{bl}(\boldsymbol{k}) = V_{bl}(-\boldsymbol{k}) \quad , \quad \widetilde{E}_l(\boldsymbol{k}) = E_l(-\boldsymbol{k}) \quad . \tag{14.332}$$

The spin wave Hamiltonian is then

$$H_{\rm sw} = \tfrac{1}{2} S \sum_{\boldsymbol{k}} \sum_{l=1}^{q} \Big[E_l(\boldsymbol{k}) \, \big(\beta_{l,\boldsymbol{k}}^\dagger \, \beta_{l,\boldsymbol{k}} + \tfrac{1}{2} \big) - \tfrac{1}{2} \Lambda_l \Big] \quad . \tag{14.333}$$

The ground state energy is then

$$E^0_{\rm SW} = \tfrac{1}{4} S \sum_{\boldsymbol{k}} \sum_{l=1}^{q} \Big[E_l(\boldsymbol{k}) - \Lambda_l \Big] \quad . \tag{14.334}$$

14.11 Appendix: The Foldy-Wouthuysen Transformation

Let us write

$$\hat{H} = mc^2\,\gamma^0 + c\gamma^0\,\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + V \quad , \tag{14.335}$$

where

$$\boldsymbol{\pi} = \boldsymbol{p} + \tfrac{e}{c}\boldsymbol{A} \tag{14.336}$$

is the dynamical momentum and where the γ^μ are the Dirac matrices,

$$\gamma^0 = \begin{pmatrix} 1_{2\times2} & 0_{2\times2} \\ 0_{2\times2} & -1_{2\times2} \end{pmatrix} \qquad , \qquad \boldsymbol{\gamma} = \begin{pmatrix} 0_{2\times2} & \boldsymbol{\sigma}_{2\times2} \\ -\boldsymbol{\sigma}_{2\times2} & 0_{2\times2} \end{pmatrix} \quad . \tag{14.337}$$

Here $\boldsymbol{\sigma}$ is the vector of Pauli matrices.

The idea behind the FW transformation is to unitarily transform to a different Hilbert space basis such that the coupling in $\hat{H}$ between the upper and lower components of the Dirac spinor vanishes. This may be done systematically as an expansion in inverse powers of the electron mass m. We begin by defining $K \equiv c\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + V$ so that $\hat{H} = mc^2\,\gamma^0 + K$. Note that K is of order m^0. We then write

$$\begin{aligned} \hat{\tilde{H}} &= e^{iS}\,\hat{H}\,e^{-iS} \\ &= \hat{H} + i\big[S,\hat{H}\big] + \frac{(i)^2}{2!}\big[S,[S,\hat{H}]\big] + \ldots \quad , \end{aligned} \tag{14.338}$$

where S itself is written as a power series in $(mc^2)^{-1}$:

$$S = \frac{S_0}{mc^2} + \frac{S_1}{(mc^2)^2} + \ldots \quad . \tag{14.339}$$

The job now is to write $\hat{\tilde{H}}$ as a power series in m^{-1}. The first few terms are easy to find:

$$\hat{\tilde{H}} = mc^2\,\gamma^0 + K + i\big[S_0,\gamma^0\big] + \frac{1}{mc^2}\left(i\big[S_0,K\big] + i\big[S_1,\gamma^0\big] - \tfrac{1}{2}\big[S_0,[S_0,\gamma^0]\big] \right) + \ldots \tag{14.340}$$

We choose the operators S_n so as to cancel, at each order in m^{-1}, the off-diagonal terms in $\hat{\tilde{H}}$ that couple the upper two components of Ψ to the lower two components of Ψ. To order m^0, we then demand

$$c\gamma^0\boldsymbol{\gamma}\cdot\boldsymbol{\pi} + i\big[S_0,\gamma^0\big] = 0 \quad . \tag{14.341}$$

Note that we do not demand that $i\left[S_0, \gamma^0\right]$ completely cancel K – indeed it is impossible to find such an S_0, and one way to see this is to take the trace. The trace of any commutator must vanish, but $\mathrm{Tr}\, K = 4V$, which is in general nonzero. But this is of no concern to us, since we only need cancel the (traceless) *off-diagonal* part of K, which is to say $c\gamma^0 \boldsymbol{\gamma} \cdot \boldsymbol{\pi}$.

To solve for S_0, one can write it in terms of its four 2×2 subblocks, compute the commutator with γ^0, and then impose eqn. 14.341. One then finds $S_0 = -\frac{i}{2} c\, \boldsymbol{\gamma} \cdot \boldsymbol{\pi}$.

STUDENT EXERCISE: Derive the result $S_0 = -\frac{i}{2} c\, \boldsymbol{\gamma} \cdot \boldsymbol{\pi}$.

At the next level, we have to deal with the term in the round brackets in eqn. 14.340. Since we know S_0, we can compute the first and the third terms therein. In general, this will leave us with an off-diagonal term coupling upper and lower components of Ψ. We then choose S_1 so as to cancel this term. This calculation already is tedious, and we haven't even gotten to the spin-orbit interaction term yet, since it is of order m^{-2} – yecch!

14.11.1 Derivation of the spin-orbit interaction

Here's a simpler way to proceed to order m^{-2}. Let a, b be block indices and i, j be indices within each block. Thus, the component Ψ_{ai} is the $i^{\rm th}$ component of the $a^{\rm th}$ block; $\Psi_{a=1,i=2}$ is the lower component of the upper block, *i.e.* the second component of the four-vector Ψ.

Write the Hamiltonian as

$$\hat{H} = mc^2\, \tau^z + c\boldsymbol{\sigma} \cdot \boldsymbol{\pi}\, \tau^x + V(\boldsymbol{r}) \quad , \tag{14.342}$$

where τ^μ are Pauli matrices with indices a, b and σ^ν are Pauli matrices with indices i, j. The σ and τ matrices commute because they act on different indices.

A very important result regarding Pauli matrices:

$$e^{i\theta\, \hat{\boldsymbol{n}} \cdot \boldsymbol{\tau}/2}\, \tau^\alpha\, e^{-i\theta\, \hat{\boldsymbol{n}} \cdot \boldsymbol{\tau}/2} = n^\alpha n^\beta\, \tau^\beta + \cos\theta\, (\delta^{\alpha\beta} - n^\alpha n^\beta)\, \tau^\beta + \sin\theta\, \epsilon^{\alpha\beta\gamma}\, n^\beta\, \tau^\gamma \quad . \tag{14.343}$$

STUDENT EXERCISE: Verify and interpret the above result.

Using this result, we can write

$$A\, \tau^z + B\, \tau^x = \sqrt{A^2 + B^2} \cdot e^{-i \tan^{-1}(B/A)\, \tau^y/2}\, \tau^z\, e^{i \tan^{-1}(B/A)\, \tau^y/2} \quad , \tag{14.344}$$

and, for our specific purposes,

$$mc^2\, \tau^z + c\boldsymbol{\sigma} \cdot \boldsymbol{\pi}\, \tau^x = \sqrt{(mc^2)^2 + (c\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2} \cdot U\, \tau^z\, U^\dagger \quad , \tag{14.345}$$

where

$$U = e^{-i \tan^{-1}\left(\frac{\boldsymbol{\sigma} \cdot \boldsymbol{\pi}}{mc}\right) \tau^y/2} \quad . \tag{14.346}$$

The fact that $\boldsymbol{\sigma}\cdot\boldsymbol{\pi}$ is an operator is no obstacle here, since it commutes with the τ^μ matrices. We can give meaning to expressions like $\tan^{-1}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}/mc)$ in terms of their Taylor series expansions.

We therefore have the result,

$$U^\dagger\,\hat{H}\,U = \sqrt{(mc^2)^2 + (c\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}\cdot\tau^z + U^\dagger\,V(\boldsymbol{r})\,U \quad . \tag{14.347}$$

The first term is diagonal in the block indices. Expanding the square root, we have

$$\begin{aligned} mc^2\sqrt{1+\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right)^2} &= mc^2 + \frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}{2m} + \mathcal{O}(m^{-3}) \\ &= mc^2 + \frac{\boldsymbol{\pi}^2}{2m} + \frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma} + \mathcal{O}(m^{-3}) \quad , \end{aligned} \tag{14.348}$$

since

$$\begin{aligned} (\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2 &= \sigma^\mu\sigma^\nu\,\pi^\mu\pi^\nu \\ &= (\delta^{\mu\nu} + i\epsilon^{\mu\nu\lambda}\sigma^\lambda)\,\pi^\mu\pi^\nu \\ &= \boldsymbol{\pi}^2 + \tfrac{i}{2}\epsilon^{\mu\nu\lambda}\left[p^\mu + \tfrac{e}{c}A^\mu, p^\nu + \tfrac{e}{c}A^\nu\right] \\ &= \boldsymbol{\pi}^2 + \frac{e\hbar}{c}\,\boldsymbol{B}\cdot\boldsymbol{\sigma} \quad . \end{aligned} \tag{14.349}$$

We next need to compute $U^\dagger\,V(\boldsymbol{r})\,U$ to order m^{-2}. To do this, first note that

$$U = 1 - \frac{i}{2}\,\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\,\tau^y - \frac{1}{8}\left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right)^2 + \ldots \quad , \tag{14.350}$$

Thus,

$$U^\dagger\,V\,U = V + \frac{i}{2mc}\big[\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, V\big]\,\tau^y - \frac{1}{8m^2c^2}\big[\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, [\boldsymbol{\sigma}\cdot\boldsymbol{\pi}, V]\big] + \ldots \quad . \tag{14.351}$$

Upon reflection, one realizes that, to this order, it suffices to take the first term in the Taylor expansion of $\tan^{-1}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}/mc)$ in eqn. 14.346, in which case one can then invoke eqn. 14.338 to obtain the above result. The second term on the RHS of eqn. 14.351 is simply $\frac{\hbar}{2mc}\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y$. The third term is

$$\begin{aligned} \frac{i\hbar}{8m^2c^2}\big[\sigma^\mu\pi^\mu, \sigma^\nu\partial^\nu V\big] &= \frac{i\hbar}{8m^2c^2}\Big\{\sigma^\mu\left[\pi^\mu, \sigma^\nu\partial^\nu V\right] + \left[\sigma^\mu, \sigma^\nu\partial^\nu V\right]\pi^\mu\Big\} \\ &= \frac{i\hbar}{8m^2c^2}\left\{\frac{\hbar}{i}\partial^\mu\partial^\nu V\,\sigma^\mu\sigma^\nu + 2i\epsilon^{\mu\nu\lambda}\sigma^\lambda\partial^\nu V\,\pi^\mu\right\} \\ &= \frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2 V + \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi} \quad . \end{aligned} \tag{14.352}$$

Therefore,

$$\begin{aligned} U^\dagger\,\hat{H}\,U = &\left(mc^2 + \frac{\boldsymbol{\pi}^2}{2m} + \frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma}\right)\tau^z + V + \frac{\hbar}{2mc}\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y \\ &+ \frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2 V + \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi} + \mathcal{O}(m^{-3}) \quad . \end{aligned} \tag{14.353}$$

This is not block-diagonal, owing to the last term on the RHS of the top line. We can eliminate this term by effecting yet another unitary transformation. However, this will result in a contribution to the energy of order m^{-3}, so we can neglect it. To substantiate this last claim, drop all the block-diagonal terms except for the leading order one, $mc^2\,\tau^z$, and consider the Hamiltonian

$$\mathcal{K} = mc^2\,\tau^z + \frac{\hbar}{2mc}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\,\tau^y \quad . \tag{14.354}$$

We now know how to bring this to block-diagonal form. The result is

$$\begin{aligned}\tilde{\mathcal{K}} &= mc^2\sqrt{1+\left(\frac{\hbar\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V}{2m^2c^3}\right)^2}\,\tau^z \\ &= \left(mc^2 + \frac{\hbar^2(\boldsymbol{\nabla}\,V)^2}{8m^3c^4} + \ldots\right)\tau^z \quad ,\end{aligned} \tag{14.355}$$

and the correction is of order m^{-3}, as promised.

We now assume all the negative energy ($\tau^z = -1$) states are filled. The Hamiltonian for the electrons, valid to $\mathcal{O}(m^{-3})$, is then

$$\tilde{\hat{H}} = mc^2 + V + \frac{\boldsymbol{\pi}^2}{2m} + \frac{e\hbar}{2mc}\,\boldsymbol{B}\cdot\boldsymbol{\sigma} + \frac{\hbar^2}{8m^2c^2}\,\boldsymbol{\nabla}^2 V + \frac{\hbar}{4m^2c^2}\,\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\,V\times\boldsymbol{\pi} \quad . \tag{14.356}$$

Chapter 15

Spins, Coherent States, Path Integrals, and Applications

15.1 The Coherent State Path Integral

15.1.1 Feynman path integral

The path integral formulation of quantum mechanics is both beautiful and powerful. It is useful in elucidating the quantum-classical correspondence and the semiclassical approximation, in accounting for interference effects, in treatments of tunneling problems via the method of instantons, *etc.* Our goal is to derive and to apply a path integral method for quantum spin. We begin by briefly reviewing the derivation of the usual Feynman path integral.

Consider the propagator $K(x_{\rm i}, x_{\rm f}, T)$, which is the probability amplitude that a particle located at $x = x_{\rm i}$ at time $t = 0$ will be located at $x = x_{\rm f}$ at time $t = T$. We may write

$$\begin{aligned} K(x_{\rm i}, x_{\rm f}, T) &= \big\langle\, x_{\rm f} \,\big|\, e^{-iHT/\hbar} \,\big|\, x_{\rm i} \,\big\rangle \\ &= \big\langle\, x_N \,\big|\, e^{-i\epsilon H/\hbar}\, 1\, e^{-i\epsilon H/\hbar}\, 1 \cdots 1\, e^{-i\epsilon H/\hbar} \,\big|\, x_0 \,\big\rangle \end{aligned} \tag{15.1}$$

where $\epsilon = T/N$, and where we have defined $x_0 \equiv x_{\rm i}$ and $x_N \equiv x_{\rm f}$. We are interested in the limit $N \to \infty$. Inserting $(N-1)$ resolutions of the identity of the form

$$1 = \int\limits_{-\infty}^{\infty} dx_j \,\big|\, x_j \,\big\rangle\big\langle\, x_j \,\big| \quad , \tag{15.2}$$

we find that we must evaluate matrix elements of the form

$$\begin{aligned}\left\langle\, x_{j+1} \,\middle|\, e^{-iH\epsilon/\hbar} \,\middle|\, x_j \,\right\rangle &\approx \int\limits_{-\infty}^{\infty} dp_j \left\langle\, x_{j+1} \,\middle|\, p_j \,\right\rangle \left\langle\, p_j \,\middle|\, e^{-iT\epsilon/\hbar}\, e^{-iV\epsilon/\hbar} \,\middle|\, x_j \,\right\rangle \\ &= \int\limits_{-\infty}^{\infty} dp_j\, e^{ip_j(x_{j+1}-x_j)}\, e^{-i\epsilon p_j^2/2m\hbar}\, e^{-i\epsilon V(x_j)/\hbar} \quad .\end{aligned} \tag{15.3}$$

The propagator may now be written as

$$\begin{aligned}\left\langle\, x_N \,\middle|\, e^{-iHT/\hbar} \,\middle|\, x_0 \,\right\rangle &\approx \int\limits_{-\infty}^{\infty} > \prod_{j=1}^{N-1} dx_j \int\limits_{-\infty}^{\infty} \prod_{k=0}^{N-1} dp_k\, \exp\left\{ i \sum_{k=0}^{N-1} \left[p_k(x_{k+1}-x_k) - \frac{\epsilon}{2m\hbar}\, p_k^2 - \frac{\epsilon}{\hbar}\, V(x_k) \right] \right\} \\ &= \left(\frac{2\pi\hbar m}{i\epsilon} \right)^N \int\limits_{-\infty}^{\infty} \prod_{j=1}^{N-1} dx_j\, \exp\left\{ \frac{i\epsilon}{\hbar} \sum_{k=1}^{N-1} \left[\tfrac{1}{2} m \Big(\frac{x_{j+1}-x_j}{\epsilon} \Big)^2 - V(x_j) \right] \right\} \\ &\equiv \int\limits_{\substack{x(0)=x_{\rm i} \\ x(T)=x_{\rm f}}} \mathcal{D}x(t)\, \exp\left\{ \frac{i}{\hbar} \int\limits_0^T dt \left[\tfrac{1}{2} m\dot{x}^2 - V(x) \right] \right\} \quad ,\end{aligned} \tag{15.4}$$

where we absorb the prefactor into the measure $\mathcal{D}x(t)$. Note the boundary conditions on the path integral at $t = 0$ and $t = T$. In the semiclassical approximation, we assume that the path integral is dominated by trajectories $x(t)$ which extremize the argument of the exponential in the last term above. This quantity is (somewhat incorrectly) identified as the classical action, $\mathcal{S}$, and the action-extremizing equations are of course the Euler-Lagrange equations. Setting $\delta\mathcal{S} = 0$ yields Newton's second law, $m\ddot{x} = -\partial V/\partial x$, which is to be solved subject to the two boundary conditions.

The 'imaginary time' version, which yields the 'thermal propagator', is obtained by writing $T = -i\hbar\beta$ and $t = -i\tau$, in which case

$$\left\langle\, x_{\rm f} \,\middle|\, e^{-\beta H} \,\middle|\, x_{\rm i} \,\right\rangle = \int\limits_{\substack{x(0)=x_{\rm i} \\ x(\hbar\beta)=x_{\rm f}}} \mathcal{D}x(\tau)\, \exp\left\{ -\frac{1}{\hbar} \overbrace{\int\limits_0^{\hbar\beta} d\tau \left[\tfrac{1}{2} m\dot{x}^2 + V(x) \right]}^{\text{'Euclidean action' } \mathcal{S}_{\rm E}} \right\} \quad . \tag{15.5}$$

The partition function is the trace of the thermal propagator, *viz.*

$$Z = \mathrm{Tr}\; e^{-\beta H} = \int\limits_{-\infty}^{\infty} dx \left\langle\, x \,\middle|\, e^{-\beta H} \,\middle|\, x \,\right\rangle = \int\limits_{x(0)=x(\hbar\beta)} \mathcal{D}x(\tau)\, \exp\big(-\mathcal{S}_{\rm E}[x(\tau)]/\hbar \big) \tag{15.6}$$

The equations of motion derived from $\mathcal{S}_{\rm E}$ are $m\ddot{x} = +\partial V/\partial x$, corresponding to motion in the 'inverted potential'.

15.1.2 Coherent state path integral for the 'Heisenberg-Weyl' hroup

We now turn to the method of coherent state path integration. In order to discuss this, we must first introduce the notion of coherent states. This is most simply done by appealing to the one-dimensional simple harmonic oscillator,

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega_0^2 x^2 = \hbar\omega_0\,(a^\dagger a + \tfrac{1}{2}) \quad , \tag{15.7}$$

where a and $a^\dagger$ are ladder operators,

$$a = \ell\,\partial_x + \frac{x}{2\ell} \qquad , \qquad a^\dagger = -\ell\,\partial_x + \frac{x}{2\ell} \tag{15.8}$$

with $\ell \equiv \sqrt{\hbar/2m\omega_0}$. *Exercise: Check that* $[a, a^\dagger] = 1$.

The ground state satisfies $a\,\psi_0(x) = 0$, which yields

$$\psi_0(x) = (2\pi\ell^2)^{-1/4}\,\exp(-x^2/4\ell^2) \quad . \tag{15.9}$$

The normalized coherent state $\big|\, z\,\big\rangle$ is defined as

$$\big|\, z\,\big\rangle = e^{-\frac{1}{2}|z|^2}\,e^{za^\dagger}\big|\, 0\,\big\rangle = e^{-\frac{1}{2}|z|^2}\sum_{n=0}^{\infty}\frac{z^n}{\sqrt{n!}}\big|\, n\,\big\rangle \quad . \tag{15.10}$$

The coherent state is an eigenstate of the annihilation operator a:

$$a\big|\, z\,\big\rangle = z\big|\, z\,\big\rangle \qquad \Longleftrightarrow \qquad \big\langle\, z\,\big|\, a^\dagger = \big\langle\, z\,\big|\, \bar{z} \quad . \tag{15.11}$$

The overlap of coherent states is given by

$$\big\langle\, z_1\,\big|\, z_2\,\big\rangle = e^{-\frac{1}{2}|z_1|^2}\,e^{-\frac{1}{2}|z_2|^2}\,e^{\bar{z}_1 z_2} \quad , \tag{15.12}$$

hence different coherent states are not orthogonal. Despite this nonorthogonality, the coherent states allow a simple resolution of the identity,

$$1 = \int\frac{d^2z}{2\pi i}\big|\, z\,\big\rangle\big\langle\, z\,\big| \qquad , \qquad \frac{d^2z}{2\pi i} \equiv \frac{d\,\mathrm{Re}\,z\,d\,\mathrm{Im}\,z}{\pi} \tag{15.13}$$

which is straightforward to establish.

To gain some physical intuition about the coherent states, define

$$z \equiv \frac{Q}{2\ell} + \frac{i\ell P}{\hbar} \quad . \tag{15.14}$$

One finds (*exercise!*)

$$\psi_{P,Q}(x) = \left\langle\, x \,\middle|\, z \,\right\rangle = (2\pi\ell^2)^{-1/4}\, e^{-iPQ/2\hbar}\, e^{iPx/\hbar}\, e^{-(x-Q)^2/4\ell^2} \quad , \tag{15.15}$$

hence the coherent state $\psi_{P,Q}(x)$ is a wavepacket Gaussianly localized about $x = Q$, but oscillating with momentum P. *Exercise: Compute* $\langle (q-Q)^2 \rangle$ *and* $\langle (p-P)^2 \rangle$.

Now we derive the imaginary time path integral. We write

$$\left\langle\, z_{\rm f} \,\middle|\, e^{-\beta H} \,\middle|\, z_{\rm i} \,\right\rangle = \left\langle\, z_N \,\middle|\, e^{-\epsilon H/\hbar}\, 1\, e^{-\epsilon H/\hbar} \cdots 1\, e^{-\epsilon H/\hbar} \,\middle|\, z_0 \,\right\rangle \quad , \tag{15.16}$$

inserting resolutions of the identity at $N-1$ points, as before. We next evaluate the matrix element

$$\begin{aligned} \left\langle\, z_j \,\middle|\, e^{-\epsilon H/\hbar} \,\middle|\, z_{j-1} \,\right\rangle &= \left\langle\, z_j \,\middle|\, z_{j-1} \,\right\rangle \cdot \left\{ 1 - \frac{\epsilon}{\hbar} \frac{\left\langle\, z_j \,\middle|\, H \,\middle|\, z_{j-1} \,\right\rangle}{\left\langle\, z_j \,\middle|\, z_{j-1} \,\right\rangle} + \ldots \right\} \\ &\simeq \left\langle\, z_j \,\middle|\, z_{j-1} \,\right\rangle \exp\left\{ -\frac{\epsilon}{\hbar} H(\bar{z}_j | z_{j-1}) \right\} \end{aligned} \tag{15.17}$$

where

$$H(\bar{z}|w) \equiv \frac{\left\langle\, z \,\middle|\, H \,\middle|\, w \,\right\rangle}{\left\langle\, z \,\middle|\, w \,\right\rangle} = e^{-\bar{z}w} \left\langle\, 0 \,\middle|\, e^{\bar{z}a}\, H(a^\dagger, a)\, e^{wa^\dagger} \,\middle|\, 0 \,\right\rangle \quad . \tag{15.18}$$

This last equation is extremely handy. It says, upon invoking eqn. 15.11, that if $H(a, a^\dagger)$ is *normal ordered* such that all creation operators $a^\dagger$ appear to the *left* of all destruction operators a, then $H(\bar{z}|w)$ is obtained from $H(a^\dagger, a)$ simply by sending $a^\dagger \to \bar{z}$ and $a \to w$. This is because a acting to the right on $\left|\, w \,\right\rangle$ yields its eigenvalue w, while $a^\dagger$ acting to the left on $\left\langle\, z \,\right|$ generates $\bar{z}$. Note that the function $H(\bar{z}|w)$ is holomorphic in w and in $\bar{z}$, but is completely independent of their complex conjugates $\bar{w}$ and z.

The overlap between coherent states at consecutive time slices may be written

$$\left\langle\, z_j \,\middle|\, z_{j-1} \,\right\rangle = \exp\left\{ -\tfrac{1}{2} \Big[\bar{z}_j (z_j - z_{j-1}) - z_{j-1} (\bar{z}_j - \bar{z}_{j-1}) \Big] \right\} \quad , \tag{15.19}$$

hence

$$\begin{aligned} \left\langle\, z_N \,\middle|\, z_{N-1} \,\right\rangle \cdots \left\langle\, z_1 \,\middle|\, z_0 \,\right\rangle = \exp\left\{ \tfrac{1}{2} \sum_{j=1}^{N-1} \Big[z_j (\bar{z}_{j+1} - \bar{z}_j) - \bar{z}_j (z_j - z_{j-1}) \Big] \right\} \\ \times \exp\left\{ \tfrac{1}{2}\, z_0 (\bar{z}_1 - \bar{z}_0) - \tfrac{1}{2}\, \bar{z}_N (z_N - z_{N-1}) \right\} \quad , \end{aligned} \tag{15.20}$$

which allows us to write down the path integral expression for the propagator,

$$\begin{aligned}\big\langle\, z_{\rm f}\,\big|\, e^{-\beta H}\,\big|\, z_{\rm i}\,\big\rangle &= \int \prod_{j=1}^{N-1} \frac{d^2 z_j}{2\pi i}\, \exp\Big(-\mathcal{S}_{\rm E}[\{z_j,\bar z_j\}]/\hbar\Big) \\ \mathcal{S}_{\rm E}[\{z_j,\bar z_j\}]/\hbar &= \sum_{j=1}^{N-1}\Big[\tfrac{1}{2}\,\bar z_j(z_j - z_{j-1}) - \tfrac{1}{2}\,z_j(\bar z_{j+1}-\bar z_j)\Big] + \frac{\epsilon}{\hbar}\sum_{j=1}^{N} H(\bar z_j|z_{j-1}) \\ &\qquad + \tfrac{1}{2}\,\bar z_{\rm f}\big(z_{\rm f}-z_{N-1}\big) - \tfrac{1}{2}\,z_{\rm i}\big(\bar z_1 - \bar z_{\rm i}\big) \quad .\end{aligned} \tag{15.21}$$

In the limit $N\to\infty$, we identify the continuum Euclidean action

$$\begin{aligned}\mathcal{S}_{\rm E}[\{z(\tau),\bar z(\tau)\}]/\hbar = \int_0^{\hbar\beta} d\tau\,\left\{\frac{1}{2}\left(\bar z\,\frac{\partial z}{\partial\tau} - z\,\frac{\partial\bar z}{\partial\tau}\right) + \frac{1}{\hbar}\,H(\bar z|z)\right\} \\ + \tfrac{1}{2}\,\bar z_{\rm f}\big[z_{\rm f}-z(\hbar\beta)\big] - \tfrac{1}{2}\,z_{\rm i}\big[\bar z(0)-\bar z_{\rm i}\big] \quad ,\end{aligned} \tag{15.22}$$

and write the continuum expression for the path integral,

$$\big\langle\, z_{\rm f}\,\big|\, e^{-\beta H}\,\big|\, z_{\rm i}\,\big\rangle = \int \mathcal{D}[z(\tau),\bar z(\tau)]\, e^{-\mathcal{S}_{\rm E}[\{z(\tau),\bar z(\tau)\}]/\hbar} \quad . \tag{15.23}$$

The continuum limit is in a sense justified by examining the discrete equations of motion,

$$\begin{aligned}\frac{1}{\hbar}\frac{\partial\mathcal{S}_{\rm E}}{\partial z_k} &= \bar z_k - \bar z_{k+1} + \frac{\epsilon}{\hbar}\frac{\partial H(\bar z_{k+1}|z_k)}{\partial z_k} \\ \frac{1}{\hbar}\frac{\partial\mathcal{S}_{\rm E}}{\partial \bar z_k} &= z_k - z_{k-1} + \frac{\epsilon}{\hbar}\frac{\partial H(\bar z_k|z_{k-1})}{\partial \bar z_k} \quad ,\end{aligned} \tag{15.24}$$

which have the sensible continuum limit

$$\hbar\frac{\partial\bar z}{\partial\tau} = \frac{\partial H(\bar z|z)}{\partial z} \qquad , \qquad \hbar\frac{\partial z}{\partial\tau} = -\frac{\partial H(\bar z|z)}{\partial\bar z} \tag{15.25}$$

with boundary conditions $\bar z(\hbar\beta)=\bar z_{\rm f}$ and $z(0)=z_{\rm i}$. Note that there are only two boundary conditions – one on $z(0)$ and the other on $\bar z(\hbar\beta)$. The function $z(\tau)$ (or its discrete version z_j) is evolved forward from initial data $z_{\rm i}$, while $\bar z(\tau)$ (or $\bar z_j$) is evolved backward from final data $\bar z_{\rm f}$. This is the proper number of boundary conditions to place on two first order differential (or finite difference) equations. It is noteworthy that the action of eqn. 15.21 or eqn. 15.22 imposes only a *finite* penalty on *discontinuous* paths.[1] Nevertheless, the paths which extremize the action are continuous throughout the interval $\tau\in(0,\hbar\beta)$. As $z(\tau)$ is integrated forward from $z_{\rm i}$, its

[1] In the Feynman path integral, discontinuous paths contribute an infinite amount to the action, and are therefore suppressed.

final value $z(\hbar\beta)$ will in general be different from $z_{\rm f}$. Similarly, $\bar z(\tau)$ integrated backward from $\bar z_{\rm f}$ will in general yield an endpoint value $\bar z(0)$ which differs from $\bar z_{\rm i}$. The differences $z(\hbar\beta) - z_{\rm f}$ and $\bar z(0) - \bar z_{\rm i}$ are often identified as path discontinuities, but the fact is that the equations of motion know nothing about either $z_{\rm f}$ or $\bar z_{\rm i}$. These difference terms do enter in a careful accounting of the action formulae of eqns. 15.21 and 15.22, however.

The importance of the boundary terms is nicely illustrated in a computation of the semiclassical imaginary time propagator for the harmonic oscillator. With $H = \hbar\omega_0\, a^\dagger a$ (dropping the constant term for convenience), we have

$$\begin{aligned}\big\langle\, z_{\rm f}\,\big|\,\exp(-\beta\hbar\omega_0\, a^\dagger a)\,\big|\,z_{\rm i}\,\big\rangle &= e^{-\frac{1}{2}|z_{\rm f}|^2-\frac{1}{2}|z_{\rm i}|^2}\sum_{m,n=0}^{\infty}\frac{\bar z_{\rm f}^{\,m}\,z_{\rm i}^{\,n}}{\sqrt{m!\,n!}}\,\big\langle\, m\,\big|\,\exp(-\beta\hbar\omega_0\, a^\dagger a)\,\big|\,n\,\big\rangle \\ &= \exp\left\{-\tfrac{1}{2}|z_f|^2-\tfrac{1}{2}|z_i|^2+\bar z_f z_i\, e^{-\beta\hbar\omega_0}\right\}\quad .\end{aligned} \tag{15.26}$$

The Euclidean action is $L_{\rm E} = \frac{1}{2}\hbar(\bar z\dot z - z\dot{\bar z}) + \hbar\omega_0\,\bar z z$, so the equations of motion are

$$\hbar\dot{\bar z} = \frac{\partial H}{\partial z} = \hbar\omega_0\,\bar z \qquad , \qquad \hbar\dot z = -\frac{\partial H}{\partial \bar z} = -\hbar\omega_0\, z \tag{15.27}$$

subject to boundary conditions $z(0) = z_{\rm i}$, $\bar z(\hbar\beta) = \bar z_{\rm f}$. The solution is

$$z(\tau) = z_{\rm i}\, e^{-\omega_0\tau} \qquad , \qquad \bar z(\tau) = \bar z_{\rm f}\, e^{\omega_0(\tau-\hbar\beta)} \quad . \tag{15.28}$$

Along the 'classical path' the Euclidean Lagrangian vanishes: $L_{\rm E} = 0$. The entire contribution to the action therefore comes from the boundary terms:

$$\begin{aligned}\mathcal{S}_{\rm E}^{\rm cl}/\hbar &= 0 + \tfrac{1}{2}\bar z_{\rm f}(z_{\rm f} - z_{\rm i}\, e^{-\beta\hbar\omega_0}) - \tfrac{1}{2}z_i(\bar z_f\, e^{-\beta\hbar\omega_0} - \bar z_i) \\ &= \tfrac{1}{2}|z_{\rm f}|^2 + \tfrac{1}{2}|z_{\rm i}|^2 - \bar z_{\rm f} z_{\rm i}\, e^{-\beta\hbar\omega_0} \quad ,\end{aligned} \tag{15.29}$$

What remains is to compute the fluctuation determinant. We write

$$z_j = z_j^{\rm cl} + \eta_j \qquad , \qquad \bar z_j = \bar z_j^{\rm cl} + \bar\eta_j \quad , \tag{15.30}$$

and expand the action as

$$\begin{aligned}\mathcal{S}_{\rm E}[\{z_j,\bar z_j\}] &= \mathcal{S}_{\rm E}[\{z_j^{\rm cl},\bar z_j^{\rm cl}\}] + \frac{\partial^2\mathcal{S}_{\rm E}}{\partial\bar z_i\partial z_j}\,\bar\eta_i\eta_j + \frac{1}{2}\,\frac{\partial^2\mathcal{S}_{\rm E}}{\partial z_i\partial z_j}\,\eta_i\eta_j + \frac{1}{2}\,\frac{\partial^2\mathcal{S}_{\rm E}}{\partial\bar z_i\partial\bar z_j}\,\bar\eta_i\bar\eta_j + \ldots \\ &\equiv \mathcal{S}_{\rm E}^{\rm cl} + \frac{\hbar}{2}\begin{pmatrix}\bar z_i & z_i\end{pmatrix}\begin{pmatrix}A_{ij} & B_{ij}\\ C_{ij} & A^{\rm t}_{ij}\end{pmatrix}\begin{pmatrix}z_j\\ \bar z_j\end{pmatrix} + \ldots \quad .\end{aligned} \tag{15.31}$$

For general H, we obtain

$$\begin{aligned} A_{ij} &= \delta_{ij} - \delta_{i,j+1} + \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_i|z_j)}{\partial\bar{z}_i\,\partial z_j}\,\delta_{i,j+1} \\ B_{ij} &= \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_i|z_{i-1})}{\partial\bar{z}_i^2}\,\delta_{i,j} \\ C_{ij} &= \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_{i+1}|z_i)}{\partial z_i^2}\,\delta_{i,j} \quad . \end{aligned} \tag{15.32}$$

with i and j running from 1 to $N-1$. The contribution of the fluctuation determinant to the matrix element is then

$$\int\prod_{j=1}^{N-1}\frac{d^2\eta_i}{2\pi i}\,\exp\left\{-\frac{1}{2}\begin{pmatrix}\mathsf{Re}\;\eta_k & \mathsf{Im}\;\eta_k\end{pmatrix}\begin{pmatrix}1 & 1\\ -i & i\end{pmatrix}\begin{pmatrix}A_{kl} & B_{kl}\\ C_{kl} & A_{lk}\end{pmatrix}\begin{pmatrix}1 & i\\ 1 & -i\end{pmatrix}\begin{pmatrix}\mathsf{Re}\;\eta_l\\ \mathsf{Im}\;\eta_l\end{pmatrix}\right\} = \mathsf{det}^{-1/2}\begin{pmatrix}A & B\\ C & A^{\rm t}\end{pmatrix} \quad . \tag{15.33}$$

In the case of the harmonic oscillator discussed above, we have $B_{ij} = C_{ij} = 0$, and since A_{ij} has no elements above its diagonal and $A_{ii} = 1$ for all i, we simply have that the determinant contribution is unity.

15.2 Coherent States for Spin

For the pros: A. Perelomov, *Generalized Coherent States and their Applications* (Springer-Verlag, NY, 1986).

A spin-coherent state $\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle$ is simply a rotation of the 'highest weight' state $\big|\,m=+S\,\big\rangle$, such that the spin is maximally polarized along $\hat{\boldsymbol{\Omega}}$, *i.e.*

$$\hat{\boldsymbol{\Omega}}\cdot\boldsymbol{S}\,\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle = +S\,\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle \quad . \tag{15.34}$$

Note that $\big|\,m=+S\,\big\rangle$ is itself a coherent state with $\hat{\boldsymbol{\Omega}}=\hat{\boldsymbol{z}}$. We can effect this rotation by means of an element $\mathcal{R}$ of the group SU(2):

$$\begin{aligned} \mathcal{R} &\equiv \exp(i\psi S^z)\,\exp(i\theta S^y)\,\exp(i\phi S^z) \\ \big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle &= \mathcal{R}^\dagger\,\big|\,\hat{\boldsymbol{z}}\,\big\rangle \quad . \end{aligned} \tag{15.35}$$

To define and manipulate the spin coherent states, it is useful to introduce the Schwinger representation of quantum spin. You are probably already familiar with the Holstein-Primakoff transformation,

$$S^+ = h^\dagger\,(2S - h^\dagger h)^{1/2} \qquad , \qquad S^- = (2S-h^\dagger h)^{1/2}\,h \qquad , \qquad S^z = h^\dagger h - S \quad , \tag{15.36}$$

with $0 \leqslant h^\dagger h \leqslant 2S$. by which a quantum spin can be represented by a single boson. Note that the eigenvalues of the boson number operator $n_h = h^\dagger h$ range over the nonnegative integers. There are thus an infinite number of allowed states, but only a finite number $(2S+1)$ of states in the Hilbert space for spin. But the factor $\sqrt{2S - h^\dagger h}$ in S^+ annihilates the state of maximal polarization, $\big|\, m = +S \,\big\rangle$, and thus for any Hamiltonian which can be written in terms of the spin algebra operators, the infinite-dimensional boson Hilbert space is effectively divided into two parts. The 'physical' states all have $0 \leqslant n_h \leqslant 2S$, and there are no matrix elements connecting this subspace to the 'unphysical' one where $n_h > 2S$.

The square roots are unwieldy, however, and in practice one expands them in powers of $(n_h/2S)$, *viz.*

$$(2S - h^\dagger h)^{1/2} = \sqrt{2S} \cdot \left\{ 1 - \frac{1}{2}\Big(\frac{h^\dagger h}{2S}\Big) + \frac{1}{8}\Big(\frac{h^\dagger h}{2S}\Big)^2 + \dots \right\} \quad . \tag{15.37}$$

This expansion forms the basis of spin wave theory. Hence, within spin wave theory, unphysical states are allowed. For example, an interaction like $S_i^+ \, S_j^-$ between spins on sites i and j takes the form $2S \, h_i^\dagger \, h_j$ within the spin wave expansion. But such a term knows nothing of the border lying at $n_h = 2S$ separating physical from unphysical states.

In the Schwinger representation, two bosons are used, and the constraint is a holonomic one (*i.e.* one which can be written as an equality):

$$S^+ = a^\dagger \, b \qquad , \qquad S^- = a \, b^\dagger \qquad , \qquad S^z = \tfrac{1}{2}(a^\dagger a - b^\dagger b) \quad , \tag{15.38}$$

and subject to the constraint $a^\dagger a + b^\dagger b = 2S$. The constraint simply says $n_a + n_b = 2S$, *i.e.* there are a total of $2S$ bosons present. Note that the operators $S^\pm$ change the number of a and b bosons, but preserve the total $n_a + n_b$, hence they *commute with the constraint.*

Exercise: Verify that $[S^+, S^-] = 2S^z$ *and* $[S^z, S^\pm] = \pm S^\pm$ *for both the Holstein-Primakoff and Schwinger representations.*

A shorthand way of rendering the spin operators in the Schwinger representation is

$$\boldsymbol{S} = \tfrac{1}{2} \begin{pmatrix} a^\dagger & b^\dagger \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad . \tag{15.39}$$

We now investigate the action of the SU(2) rotation $\mathcal{R}$ on the Schwinger bosons. We wish to evaluate the expression

$$\mathcal{R}^\dagger \begin{pmatrix} a \\ b \end{pmatrix} \mathcal{R} = e^{-i\phi S^z} \, e^{-i\theta S^y} \, e^{-i\psi S^z} \begin{pmatrix} a \\ b \end{pmatrix} e^{i\psi S^z} \, e^{i\theta S^y} \, e^{i\phi S^z} \quad . \tag{15.40}$$

Let's work this out:

- Rotation about the $\hat{\boldsymbol{z}}$-axis:

$$e^{-i\psi S^z} \begin{pmatrix} a \\ b \end{pmatrix} e^{i\psi S^z} = e^{-i\frac{\psi}{2} a^\dagger a} e^{i\frac{\psi}{2} b^\dagger b} \begin{pmatrix} a \\ b \end{pmatrix} e^{-i\frac{\psi}{2} b^\dagger b} e^{i\frac{\psi}{2} a^\dagger a} = \begin{pmatrix} e^{+i\frac{\psi}{2}} \, a \\ e^{-i\frac{\psi}{2}} \, b \end{pmatrix} \quad . \tag{15.41}$$

- Rotation about the $\hat{\boldsymbol{y}}$ axis:

$$e^{-i\theta S^y}\begin{pmatrix}a\\b\end{pmatrix}e^{i\theta S^y}=e^{\frac{\theta}{2}(ab^\dagger-a^\dagger b)}\begin{pmatrix}a\\b\end{pmatrix}e^{-\frac{\theta}{2}(ab^\dagger-a^\dagger b)}=\begin{pmatrix}\cos(\theta/2)\,a+\sin(\theta/2)\,b\\-\sin(\theta/2)\,a+\cos(\theta/2)\,b\end{pmatrix}\quad . \tag{15.42}$$

We are also licensed to make an additional rotation $\mathcal{U}=\exp(i\xi S)$, where $S=\frac{1}{2}(a^\dagger a+b^\dagger b)$. The final result of the combined transformation $\mathcal{R}\mathcal{U}$ is

$$\begin{pmatrix}\tilde a\\ \tilde b\end{pmatrix}\equiv\mathcal{U}^\dagger\,\mathcal{R}^\dagger\begin{pmatrix}a\\b\end{pmatrix}\mathcal{R}\mathcal{U}=e^{i\xi/2}\begin{pmatrix}\bar u & v\\ -\bar v & u\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix} \tag{15.43}$$

where u and v are spinor coordinates,

$$\begin{aligned}u&=e^{-i\psi/2}\,e^{-i\phi/2}\,\cos\left(\tfrac{1}{2}\theta\right)\\ v&=e^{-i\psi/2}\,e^{+i\phi/2}\,\sin\left(\tfrac{1}{2}\theta\right)\quad .\end{aligned} \tag{15.44}$$

The phase ξ is unphysical, and without loss of generality we are free to define $\xi\equiv-(\phi+\psi)$, in which case

$$\begin{pmatrix}\tilde a\\ \tilde b\end{pmatrix}=\begin{pmatrix}\cos(\frac{1}{2}\theta) & \sin(\frac{1}{2}\theta)\,e^{i\psi}\\ -\sin(\frac{1}{2}\theta)\,e^{i\phi} & \cos(\frac{1}{2}\theta)\,e^{-i(\phi+\psi)}\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}\quad . \tag{15.45}$$

Now that we know how the Schwinger bosons themselves transform under SU(2), we investigate the transformation properties of the spin operators S^α, which are bilinear in the Schwinger bosons. We find

$$\begin{aligned}\mathcal{R}^\dagger\,S^z\,\mathcal{R}&=\begin{pmatrix}a^\dagger & b^\dagger\end{pmatrix}\begin{pmatrix}u & -\bar v\\ v & \bar u\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\bar u & \bar v\\ -v & u\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}\\ &=\tfrac{1}{2}\begin{pmatrix}a^\dagger & b^\dagger\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\,e^{-i\phi}\\ \sin\theta\,e^{i\phi} & -\cos\theta\end{pmatrix}\begin{pmatrix}a\\b\end{pmatrix}\\ &=\sin\theta\cos\phi\,S^x+\sin\theta\sin\phi\,S^z+\cos\theta\,S^z\quad ,\end{aligned}$$

hence $\mathcal{R}^\dagger S^z\mathcal{R}=\hat{\boldsymbol{\Omega}}\cdot\boldsymbol{S}$, and

$$S\,\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle=\mathcal{R}^\dagger S^z\,\big|\,\hat{\boldsymbol{z}}\,\big\rangle=(\mathcal{R}^\dagger S^z\mathcal{R})\,\mathcal{R}^\dagger\big|\,\hat{\boldsymbol{z}}\,\big\rangle=\hat{\boldsymbol{\Omega}}\cdot\boldsymbol{S}\,\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle\quad . \tag{15.46}$$

Explicitly, then,

$$\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle=\big[(2S)!\big]^{-1/2}(ua^\dagger+vb^\dagger)^{2S}\big|\,0\,\big\rangle=\sum_{k=0}^{2S}\binom{2S}{k}^{1/2}u^k\,v^{2S-k}\,\big|\,k-S\,\big\rangle\quad . \tag{15.47}$$

Example: $S=\frac{1}{2}$, $\theta=\frac{1}{2}\pi$, $\phi=\frac{1}{2}\pi$ gives $\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle=\frac{1}{\sqrt{2}}|\uparrow\rangle+\frac{i}{\sqrt{2}}|\downarrow\rangle=\big|\,\hat{\boldsymbol{y}}\,\big\rangle$.

A useful property of the coherent states: if

$$\big|\,\psi\,\big\rangle = f(a^\dagger, b^\dagger)\big|\,0\,\big\rangle \equiv \sum_{k=0}^{2S} f_k\,(a^\dagger)^k\,(b^\dagger)^{2S-k}\,\big|\,0\,\big\rangle \quad , \tag{15.48}$$

then

$$\big\langle\,\hat{\boldsymbol{\Omega}}\,\big|\,\psi\,\big\rangle = \sqrt{(2S)!}\,f(\bar{u},\bar{v}) \quad , \tag{15.49}$$

i.e. replace $a^\dagger \to \bar{u}$ and $b^\dagger \to \bar{v}$ as arguments of f. The overlap of the coherent states is

$$\begin{aligned}\big\langle\,\hat{\boldsymbol{\Omega}}\,\big|\,\hat{\boldsymbol{\Omega}}'\,\big\rangle &= (\bar{u}u' + \bar{v}v')^{2S} \\ &= \left[\tfrac{1}{2}\,(1 + \hat{\boldsymbol{\Omega}}\cdot\hat{\boldsymbol{\Omega}}')\right]^S e^{iS\gamma(\hat{\boldsymbol{\Omega}},\hat{\boldsymbol{\Omega}}')}\end{aligned} \tag{15.50}$$

where

$$\begin{aligned}\gamma(\hat{\boldsymbol{\Omega}},\hat{\boldsymbol{\Omega}}') &\equiv 2\,\mathrm{arg}\,(\bar{u}u' + \bar{v}v') \\ &= 2\tan^{-1}\left[\frac{\sin\frac{1}{2}\theta\,\sin\frac{1}{2}\theta'\,\sin(\phi'-\phi)}{\cos\frac{1}{2}\theta\,\cos\frac{1}{2}\theta' + \sin\frac{1}{2}\theta\,\sin\frac{1}{2}\theta'\,\cos(\phi'-\phi)}\right]\end{aligned} \tag{15.51}$$

Perhaps the most important result, for our purposes, is the resolution of the identity:

$$1 = \frac{2S+1}{4\pi}\int d\Omega\,\big|\,\hat{\boldsymbol{\Omega}}\,\big\rangle\big\langle\,\hat{\boldsymbol{\Omega}}\,\big| \quad . \tag{15.52}$$

As with the case of coherent states for the harmonic oscillator, the spin coherent states permit a simple resolution of the identity despite their nonorthogonality.

The last step, before we tackle the derivation of the spin path integral, is to compute matrix elements in the coherent state basis. We assume that the Hamiltonian commutes with the constraint, *i.e.* it preserves total spin. The most general such Hamiltonian may be written

$$H = \sum_{m,n,j} \mathcal{C}_{mnj}\,(a^\dagger)^m\,(b^\dagger)^n\,(a)^{m+j}\,(b)^{n-j} \quad , \tag{15.53}$$

and its matrix elements may be evaluated using

$$\big\langle\,\hat{\boldsymbol{\Omega}}_1\,\big|\,\overbrace{(a^\dagger)^m\,(b^\dagger)^n\,(a)^{m+j}\,(b)^{n-j}}^{\text{preserves total S}}\,\big|\,\hat{\boldsymbol{\Omega}}_2\,\big\rangle = \frac{(2S)!}{(2S-m-n)!}\,(\bar{u}_1u_2 + \bar{v}_1v_2)^{2S-m-n}\,\bar{u}_1^m\,\bar{v}_1^n\,u_2^{m+j}\,v_2^{n-j} \quad . \tag{15.54}$$

Note that the above operator product must be *normal-ordered*, with annihilation operators a, b appearing to the right of creation operators $a^\dagger$, $b^\dagger$.

Exercise: Verify eqn. 15.54 by finding the $\mathcal{O}\big(\bar{z}_1^{2S} z_2^{2S}\big)$ term of the matrix element in the (unnormalized) generalized coherent state

$$\big|\,z,\hat{\boldsymbol{\Omega}}\,\big\rangle \equiv e^{zua^\dagger}\,e^{zvb^\dagger}\big|\,0\,\big\rangle \quad , \tag{15.55}$$

where z is a complex number. Show that $a \,|\, z, \hat{\boldsymbol{\Omega}} \,\rangle = zu \,\big|\, z, \hat{\boldsymbol{\Omega}} \,\big\rangle$, $b \,\big|\, z, \hat{\boldsymbol{\Omega}} \,\big\rangle = zv \,|\, z, \hat{\boldsymbol{\Omega}} \,\rangle$, *and*

$$\big\langle\, z, \hat{\boldsymbol{\Omega}} \,\big|\, z', \hat{\boldsymbol{\Omega}}' \,\big\rangle = \exp\left[\bar{z}z'(\bar{u}u' + \bar{v}v')\right] \quad . \tag{15.56}$$

Use these results to verify eqn. 15.54.

As with the case of the coherent state path integral for the Heisenberg-Weyl group, only diagonal matrix elements are needed. In this case the expression eqn. 15.54 simplifies to

$$\big\langle\, \hat{\boldsymbol{\Omega}} \,\big|\, (a^\dagger)^m (b^\dagger)^n (a)^{m+j} (b)^{n-j} \,\big|\, \hat{\boldsymbol{\Omega}} \,\big\rangle = \frac{(2S)!}{(2S-m-n)!}\, \bar{u}^m\, \bar{v}^n\, u^{m+j}\, v^{n-j} \quad . \tag{15.57}$$

Two examples of matrix element computation:

- $\mathcal{O} = S^+ = a^\dagger b$. Here, $(m, n, j) = (1, 0, -1)$, so

$$\big\langle\, \hat{\boldsymbol{\Omega}} \,\big|\, a^\dagger b \,\big|\, \hat{\boldsymbol{\Omega}} \,\big\rangle = 2S\, \bar{u}v = S \sin\theta\, e^{i\phi} \quad . \tag{15.58}$$

- $\mathcal{O} = (S_x)^2$. First we normal order:

$$\begin{aligned} S_x^2 &= \left(\frac{a^\dagger b + a b^\dagger}{2}\right)^2 \\ &= \tfrac{1}{4}\Big(a^\dagger b\, a^\dagger b + a\, b^\dagger a\, b^\dagger + a^\dagger b\, b^\dagger a + a\, b^\dagger a^\dagger b\Big) \\ &= \tfrac{1}{4}\Big(\overbrace{a^\dagger a^\dagger\, b\, b}^{(2,0,-2)} + \overbrace{b^\dagger b^\dagger a\, a}^{(0,2,2)} + \overbrace{2\, a^\dagger b^\dagger a\, b}^{(1,1,0)} + \overbrace{a^\dagger a}^{(1,0,0)} + \overbrace{b^\dagger b}^{(0,1,0)}\Big) \quad , \end{aligned} \tag{15.59}$$

  which, following the rules in Eqn. 15.57, yields

$$\begin{aligned} \big\langle\, \hat{\boldsymbol{\Omega}} \,\big|\, S_z^2 \,\big|\, \hat{\boldsymbol{\Omega}} \,\big\rangle &= \tfrac{1}{4}(2S)(2S-1)(\bar{u}^2 v^2 + \bar{v}^2 u^2 + 2\bar{u}\bar{v}uv) + \tfrac{1}{4}(2S)(\bar{u}u + \bar{v}v) \\ &= S(S - \tfrac{1}{2})(\sin\theta\cos\phi)^2 + \tfrac{1}{2}S \quad . \end{aligned} \tag{15.60}$$

Exercise: Prove that

$$\big\langle\, \hat{\boldsymbol{\Omega}} \,\big|\, S^\alpha S^\beta \,\big|\, \hat{\boldsymbol{\Omega}} \,\big\rangle = S(S - \tfrac{1}{2})\, \Omega^\alpha \Omega^\beta + \tfrac{1}{2} S\, \delta_{\alpha\beta} + \frac{i}{2} S\, \epsilon_{\alpha\beta\gamma}\, \Omega^\gamma \quad . \tag{15.61}$$

15.2.1 Coherent state wavefunctions

Consider a state

$$\big|\, \Psi \,\big\rangle = \tfrac{1}{\sqrt{2S!}}\, \Psi(a^\dagger, b^\dagger) \big|\, 0 \,\big\rangle \quad , \tag{15.62}$$

where $\Psi(a^\dagger, b^\dagger)$ is homogeneous of degree $2S$. Then

$$\left\langle\, \hat{\boldsymbol{\Omega}} \,\middle|\, \Psi \,\right\rangle = \Psi(\bar{u}, \bar{v}) \quad , \tag{15.63}$$

where $\Psi(\bar{u}, \bar{v})$ is obtained from $\Psi(a^\dagger, b^\dagger)$ simply by substituting $a^\dagger \to \bar{u}$ and $b^\dagger \to \bar{v}$.

Now suppose we wish to calculate the matrix element of some operator $\hat{A}$ between states $\left|\, \Psi \,\right\rangle$ and $\left|\, \Phi \,\right\rangle$. We assume that $\hat{A}$ preserves total spin, in which case it may be written

$$\begin{aligned} \hat{A} &= \sum_{k,l,j} A_{klj}\, \hat{T}_{klj} \\ \hat{T}_{klj} &= (a)^k\, (b)^l\, (a^\dagger)^{k+j}\, (b^\dagger)^{l-j} \quad . \end{aligned} \tag{15.64}$$

Note here that we have written $\hat{A}$ in normal-ordered form, but this time with the creation operators appearing to the *right*. One then has

$$\left\langle\, \Psi \,\middle|\, \hat{A} \,\middle|\, \Phi \,\right\rangle = \frac{2S+1}{4\pi} \int d\Omega \left\langle\, \Psi \,\middle|\, \hat{\boldsymbol{\Omega}} \,\right\rangle \left\langle\, \hat{\boldsymbol{\Omega}} \,\middle|\, \hat{A} \,\middle|\, \Phi \,\right\rangle \quad . \tag{15.65}$$

It can further be shown that

$$\left\langle\, \hat{\boldsymbol{\Omega}} \,\middle|\, \hat{T}_{klj} \,\middle|\, \Phi \,\right\rangle = \left(\frac{\partial}{\partial \bar{u}}\right)^k \left(\frac{\partial}{\partial \bar{v}}\right)^l \bar{u}^{k+j}\, \bar{v}^{l-j}\, \Phi(\bar{u}, \bar{v}) \tag{15.66}$$

and that

$$\left\langle\, \Psi \,\middle|\, \hat{T}_{klj} \,\middle|\, \Phi \,\right\rangle = \frac{(2S+k+l+1)!}{(2S)!} \cdot \int \frac{d\Omega}{4\pi}\, \Psi^*(u, v)\, u^k\, v^l\, \bar{u}^{k+j}\, \bar{v}^{l-j}\, \Phi(\bar{u}, \bar{v}) \quad . \tag{15.67}$$

15.2.2 Valence bond states

The operator $\mathcal{A}^\dagger_{ij} \equiv a^\dagger_i b^\dagger_j - b^\dagger_i a^\dagger_j$ creates a singlet 'valence bond' between sites i and j.

Exercise: Show that $\mathcal{A}^\dagger_{ij}$ transforms as an SU(2) singlet, i.e. $\mathcal{R}^\dagger\, \mathcal{A}^\dagger_{ij}\, \mathcal{R} = \mathcal{A}^\dagger_{ij}$.

Now consider the valence bond solid (VBS) state

$$\left|\, \Psi(\mathcal{L}, m) \,\right\rangle \equiv \prod_{\langle ij \rangle \in \mathcal{L}} (a^\dagger_i b^\dagger_j - b^\dagger_i a^\dagger_j)^m \left|\, 0 \,\right\rangle \quad , \tag{15.68}$$

where $\left|\, 0 \,\right\rangle$ is the Schwinger boson vacuum. Here, the product is over all links $\langle ij \rangle$ of some regular lattice $\mathcal{L}$. The state $\left|\, \Psi(\mathcal{L}, m) \,\right\rangle$ possesses the following properties:

- $\left|\, \Psi(\mathcal{L}, m) \,\right\rangle$ is a singlet, *i.e.* it has total spin zero.

- For every site i, we have $(a_i^\dagger\, a_i + b_i^\dagger\, b_i)\,\big|\,\Psi(\mathcal{L},m)\,\big\rangle = mz\,\big|\,\Psi(\mathcal{L},m)\,\big\rangle$, where z is the coordination number of $\mathcal{L}$. *I.e.* there is a quantum spin $S = \frac{1}{2}mz$ at every site.

- The maximum eigenvalue of the total link spin $\boldsymbol{J}_{ij} \equiv \boldsymbol{S}_i + \boldsymbol{S}_j$ is $J_{ij}^{\rm max} = 2S - m$. This is significant because with two spin-S objects the total spin will in general range from 0 to $2S$. What is special about the VBS states is that they have zero weight in the sector $J_{ij} > 2S - m$ for every link.

Consequently, $\big|\,\Psi(\mathcal{L},m)\,\big\rangle$ is annihilated by any link spin projection operator $\mathcal{P}_S^J(ij)$, so long as $J > 2S - m$. The projector $\mathcal{P}_S^J(ij)$ may be written as an order $2S$ polynomial in $\boldsymbol{S}_i \cdot \boldsymbol{S}_j$, *viz.*

$$\mathcal{P}_S^J(ij) = \prod_{\substack{k=0\\(k\neq J)}}^{2S} \frac{\boldsymbol{S}_i\cdot\boldsymbol{S}_j + S(S+1) - \frac{1}{2}k(k+1)}{\frac{1}{2}J(J+1) - \frac{1}{2}k(k+1)} \quad . \tag{15.69}$$

Therefore, if one writes a Hamiltonian of the form

$$H = \sum_{\langle ij\rangle} \sum_{J=2S-m+1}^{2S} \lambda_J\, \mathcal{P}_S^J(ij) \tag{15.70}$$

with each $\lambda_J > 0$, then $H\big|\,\Psi(\mathcal{L},m)\,\big\rangle = 0$ and $\big|\,\Psi(\mathcal{L},m)\,\big\rangle$ is an exact, zero energy ground state for H.[2]

The simplest example is for the $S = 1$ linear chain, where

$$\mathcal{P}^{J=1}_{S=\frac{1}{2}}(ij) = \tfrac{1}{6}(\boldsymbol{S}_i\cdot\boldsymbol{S}_j)^2 + \tfrac{1}{2}\boldsymbol{S}_i\cdot\boldsymbol{S}_j + \tfrac{1}{3} \quad . \tag{15.71}$$

We conclude that the bilinear-biquadratic $S = 1$ chain with Hamiltonian

$$H = J\sum_n \Big[\boldsymbol{S}_n\cdot\boldsymbol{S}_{n+1} + \tfrac{1}{3}(\boldsymbol{S}_n\cdot\boldsymbol{S}_{n+1})^2\Big] \tag{15.72}$$

has as its exact ground state $\big|\,\Psi(\mathcal{L},m=1)\,\big\rangle$, where $\mathcal{L}$ is the linear chain. The energy per site is $-\frac{2}{3}J$.

The states $\big|\,\Psi(\mathcal{L},m)\,\big\rangle$ are easily generalized to ones of broken translational or lattice point group symmetry, even while maintaining the constraint that zm link operators $\mathcal{A}_{ij}^\dagger$ are associated with each site i (with different values of j). [3]

[2] If every λ_J is nonnegative, then it is simple to prove that H itself can have no negative eigenvalues.

[3] Were this not the case, then some sites would have different total spin than others. It is perfectly sensible from a mathematical point of view to consider models where the total spin varies from site to site. Most (but by no means all) models of physical interest, however, have one value of S for each magnetic site.

For example, on the honeycomb lattice, where we have links oriented along 0°, 120°, and 240°, we can define the state

$$\big|\,\Psi(m,m',m'')\,\big\rangle \equiv \prod_{\langle ij\rangle\in 0^\circ} (\mathcal{A}^\dagger_{ij})^m \prod_{\langle kl\rangle\in 120^\circ} (\mathcal{A}^\dagger_{kl})^{m'} \prod_{\langle rs\rangle\in 240^\circ} (\mathcal{A}^\dagger_{rs})^{m''} \,\big|\,0\,\big\rangle \quad . \tag{15.73}$$

This state therefore has $S = \frac{1}{2}(m+m'+m'')$ on each site, but it breaks the point group symmetry of the underlying triangular Bravais lattice. Similarly, one can define 'columnar' states on the square lattice which break both translational and point group symmetry, *e.g.*

$$\begin{aligned}\big|\,\Psi_{\rm A}\,\big\rangle &= \prod_{m,n}(a^\dagger_{m,n}b^\dagger_{m+1,n} - b^\dagger_{m,n}a^\dagger_{m+1,n})\,\big|\,0\,\big\rangle \\ \big|\,\Psi_{\rm B}\,\big\rangle &= \prod_{j,n}(a^\dagger_{2j,n}b^\dagger_{2j+1,n} - b^\dagger_{2j,n}a^\dagger_{2j+1,n})^2\,\big|\,0\,\big\rangle \quad .\end{aligned} \tag{15.74}$$

Exercise: Compare and contrast the states $\big|\,\Psi_{\rm A}\,\big\rangle$ *and* $\big|\,\Psi_{\rm B}\,\big\rangle$.

15.2.3 Derivation of spin path integral

Let us compute the real time propagator in the coherent state basis. We begin, as usual, by writing

$$\big\langle\,\hat{\boldsymbol{\Omega}}_N\,\big|\,e^{-iHT/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_0\,\big\rangle = \big\langle\,\hat{\boldsymbol{\Omega}}_N\,\big|\,e^{-i\epsilon H/\hbar}\,1\,e^{-i\epsilon H/\hbar}\,1\cdots 1\,e^{-i\epsilon H/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_0\,\big\rangle \quad , \tag{15.75}$$

where each symbol 1 stands for an insertion of the resolution of the identity, eqn. 15.52. We next compute

$$\begin{aligned}\big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,e^{-i\epsilon H/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle &= \big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle\cdot\left\{1 - \frac{i\epsilon}{\hbar}\,\frac{\big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,H\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle}{\big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle} + \mathcal{O}(\epsilon^2)\right\} \\ &\simeq \big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle\,\exp\Big(-i\epsilon H(\hat{\boldsymbol{\Omega}}_j\,|\,\hat{\boldsymbol{\Omega}}_{j-1})/\hbar\Big) \quad ,\end{aligned} \tag{15.76}$$

where the Hamiltonian is replaced by its coherent state matrix element,

$$H(\hat{\boldsymbol{\Omega}}_j\,|\,\hat{\boldsymbol{\Omega}}_{j-1}) = \frac{\big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,H\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle}{\big\langle\,\hat{\boldsymbol{\Omega}}_j\,\big|\,\hat{\boldsymbol{\Omega}}_{j-1}\,\big\rangle} \quad . \tag{15.77}$$

Exercise: Show that $H(\hat{\boldsymbol{\Omega}}_j\,|\,\hat{\boldsymbol{\Omega}}_{j-1}) = H(\bar{u}_j,\bar{v}_j\,|\,u_{j-1},v_{j-1})$ *is a holomorphic function of its arguments.*

We therefore have

$$\big\langle\,\hat{\boldsymbol{\Omega}}_N\,\big|\,e^{-iHT/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_0\,\big\rangle = \Big(\frac{2S+1}{4\pi}\Big)^{N-1}\int d\Omega_1\cdots\int d\Omega_{N-1}\,e^{i\mathcal{A}[\{\hat{\boldsymbol{\Omega}}_j\}]} \quad , \tag{15.78}$$

where $\mathcal{A} \equiv \mathcal{S}/\hbar$ is given by

$$\mathcal{A} = -i \sum_{j=1}^{N} \ln \left\langle \, \hat{\boldsymbol{\Omega}}_j \, \middle| \, \hat{\boldsymbol{\Omega}}_{j-1} \, \right\rangle - \frac{\epsilon}{\hbar} \sum_{j=1}^{N} H(\hat{\boldsymbol{\Omega}}_j \,|\, \hat{\boldsymbol{\Omega}}_{j-1}) \quad . \tag{15.79}$$

Expanding in the difference between $\hat{\boldsymbol{\Omega}}_j$ and $\hat{\boldsymbol{\Omega}}_{j-1}$, we may write

$$\begin{aligned} \ln \left\langle \, \hat{\boldsymbol{\Omega}}_j \, \middle| \, \hat{\boldsymbol{\Omega}}_{j-1} \, \right\rangle &= 2S \, \ln \Big\{ 1 - \bar{u}_j (u_j - u_{j-1}) - \bar{v}_j (v_j - v_{j-1}) \Big\} \\ &= -2S \, \epsilon \left\{ \bar{u}_j \left(\frac{u_j - u_{j-1}}{\epsilon} \right) - \bar{v}_j \left(\frac{v_j - v_{j-1}}{\epsilon} \right) + \dots \right\} \\ &\simeq -2S \, \epsilon \, (\bar{u}_j \dot{u}_j + \bar{v}_j \dot{v}_j) + \mathcal{O}\big((\hat{\boldsymbol{\Omega}}_j - \hat{\boldsymbol{\Omega}}_{j-1})^2 \big) \quad . \end{aligned} \tag{15.80}$$

The continuum limit is

$$\mathcal{A}[\hat{\boldsymbol{\Omega}}(t)] = \int_0^T dt \left\{ 2iS(\bar{u}\dot{u} + \bar{v}\dot{v}) - \frac{1}{\hbar} \, H(\hat{\boldsymbol{\Omega}}) \right\} \quad , \tag{15.81}$$

where $H(\hat{\boldsymbol{\Omega}}) \equiv H(\hat{\boldsymbol{\Omega}} \,|\, \hat{\boldsymbol{\Omega}})$. Substituting $u = \cos(\theta/2)$ and $v = \sin(\theta/2) \, \exp(i\phi)$, we obtain

$$\bar{u}\dot{u} + \bar{v}\dot{v} = i \sin^2(\theta/2) \, \dot{\phi} = \frac{i}{2}(1 - \cos\theta) \, \dot{\phi} = \frac{i}{2} \, \dot{\omega} \tag{15.82}$$

where $d\omega = (1 - \cos\theta) \, d\phi$ is the differential element of solid angle. We may now, finally, write the spin path integral as

$$\left\langle \, \hat{\boldsymbol{\Omega}}_{\rm f} \, \middle| \, e^{-iHT/\hbar} \, \middle| \, \hat{\boldsymbol{\Omega}}_{\rm i} \, \right\rangle = \int\limits_{\substack{w(0)=w_{\rm i} \\ \bar{w}(T)=\bar{w}_{\rm f}}} \mathcal{D}\hat{\boldsymbol{\Omega}}(t) \, \exp \left\{ -i \int_0^T dt \left[S \, \frac{d\omega}{dt} + \frac{1}{\hbar} H(\hat{\boldsymbol{\Omega}}) \right] \right\} \cdot \left\langle \, \hat{\boldsymbol{\Omega}}_{\rm f} \, \middle| \, \hat{\boldsymbol{\Omega}}(T) \, \right\rangle \left\langle \, \hat{\boldsymbol{\Omega}}(0) \, \middle| \, \hat{\boldsymbol{\Omega}}_{\rm i} \, \right\rangle \tag{15.83}$$

where $w \equiv v/u = \tan(\theta/2) \, \exp(i\phi)$ is the stereographic projection of the spinor coordinates (u, v) onto the complex plane.

The inclusion of the overlap terms inside the path integral is necessary if we are to allow for the possibility of so-called discontinuous paths. Within the semiclassical approximation, $u(t)$ and $v(t)$ are integrated forward from initial data $u_{\rm i}$ and $v_{\rm i}$ while $\bar{u}(t)$ and $\bar{v}(t)$ are integrated backward from final data $\bar{u}_{\rm f}$ and $\bar{v}_{\rm f}$.[4] We encountered an analogous situation with the coherent state path integral for the Heisenberg-Weyl group, where $z(t)$ was integrated forward from initial data $z_{\rm i}$ and $\bar{z}(t)$ integrated backward from final data $\bar{z}_{\rm f}$. In fact, these paths are perfectly

[4] The equations of motion may also be written in terms of the stereographic coordinate $w = v/u$, in which case $w(t)$ is integrated forward from initial data $w_{\rm i}$ and $\bar{w}(t)$ is integrated backward from final data $\bar{w}_{\rm f}$.

continuous; there simply is no reason why $z(T)$ should have any resemblance to $z_{\rm f}$, or $\bar{z}(0)$ to $\bar{z}_{\rm i}$, since the equations of motion know nothing about either $z_{\rm f}$ or $\bar{z}_{\rm i}$.

The thermal, or imaginary time, propagator in the coherent state representation is

$$\left\langle\, \hat{\boldsymbol{\Omega}}_{\rm f} \,\middle|\, e^{-\beta H} \,\middle|\, \hat{\boldsymbol{\Omega}}_{\rm i} \,\right\rangle = \int\limits_{\substack{w(0)=w_{\rm i}\\ \bar{w}(T)=\bar{w}_{\rm f}}} \mathcal{D}\hat{\boldsymbol{\Omega}}(\tau)\, \exp\left\{ -\int\limits_0^{\hbar\beta} d\tau \left[iS\,\frac{d\omega}{d\tau} + \frac{1}{\hbar} H(\hat{\boldsymbol{\Omega}}) \right] \right\} \cdot \left\langle\, \hat{\boldsymbol{\Omega}}_{\rm f} \,\middle|\, \hat{\boldsymbol{\Omega}}(\hbar\beta) \,\right\rangle \left\langle\, \hat{\boldsymbol{\Omega}}(0) \,\middle|\, \hat{\boldsymbol{\Omega}}_{\rm i} \,\right\rangle \quad . \tag{15.84}$$

15.2.4 Gauge field and geometric phase

The solid angle functional $\omega[\hat{\boldsymbol{\Omega}}(t)]$ may be written

$$\omega[\hat{\boldsymbol{\Omega}}(t)] = \int\limits_0^T dt\, \boldsymbol{A}(\hat{\boldsymbol{\Omega}}) \cdot \frac{d\hat{\boldsymbol{\Omega}}}{dt} \tag{15.85}$$

for any $\boldsymbol{A}(\hat{\boldsymbol{\Omega}})$ which satisfies $\nabla \times \boldsymbol{A} = \hat{\boldsymbol{\Omega}}$, *i.e.*

$$\Omega^a = \epsilon_{abc}\, \frac{\partial}{\partial \Omega^b}\, A^c(\hat{\boldsymbol{\Omega}}) \tag{15.86}$$

(To see this, use Stokes' theorem.) We now derive a useful result:

$$\begin{aligned} \delta\omega[\hat{\boldsymbol{\Omega}}(t)] &= \int dt \left\{ \frac{\partial A^b}{\partial \Omega^a}\, \frac{d\Omega^b}{dt}\, \delta\Omega^a + A^a\, \frac{d}{dt}\, \delta\Omega^a \right\} \\ &= \int dt \left(\frac{\partial A^b}{\partial \Omega^a} - \frac{\partial A^a}{\partial \Omega^b} \right) \frac{d\Omega^b}{dt}\, \delta\Omega^a \\ &= \int dt\, \delta\Omega^a\, \epsilon_{abc} \dot{\Omega}^b\, \Omega^c = \int dt\, \delta\hat{\boldsymbol{\Omega}} \cdot \frac{\partial \hat{\boldsymbol{\Omega}}}{\partial t} \times \hat{\boldsymbol{\Omega}} \quad , \end{aligned} \tag{15.87}$$

and hence the functional derivative is

$$\frac{\delta\omega[\hat{\boldsymbol{\Omega}}]}{\delta\hat{\boldsymbol{\Omega}}(t)} = \frac{\partial\hat{\boldsymbol{\Omega}}}{\partial t} \times \hat{\boldsymbol{\Omega}} \quad . \tag{15.88}$$

15.2.5 Semiclassical dynamics

We begin with the action functional,

$$\tilde{\mathcal{A}}[\hat{\boldsymbol{\Omega}}(t), \lambda(t)] \equiv \mathcal{A}[\hat{\boldsymbol{\Omega}}(t)] + \int\limits_0^T dt\, \lambda(t) \big(\hat{\boldsymbol{\Omega}}^2(t) - 1 \big) \quad . \tag{15.89}$$

Here, $\lambda(t)$ is a Lagrange multiplier field which enforces the constraint $\hat{\boldsymbol{\Omega}}(t) \cdot \hat{\boldsymbol{\Omega}}(t) = 1$ at all times. We next vary with respect to $\hat{\boldsymbol{\Omega}}(t)$ and $\lambda(t)$:

$$\begin{aligned} \frac{\delta \tilde{\mathcal{A}}}{\delta \hat{\boldsymbol{\Omega}}(t)} &= -S \frac{\partial \hat{\boldsymbol{\Omega}}}{\partial t} \times \hat{\boldsymbol{\Omega}} - \frac{1}{\hbar} \frac{\partial H}{\partial \hat{\boldsymbol{\Omega}}} + 2\,\lambda\,\hat{\boldsymbol{\Omega}} \\ \frac{\delta \tilde{\mathcal{A}}}{\delta \lambda(t)} &= \hat{\boldsymbol{\Omega}}^2(t) - 1 \quad . \end{aligned} \tag{15.90}$$

Setting these variations to zero, we solve for $\lambda(t)$ by taking the dot product of the first equation with $\hat{\boldsymbol{\Omega}}(t)$ and then substituting $\hat{\boldsymbol{\Omega}}^2(t) = 1$. In this manner, we find

$$\lambda = \frac{1}{2\hbar} \frac{\partial H}{\partial \hat{\boldsymbol{\Omega}}} \cdot \hat{\boldsymbol{\Omega}} \quad , \tag{15.91}$$

The effect of this is to render all terms on the RHS of eqn. 15.90 orthogonal to $\hat{\boldsymbol{\Omega}}$, thereby effectively projecting $\partial H/\partial \hat{\boldsymbol{\Omega}}$ onto this orthogonal subspace. It is then easy to obtain the equations of motion

$$\hbar S \frac{\partial \hat{\boldsymbol{\Omega}}}{\partial t} = \frac{\partial H}{\partial \hat{\boldsymbol{\Omega}}} \times \hat{\boldsymbol{\Omega}} \quad . \tag{15.92}$$

If we write the equations of motion in terms of the spinor coordinates $\{u, v, \bar{u}, \bar{v}\}$ themselves, it is important to recognize that they must satisfy the constraint $u\bar{u} + v\bar{v} = 1$. A Lagrange multiplier field λ is invoked to impose this constraint at every value of the time t. This results in the equations of motion

$$2i\hbar S\dot{u} = \frac{\partial H}{\partial \bar{u}} + \lambda u \qquad\qquad 2i\hbar S\dot{v} = \frac{\partial H}{\partial \bar{v}} + \lambda v \tag{15.93}$$

$$-2i\hbar S\dot{\bar{u}} = \frac{\partial H}{\partial u} + \lambda \bar{u} \qquad\qquad -2i\hbar S\dot{\bar{v}} = \frac{\partial H}{\partial v} + \lambda \bar{v} \quad . \tag{15.94}$$

Varying the action with respect to the Lagrange multiplier field of course yields the constraint equation. We are then left with five equations in the five unknowns $\{u, v, \bar{u}, \bar{v}, \lambda\}$, along with the four boundary conditions,

$$u(0) = u_{\rm i} \quad , \quad \bar{u}(T) = \bar{u}_{\rm f} \quad , \quad v(0) = v_{\rm i} \quad , \quad \bar{v}(T) = \bar{v}_{\rm f} \quad . \tag{15.95}$$

Implementing the constraint, one obtains an expression for λ,

$$\lambda = 2iS(\bar{u}\dot{u} + \bar{v}\dot{v}) - \bar{u}\frac{\partial H}{\partial \bar{u}} - \bar{v}\frac{\partial H}{\partial \bar{v}} \tag{15.96}$$

$$= -2iS(u\dot{\bar{u}} + v\dot{\bar{v}}) - u\frac{\partial H}{\partial u} - v\frac{\partial H}{\partial v} \quad . \tag{15.97}$$

Note that for real θ and ϕ that eqns. 15.93 and eqn. 15.94 are related by complex conjugation.

15.3 Other Useful Representations of the Spin Path Integral

15.3.1 Stereographic representation

In the stereographic representation, we write

$$w \equiv \frac{v}{u} = \tan(\theta/2)\, e^{i\phi} \quad . \tag{15.98}$$

One then finds

$$\frac{\bar{w}\dot{w}}{1+\bar{w}w} = \bar{u}\dot{u} + \bar{v}\dot{v} - \frac{d}{dt}\ln u \quad . \tag{15.99}$$

From the differential

$$dw = \tfrac{1}{2}\,\sec^2\big(\theta/2\big)\, e^{i\phi}\, d\theta + i\tan(\theta/2)\, e^{i\phi}\, d\phi \quad , \tag{15.100}$$

we obtain

$$\frac{dw\wedge d\bar{w}}{(1+\bar{w}w)^2} = \frac{1}{2i}\,\sin\theta\, d\theta\wedge d\phi \quad . \tag{15.101}$$

The Hamiltonian matrix elements may be recast in terms of w and $\bar{w}$. For example,

$$\begin{aligned} S^+ = a^\dagger b \longrightarrow 2S\,\bar{u}v &= \frac{2S\,w}{1+\bar{w}w} \\ S^z = \tfrac{1}{2}(a^\dagger a - b^\dagger b) \longrightarrow S\,(\bar{u}u - \bar{v}v) &= S\,\frac{1-\bar{w}w}{1+\bar{w}w} \quad . \end{aligned} \tag{15.102}$$

Thus, the real and imaginary time path integrals are given by

$$\big\langle\, \hat{\boldsymbol{\Omega}}_{\rm f}\,\big|\, e^{-iHT/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_{\rm i}\,\big\rangle = \int \mathcal{D}[\bar{w}(t), w(t)]\,\exp\left\{ -i\int_0^T dt\,\Big[-iS\,\frac{\bar{w}\dot{w}-\dot{\bar{w}}w}{1+\bar{w}w} + \frac{1}{\hbar}H(\bar{w},w)\Big]\right\} \tag{15.103}$$

and

$$\big\langle\, \hat{\boldsymbol{\Omega}}_{\rm f}\,\big|\, e^{-HT/\hbar}\,\big|\,\hat{\boldsymbol{\Omega}}_{\rm i}\,\big\rangle = \int \mathcal{D}[\bar{w}(t), w(t)]\,\exp\left\{ -\int_0^T dt\,\Big[S\,\frac{\bar{w}\dot{w}-\dot{\bar{w}}w}{1+\bar{w}w} + \frac{1}{\hbar}H(\bar{w},w)\Big]\right\} \quad , \tag{15.104}$$

respectively. In these above expressions, the metric $\mathcal{D}[\bar{w}, w]$ includes the $(1+\bar{w}w)^{-2}$ factor at each time step, and the Hamiltonian $H(\bar{w}, w)$ is the coherent state diagonal matrix element expressed in terms of the stereographic coordinate w and its conjugate $\bar{w}$. These expressions are incomplete, however, in that we've omitted the boundary overlap factors at $t = 0$ and $t = T$.

Exercise: Complete the expression in eqns. 15.103 and 15.104, adding the boundary terms.

15.3.2 Recovery of spin wave theory

To recover spin wave theory and the Holstein-Primakoff transformation, define $z \equiv u\bar{v}/|v| = \cos(\theta/2)\exp(-i\phi)$. Then $|z|^2 = |u|^2$ and

$$dz = -\tfrac{1}{2}\,\sin(\theta/2)\,e^{-i\phi}\,d\theta - i\cos(\theta/2)\,e^{-i\phi}\,d\phi \quad . \tag{15.105}$$

We then obtain

$$dz \wedge d\bar{z} = \frac{1}{2i}\,\sin\theta\,d\theta \wedge d\phi \quad . \tag{15.106}$$

The geometrical phase, which is responsible for the $\omega[\hat{\boldsymbol{\Omega}}(t)]$ term in the action functional, is obtained using

$$\bar{z}\dot{z} = \frac{i}{2}\,(1-\cos\theta)\,d\phi - i\,d\phi + d\cos^2(\theta/2) \tag{15.107}$$

which, after dropping the total time derivatives, yields $(i/2)\,d\omega$. As for the Hamiltonian, we have

$$\begin{aligned} S^+ = a^\dagger b \longrightarrow 2S\,\bar{u}v = 2S\,\bar{z}\,\sqrt{1-\bar{z}z} \\ S^z = \tfrac{1}{2}(a^\dagger a - b^\dagger b) \longrightarrow S\,(\bar{u}u - \bar{v}v) = 2S\,(\bar{z}z - \tfrac{1}{2}) \quad . \end{aligned} \tag{15.108}$$

This is equivalent to Holstein-Primakoff, with $h \equiv \sqrt{2S}\,z$ as the HP boson. We therefore obtain

$$\left\langle\, \hat{\boldsymbol{\Omega}}_{\rm f} \,\middle|\, e^{-iHT/\hbar} \,\middle|\, \hat{\boldsymbol{\Omega}}_{\rm i} \,\right\rangle = \int \mathcal{D}[\bar{h}(t), h(t)]\,\exp\left\{ -\,i\int_0^T dt\,\left[\bar{h}\dot{h} + \frac{1}{\hbar}H(\bar{h},h)\right]\right\} \quad , \tag{15.109}$$

where the functional integration is over a disk of area $2\pi S$ for each time t.

15.4 Quantum Tunneling of Spin

15.4.1 Model Hamiltonian

The theory of quantum spin tunneling has been developed largely by E. Chudnovsky, A. Garg, D. Loss, and others. Consider the following model Hamiltonian,

$$H = K_1 S_z^2 + K_2 S_y^2 - \gamma H S_z \quad , \tag{15.110}$$

where $K_1 > K_2 > 0$. This describes a spin-S particle with an easy axis along $\hat{\boldsymbol{x}}$ and a hard axis along $\hat{\boldsymbol{z}}$. To treat this problem by the coherent state path integral, we need to compute the diagonal matrix element of H in the coherent state basis. One finds,

$$E(\theta,\phi) = \left\langle\, \hat{\boldsymbol{\Omega}} \,\middle|\, H \,\middle|\, \hat{\boldsymbol{\Omega}} \,\right\rangle = k_1\,\cos^2\theta + k_2\,\sin^2(\theta)\,\sin^2(\phi) - h\,\cos(\theta) \quad , \tag{15.111}$$

where $k_i = S(S-\frac{1}{2})K_i$ $(i=1,2)$, $h = \gamma S H$, and where we have dropped an unimportant constant. In weak fields h, the energy function $E(\theta,\phi)$ has the following features:

- $E(\theta,\phi)$ has two degenerate minima with $\theta_0 = \cos^{-1}(h/2k_1)$ at $\phi = 0$ and at $\phi = \pi$. The minimum energy is $E_0^{\rm cl} = -h^2/4k_1$.
- There is a global maximum with $E_{\rm max} = k_1 + h$ located (assuming $h > 0$) at the South Pole ($\theta = \pi$), and a local maximum with $E'_{\rm max} = k_1 - h$ located at the North Pole ($\theta = 0$).
- There are two saddle points, located at $\theta_{\rm x} = \cos^{-1}(h/2\big(k_1 - k_2)\big)$, with $\phi = \pm\frac{1}{2}\pi$. The energy of the saddle points is $E_{\rm saddle} = k_2 - \frac{1}{4}h^2/(k_1 - k_2)$.

We therefore expect to find two low-lying states which are linear combinations of the coherent states $\big|\,\theta = \theta_0, \phi = 0\,\big\rangle$ and $\big|\,\theta = \theta_0, \phi = \pi\,\big\rangle$. Let us abbreviate these two states $\big|\,0\,\big\rangle$ and $\big|\,\pi\,\big\rangle$, respectively. The eigenstates of the system should be symmetric and antisymmetric combinations of these states: $\big|\,\pm\,\big\rangle = 2^{-1/2}\Big\{\big|\,0\,\big\rangle \pm \big|\,\pi\,\big\rangle\Big\}$. The tunnel splitting $\Delta = E_0$ may be obtained by examining the matrix elements,

$$\begin{aligned}\big\langle\,+\,\big|\,e^{-\beta H}\,\big|\,+\,\big\rangle &= \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,0\,\big\rangle + \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,\pi\,\big\rangle = e^{-\beta E_0} \\ \big\langle\,-\,\big|\,e^{-\beta H}\,\big|\,-\,\big\rangle &= \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,0\,\big\rangle - \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,\pi\,\big\rangle = e^{-\beta(E_0+\Delta)}\end{aligned} \quad , \tag{15.112}$$

where E_0 differs from $E_0^{\rm cl}$ due to 'zero-point energy', *i.e.* quantum fluctuations. In reality, there is no reason why the states $\big|\,\pm\,\big\rangle$ should necessarily be eigenstates of H. What is important, though, is that the antisymmetric combination projects out all of the ground state. By taking the $\beta \to \infty$ limit, the contribution from admixtures of higher-lying eigenstates to $\big|\,\pm\,\big\rangle$ can be suppressed. What this means is that we can calculate the exact tunnel splitting by the formula,

$$\Delta = \lim_{\beta\to\infty} \frac{1}{\beta} \ln\left\{\frac{\big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,0\,\big\rangle + \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,\pi\,\big\rangle}{\big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,0\,\big\rangle - \big\langle\,0\,\big|\,e^{-\beta H}\,\big|\,\pi\,\big\rangle}\right\} \quad . \tag{15.113}$$

Another way, of course, to compute the tunnel splitting is to simply numerically diagonalize the rank-$(2J+1)$ Hamiltonian matrix. This works without fail, but it is not particularly instructive in elucidating the physics of spin tunneling. Moreover, it may be that an instanton calculation, which we shall presently describe, yields certain analytic results which are useful and in general impossible to obtain numerically.

15.4.2 Instantons and tunnel splittings

The essence of the instanton approach to quantum tunneling is described in a beautiful article by Sidney Coleman, entitled "The Uses of Instantons". We write the imaginary time matrix element $\big\langle\,{\rm P_f}\,\big|\,\exp(-\beta H)\,\big|\,{\rm P_i}\,\big\rangle$ between points ${\rm P}_1$ and ${\rm P}_2$ as a path integral. In our case, each ${\rm P}$ labels a spin orientation $\hat{\boldsymbol{\Omega}}$, and each state $\big|\,{\rm P}\,\big\rangle$ is a spin coherent state. We extremize the action, applying the method of stationary phase. This involves solving the classical equations

of motion, subject to boundary conditions which we shall not fully specify, save to say that the most naïve boundary conditions are simply $\hat{\boldsymbol{\Omega}}(0) = \hat{\boldsymbol{\Omega}}_{\rm i}$ an $\hat{\boldsymbol{\Omega}}(\hbar\beta) = \hat{\boldsymbol{\Omega}}_{\rm f}$.[5]

There may be several instanton paths connecting P_1 and P_2. Associated with each such instanton α is a characteristic time τ_α, a classical action $Y_\alpha + i\phi_\alpha$, written in units of $\hbar$ and separating real and imaginary parts, and also a 'fluctuation determinant' prefactor D_α arising from integrating over Gaussian fluctuations about the classical instanton trajectory. If we write

$$\xi_\alpha = D_\alpha \, e^{i\phi_\alpha} \, e^{-Y_\alpha} \quad , \tag{15.114}$$

then the diagonal matrix element can be written in the 'dilute instanton gas' approximation as

$$\begin{aligned} \big\langle \, \mathrm{P}_1 \, \big| \, e^{-\beta H} \, \big| \, \mathrm{P}_1 \, \big\rangle &= \sum_{n=0}^{\infty} \sum_{\{\alpha_k, \bar\alpha_k\}} \int\limits_0^{\hbar\beta} \! d\tau_1 \cdots \int\limits_{\tau_{2n-1}}^{\hbar\beta} \! d\tau_{2n} \, \xi_{\alpha_1} \, \xi_{\bar\alpha_1} \cdots \xi_{\bar\alpha_n} \\ &= \cosh\left(\hbar\beta \Big| \sum_\alpha \xi_\alpha \Big| \right) \quad . \end{aligned} \tag{15.115}$$

Here we denote the return instantons from P_2 to P_1 with the index $\bar\alpha$. Since the return path is a time-reversed one, we have $\xi_{\bar\alpha} = \overline{\xi_\alpha}$, *i.e.* the return paths have opposite phase.

The off-diagonal matrix element, in which paths must begin at P_1 and end at P_2 requires an odd number of instanton events, and is given by

$$\begin{aligned} \big\langle \, \mathrm{P}_2 \, \big| \, e^{-\beta H} \, \big| \, \mathrm{P}_1 \, \big\rangle &= \sum_{n=0}^{\infty} \sum_{\{\alpha_k, \bar\alpha_k\}} \int\limits_0^{\hbar\beta} \! d\tau_1 \cdots \int\limits_{\tau_{2n}}^{\hbar\beta} \! d\tau_{2n+1} \, \xi_{\alpha_1} \, \xi_{\bar\alpha_1} \cdots \xi_{\bar\alpha_n} \, \xi_{\alpha_{n+1}} \\ &= \frac{\sum_\alpha \xi_\alpha}{\Big| \sum_\alpha \xi_\alpha \Big|} \cdot \sinh\left(\hbar\beta \Big| \sum_\alpha \xi_\alpha \Big| \right) \quad . \end{aligned} \tag{15.116}$$

If $\sum_\alpha \xi_\alpha$ is real, then we can read off the tunnel splitting:

$$\Delta = 2\hbar \sum_\alpha D_\alpha \, e^{i\phi_\alpha} \, e^{-Y_\alpha} \quad . \tag{15.117}$$

15.4.3 Garg's calculation (1993)

Starting from the Euclidean Lagrangian,

$$L_{\rm E} = i\hbar \, S(1 - \cos\theta) \, \dot\phi + E(\theta, \phi) \quad , \tag{15.118}$$

[5] In fact, the proper boundary conditions are $u(0) = u_{\rm i}$, $v(0) = v_{\rm i}$, $\bar u(\hbar\beta) = \bar u_{\rm f}$, and $\bar v(\hbar\beta) = \bar v_{\rm f}$, as derived above.

one derives the Euler-Lagrange equations of motion,

$$\begin{aligned}\frac{\partial L_{\rm E}}{\partial\theta} - \frac{d}{dt}\frac{\partial L_{\rm E}}{\partial\dot\theta} = 0 = \frac{1}{\hbar}\frac{\partial E}{\partial\theta} + iS\sin(\theta)\,\dot\phi \\ \frac{\partial L_{\rm E}}{\partial\phi} - \frac{d}{dt}\frac{\partial L_{\rm E}}{\partial\dot\phi} = 0 = \frac{1}{\hbar}\frac{\partial E}{\partial\phi} - iS\sin(\theta)\,\dot\theta \quad .\end{aligned} \tag{15.119}$$

Note that

$$\frac{dE}{dt} = \dot\theta\,\frac{\partial E}{\partial\theta} + \dot\phi\,\frac{\partial E}{\partial\phi} = 0 \quad , \tag{15.120}$$

which says that the energy $E(\theta,\phi)$ is conserved along the classical trajectories. One can use this result to finesse the instanton calculation and solve directly for θ as a function of ϕ. Energy conservation provides a quadratic equation in $\cos(\theta)$,

$$E_0 = -\frac{h^2}{4k_1} = k_1\,\cos^2(\theta) + k_2\,\sin^2(\theta)\,\sin^2\phi - h\,\cos(\theta) \quad , \tag{15.121}$$

the solution of which is written (Garg, 1993),

$$u(\phi) = \frac{u_0 + i\sqrt{\lambda}\,\sin\phi\,\sqrt{1-u_0^2-\lambda\,\sin^2\phi}}{1-\lambda\,\sin^2\phi} \tag{15.122}$$

where $u \equiv \cos(\theta)$, $u_0 = h/2k_1 \equiv h/h_{\rm c}$, and $\lambda = k_2/k_1$. Note that $u = \cos\theta$ is complex along the instanton path. Nevertheless, the path obeys the boundary condition that $u = u_0$ at $\phi = 0$ and $\phi = \pi$. The dimensionless instanton action is

$$\mathcal{A} = Y + i\phi = \beta E_0 + iS\int_0^{\pm\pi} d\phi\,\big\{1-u(\phi)\big\} \quad , \tag{15.123}$$

whence

$$\phi = \mathsf{Im}\;\mathcal{A} = \pm S\int_0^{\pi} d\phi\,\left\{1-\frac{u_0}{1-\lambda\,\sin^2\phi}\right\} = \pm\pi S\left\{1-\frac{u_0}{\sqrt{1-\lambda^2}}\right\} \quad . \tag{15.124}$$

Thus, there are two instantons connecting $(\theta,\phi) = (\theta_0,0)$ and $(\theta,\phi) = (\theta_0,\pi)$ which wind around the sphere in opposite directions. The tunnel splitting, according to eqn. 15.117, is

$$\Delta = 4D\,e^{-Y}\,\cos\left(\pi S\left[1-\frac{h}{2\sqrt{k_1^2-k_2^2}}\right]\right) \quad . \tag{15.125}$$

The tunnel splitting therefore vanishes at a set of dimensionless field strengths h_m, where

$$h_m = 2\sqrt{k_1^2-k_2^2}\,\left\{1-\frac{m+\frac{1}{2}}{S}\right\} \quad . \tag{15.126}$$

Note that for $h = 0$ the splitting vanishes whenever $S = m + \frac{1}{2}$, which is to say whenever the ground state is a Kramers doublet.

In Fe_8 clusters, where $S = 10$, this predicts ten values of $h > 0$ where Δ vanishes. In fact, experiments by Wernsdorfer and Sessoli see only four such vanishings. The reason for this is that the effective Hamiltonian for the experimental molecule includes a term proportional to $J_+^4 + J_-^4$ which is not included in the Hamiltonian of eqn. 15.110. This new term allows for two additional instanton solutions. Moreover, the new solutions exhibit discontinuities in $\hat{\boldsymbol{\Omega}}(\tau)$ at the boundaries $\tau = 0$ and $\tau = \hbar\beta$. This very interesting result was obtained by Keçecioğlu and Garg (2002).

15.5 Haldane's Mapping to the Nonlinear Sigma Model

The many-spin dimensionless action is

$$\mathcal{A} = -S \sum_i \omega[\hat{\boldsymbol{\Omega}}_i(t)] - \frac{1}{\hbar} \int_0^T dt\, H\big(\{\hat{\boldsymbol{\Omega}}_i(t)\}\big) \quad , \tag{15.127}$$

where the Hamiltonian is that of a Heisenberg antiferromagnet, with diagonal coherent state matrix elements given by

$$H\big(\{\hat{\boldsymbol{\Omega}}_i(t)\}\big) = \tfrac{1}{2} S^2 \sum_{i,j} J_{ij}\, \hat{\boldsymbol{\Omega}}_i \cdot \hat{\boldsymbol{\Omega}}_j \quad , \tag{15.128}$$

where $J_{ij} = J\big(|\boldsymbol{R}_{ij}|\big)$ with $\boldsymbol{R}_{ij} = \boldsymbol{R}_i - \boldsymbol{R}_j$ is a function of the distance between sites i and j in the lattice. The spin coherent state at site i is polarized along the direction

$$\hat{\boldsymbol{\Omega}}_i = \eta_i\, \hat{\boldsymbol{n}}_i \sqrt{1 - \Big(\frac{v_0\, \boldsymbol{L}_i}{\hbar S}\Big)^2} + \frac{v_0}{\hbar S}\, \boldsymbol{L}_i \quad , \tag{15.129}$$

where $\hat{\boldsymbol{n}}_i \cdot \boldsymbol{L}_i = 0$. Here, $\hat{\boldsymbol{n}}_i$ is the local *Néel field*, which varies slowly once the sublattice modulation η_i extracted from the spin field $\hat{\boldsymbol{\Omega}}_i$, $\boldsymbol{L}_i$ describes ferromagnetic fluctuations about the local Néel order; v_0 is the unit cell volume. Note that

$$\hbar S \sum_i \hat{\boldsymbol{\Omega}}_i = v_0 \sum_i \boldsymbol{L}_i = \int d^d x\, \boldsymbol{L}(\boldsymbol{x}) \quad , \tag{15.130}$$

where the RHS is obtained after taking the continuum limit.

15.5.1 Hamiltonian

We now expand the Heisenberg interaction $\hat{\boldsymbol{\Omega}}_i \cdot \hat{\boldsymbol{\Omega}}_j$ in the slowly varying quantities $\hat{\boldsymbol{n}}_i - \hat{\boldsymbol{n}}_j$ and $\boldsymbol{L}_i$. Since $\hat{n}_i$ is a unit vector, we may write

$$\hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{n}}_j = 1 - \tfrac{1}{2}(\hat{\boldsymbol{n}}_i - \hat{\boldsymbol{n}}_j) \cdot (\hat{\boldsymbol{n}}_i - \hat{\boldsymbol{n}}_j) \quad . \tag{15.131}$$

We then have

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_i \cdot \hat{\boldsymbol{\Omega}}_j = \eta_i \, \eta_j \Big\{ 1 - \tfrac{1}{2}(\hat{\boldsymbol{n}}_i - \hat{\boldsymbol{n}}_j) \cdot (\hat{\boldsymbol{n}}_i - \hat{\boldsymbol{n}}_j) \Big\} \Big\{ 1 - \tfrac{1}{2} \Big(\frac{v_0}{\hbar S} \Big) \big(\boldsymbol{L}_i^2 + \boldsymbol{L}_j^2 \big) + \ldots \Big\} \\ + \frac{v_0}{\hbar S} \, \eta_i \, \hat{\boldsymbol{n}}_i \cdot (\boldsymbol{L}_j - \boldsymbol{L}_i) + \frac{v_0}{\hbar S} \, \eta_j \, \hat{\boldsymbol{n}}_j \cdot (\boldsymbol{L}_i - \boldsymbol{L}_j) \\ + \tfrac{1}{2} \left(\frac{v_0}{\hbar S} \right)^2 \Big\{ \boldsymbol{L}_i^2 + \boldsymbol{L}_j^2 - (\boldsymbol{L}_i - \boldsymbol{L}_j)^2 \Big\} \quad .\end{aligned} \tag{15.132}$$

Lattice differences may now be expanded in derivatives, as

$$f(\boldsymbol{R}_j) - f(\boldsymbol{R}_i) = (R_j^\mu - R_i^\mu) \, \frac{\partial f(\boldsymbol{R}_i)}{\partial R_i^\mu} + \tfrac{1}{2} \, (R_j^\mu - R_i^\mu) \, (R_j^\nu - R_i^\nu) \, \frac{\partial^2 f(\boldsymbol{R}_i)}{\partial R_i^\mu \, \partial R_i^\nu} + \ldots \tag{15.133}$$

Expanding to Gaussian order in the fields $\hat{\boldsymbol{n}}$ and $\boldsymbol{L}$ and their gradients, we find

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_i \cdot \hat{\boldsymbol{\Omega}}_j = \eta_i \, \eta_j \Big\{ 1 - \tfrac{1}{2} \, (R_j^\mu - R_i^\mu) \, (R_j^\nu - R_i^\nu) \, (\partial_\mu n_i^a) \, (\partial_\nu n_i^a) + \ldots \Big\} \\ + \tfrac{1}{2} \left(\frac{v_0}{\hbar S} \right)^2 \Big\{ (1 - \eta_i \, \eta_j)(\boldsymbol{L}_i^2 + \boldsymbol{L}_j^2) - (R_j^\mu - R_i^\mu) \, (R_j^\nu - R_i^\nu) \, (\partial_\mu L_i^a) \, (\partial_\nu L_i^a) + \ldots \Big\} \\ + \frac{v_0}{\hbar S} \, \eta_i \, n_i^a \, (R_j^\mu - R_i^\mu) \, (\partial_\mu L_i^a) - \frac{v_0}{\hbar S} \, \eta_j \, n_j^a \, (R_j^\mu - R_i^\mu) \, (\partial_\mu L_j^a) + \ldots \quad .\end{aligned} \tag{15.134}$$

Upon performing the double sum over lattice sites i and j, the terms on the last line vanish, and we are left with

$$H = H_{\hat{\boldsymbol{n}}} + H_{\boldsymbol{L}} + E_0 \tag{15.135}$$

where the classical energy E_0 is given by

$$E_0 = \tfrac{1}{2} \, S^2 \sum_{i,j} J_{ij} \, \eta_i \, \eta_j \quad . \tag{15.136}$$

The Hamiltonian also contains contributions due to gradients in the Néel field,

$$H_{\hat{\boldsymbol{n}}} = - \frac{S^2}{4d\mathcal{N} v_0} \sum_{i,j} J_{ij} \, \eta_i \, \eta_j \, \big| \boldsymbol{R}_i - \boldsymbol{R}_j \big|^2 \cdot \int d^d x \, (\partial_\mu n^a)^2 \tag{15.137}$$

where d is the dimensionality of the (presumed hypercubic) lattice and $\mathcal{N}$ is the number of lattice sites, and from ferromagnetic fluctuations,

$$H_{\boldsymbol{L}} = \frac{v_0}{2\mathcal{N}\hbar^2}\sum_{i,j} J_{ij}\,(1-\eta_i\,\eta_j)\cdot\int d^dx\,\boldsymbol{L}^2(\boldsymbol{x}) - \frac{v_0}{4d\mathcal{N}\hbar^2}\sum_{i,j} J_{ij}\,\big|\boldsymbol{R}_i-\boldsymbol{R}_j\big|^2\cdot\int d^dx\,(\partial_\mu L^a)^2 \quad . \tag{15.138}$$

Retaining only terms of order $\boldsymbol{L}^2$ – and hence dropping terms of order $(\nabla\boldsymbol{L})^2$ – we obtain the Hamiltonian,

$$H = \int d^dx\,\Big\{\tfrac{1}{2}\,\rho_{\rm s}(\vec\nabla\hat{\boldsymbol{n}})^2 + \tfrac{1}{2}\,\chi^{-1}\boldsymbol{L}^2\Big\} \tag{15.139}$$

where the spin stiffness is given by

$$\rho_{\rm s} \equiv -\frac{S^2}{2d\mathcal{N}v_0}\sum_{i,j} J_{ij}\,\eta_i\,\eta_j\,\big|\boldsymbol{R}_i-\boldsymbol{R}_j\big|^2 \tag{15.140}$$

and the inverse susceptibility is given by

$$\chi^{-1} \equiv \frac{v_0}{\mathcal{N}\hbar^2}\sum_{ij} J_{ij}\,(1-\eta_i\,\eta_j) = \frac{v_0}{\hbar^2}\Big[\hat J(0)-\hat J(\boldsymbol{Q})\Big] \quad , \tag{15.141}$$

where $\boldsymbol{Q}\equiv(\pi/a,\pi/a,\ldots.\pi/a)$ is the zone corner wavevector. The dimensions of $\rho_{\rm s}$ and χ are:

$$[\rho_{\rm s}] = E\cdot L^{2-d} \qquad ; \qquad [\chi] = E\cdot T^2\cdot L^{-d} \quad . \tag{15.142}$$

15.5.2 Geometric phase

The geometric phase contribution to the dimensionless action is written

$$\mathcal{A}_{\rm B} = -S\sum_i \omega[\hat{\boldsymbol{\Omega}}_i(t)] = -S\sum_i \eta_i\omega\Big[\hat{\boldsymbol{n}}_i(t) + \eta_i\,\frac{v_0}{\hbar S}\,\boldsymbol{L}_i(t)\Big] \quad . \tag{15.143}$$

We now expand in the notionally small quantity linear in $\boldsymbol{L}_i$, using the result of eqn. 15.88:

$$\begin{aligned}\mathcal{A}_{\rm B} &= -S\sum_i \eta_i\,\omega[\hat{\boldsymbol{n}}_i] - S\int dt\,\sum_i\Big(\frac{v_0}{\hbar S}\Big)\boldsymbol{L}_i\cdot\frac{\partial\hat{\boldsymbol{n}}_i}{\partial t}\times\hat{\boldsymbol{n}}_i \\ &= -S\sum_i \eta_i\,\omega[\hat{\boldsymbol{n}}_i] - \frac{1}{\hbar}\int d^dx\int dt\,\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}}\cdot\boldsymbol{L} \quad .\end{aligned} \tag{15.144}$$

15.5.3 Emergence of the nonlinear sigma model

Let's start with the quantum action obtained thus far,

$$\mathcal{A} = -\frac{1}{\hbar}\int d^dx \int dt \left\{ \tfrac{1}{2}\,\rho_{\rm s}(\vec\nabla\hat{\boldsymbol{n}})^2 + \frac{\boldsymbol{L}^2}{2\chi} + \boldsymbol{L}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}} \right. \\ \left. + \frac{g_0\mu_{\rm B}}{\hbar S}\boldsymbol{H}_{\rm u}\cdot\boldsymbol{L} + \frac{\hbar g_0\mu_{\rm B}}{v_0}\boldsymbol{H}_{\rm s}\cdot\hat{\boldsymbol{n}}\right\} - S\sum_i \eta_i\,\omega[\hat{\boldsymbol{n}}_i] \quad . \tag{15.145}$$

We have included here an external field $\boldsymbol{H}(\boldsymbol{x},t)$ which has uniform ($\boldsymbol{k}\approx 0$) and staggered ($\boldsymbol{k}\approx\boldsymbol{Q}$) components $\boldsymbol{H}_{\rm u}$ and $\boldsymbol{H}_{\rm s}$, respectively. Now let us integrate out $\boldsymbol{L}$. In order to do so, we must introduce a Lagrange multiplier field $\lambda(\boldsymbol{x},t)$ which enforces the local constraint $\hat{\boldsymbol{n}}\cdot\boldsymbol{L}=0$. At each position $\boldsymbol{x}$, we must evaluate the functional integral

$$\begin{aligned}
\mathcal{I} &\equiv \int\mathcal{D}\lambda(t)\int\mathcal{D}\boldsymbol{L}(t)\,\exp\left\{-\frac{i}{\hbar}\int d^dx\int dt\left[\frac{\boldsymbol{L}^2}{2\chi}+\boldsymbol{L}\cdot\Big(\lambda\hat{\boldsymbol{n}}+\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}}+\frac{g_0\mu_{\rm B}}{\hbar S}\boldsymbol{H}_{\rm u}\Big)\right]\right\} \\
&= \mathcal{I}_0\int\mathcal{D}\lambda(t)\,\exp\left\{\frac{i\chi}{2\hbar}\int d^dx\int dt\left(\lambda\hat{\boldsymbol{n}}+\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}}+\frac{g_0\mu_{\rm B}}{\hbar S}\boldsymbol{H}_{\rm u}\right)^2\right\} \\
&= \tilde{\mathcal{I}}_0\,\exp\left\{\frac{i}{\hbar}\int d^dx\int dt\left[\tfrac{1}{2}\,\chi\left(\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\right)^2+\frac{g_0\mu_{\rm B}\chi}{\hbar S}\,\boldsymbol{H}_{\rm u}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}}+\tfrac{1}{2}\,\chi\left(\frac{g_0\mu_{\rm B}}{\hbar S}\right)^2(\boldsymbol{H}_{\rm u}\times\hat{\boldsymbol{n}})^2\right]\right\}
\end{aligned} \tag{15.146}$$

where $\mathcal{I}_0$ and $\tilde{\mathcal{I}}_0$ are independent of $\boldsymbol{H}_{\rm u}$ and $\hat{\boldsymbol{n}}$, and where we have suppressed the $\boldsymbol{x}$ coordinate. The complete action functional, including the geometric phase term, is then

$$\mathcal{A} = \frac{1}{\hbar}\int d^dx\int dt\left\{\tfrac{1}{2}\,\chi\left(\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\right)^2 - \tfrac{1}{2}\,\rho_{\rm s}(\vec\nabla\hat{\boldsymbol{n}})^2 + \frac{g_0\mu_{\rm B}\chi}{\hbar S}\,\boldsymbol{H}_{\rm u}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\hat{\boldsymbol{n}}\right. \\ \left. + \tfrac{1}{2}\,\chi\left(\frac{g_0\mu_{\rm B}}{\hbar S}\right)^2(\boldsymbol{H}_{\rm u}\times\hat{\boldsymbol{n}})^2 - \frac{g_0\mu_{\rm B}}{v_0}\boldsymbol{H}_{\rm s}\cdot\hat{\boldsymbol{n}}\right\} - S\sum_i\eta_i\,\omega[\hat{\boldsymbol{n}}_i] \quad . \tag{15.147}$$

Dimensional analysis reveals the spin wave velocity $c=\sqrt{\rho_{\rm s}/\chi}$. Defining $x^0=ct$, we find that the quantum field theoretic action, excluding the geometric phase term, is

$$\mathcal{A} = \frac{\rho_{\rm s}}{2\hbar c}\int d^{d+1}x\left\{(\partial_\mu n^a)(\partial^\mu n^a) + \frac{2g_0\mu_{\rm B}}{\hbar cS}\,\boldsymbol{H}_{\rm u}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial x^0}\times\hat{\boldsymbol{n}}\right. \\ \left. + \left(\frac{g_0\mu_{\rm B}}{\hbar cS}\right)^2(\boldsymbol{H}_{\rm u}\times\hat{\boldsymbol{n}})^2 - \frac{2g_0\mu_{\rm B}}{\rho_{\rm s}v_0}\boldsymbol{H}_{\rm s}\cdot\hat{\boldsymbol{n}}\right\} - S\sum_i\eta_i\,\omega[\hat{\boldsymbol{n}}_i] \quad .$$

where we adopt a Minkowski $(+,-,\ldots,-)$ metric. The Euclidean version is

$$\mathcal{A}^{\rm E} = \frac{\rho_{\rm s}}{2\hbar c}\int d^{d+1}x\left\{(\partial_\mu n^a)(\partial_\mu n^a) + \frac{2ig_0\mu_{\rm B}}{\hbar cS}\,\boldsymbol{H}_{\rm u}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial x^0}\times\hat{\boldsymbol{n}}\right. \\ \left. - \left(\frac{g_0\mu_{\rm B}}{\hbar cS}\right)^2(\boldsymbol{H}_{\rm u}\times\hat{\boldsymbol{n}})^2 + \frac{2g_0\mu_{\rm B}}{\rho_{\rm s}v_0}\boldsymbol{H}_{\rm s}\cdot\hat{\boldsymbol{n}}\right\} + iS\sum_i\eta_i\,\omega[\hat{\boldsymbol{n}}_i] \quad . \tag{15.148}$$

Notice the factor of i in the coefficient of the second term. To maximize the weight $\exp(-\mathcal{A}^{\rm E})$, the third term inside the brackets should be as large as possible. This favors a *spin flop* in which the Néel vector lies perpendicular to the (uniform) applied magnetic field $\boldsymbol{H}_{\rm u}$.

The *coupling constant* for the nonlinear sigma model is defined to be

$$g = \frac{\hbar c}{\rho_{\rm s}} = \frac{\hbar}{\sqrt{\rho_{\rm s}\chi}} = \frac{\sqrt{2d}\,v_0}{aS}\left(\frac{\sum_{i,j} J_{ij}\,(1-\eta_i\eta_j)}{\sum_{i,j} J_{ij}\,(-\eta_i\eta_j)\,|\boldsymbol{R}_i-\boldsymbol{R}_j|^2/a^2}\right)^{1/2} \quad . \tag{15.149}$$

For a nearest neighbor model on a d-dimensional cubic lattice, we have[6]

$$g = \frac{2\sqrt{d}\,a^{d-1}}{S} \quad . \tag{15.150}$$

15.5.4 Continuum limit of the geometric phase: $d = 1$

In one space dimension, we have

$$\sum_j (-1)^j\,\omega[\hat{\boldsymbol{n}}_j] = \omega[\hat{\boldsymbol{n}}_0] - \omega[\hat{\boldsymbol{n}}_1] + \omega[\hat{\boldsymbol{n}}_2] - \ldots = \tfrac{1}{2}\int_0^L dx\,\frac{\partial\omega}{\partial x} \quad . \tag{15.151}$$

We now invoke eqn. 15.88, which says

$$\delta\omega = \int_0^T dt\,\epsilon_{abc}\,\dot{n}^b\,n^c\,\delta n^a \quad , \tag{15.152}$$

to obtain the beautiful result,

$$\mathcal{A}_{\rm B} = -S\sum_j (-1)^j\,\omega[\hat{\boldsymbol{n}}_j] = \tfrac{1}{2}\,S\int dx\int dt\;\hat{\boldsymbol{n}}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\frac{\partial\hat{\boldsymbol{n}}}{\partial x} \equiv 2\pi S\,Q_{tx} \quad , \tag{15.153}$$

where Q_{tx} is an integer topological invariant, known as the *Pontrjagin index* of the field $\hat{n}(x,t)$:

$$Q_{tx} = \frac{1}{8\pi}\int d^2x\,\epsilon^{\mu\nu}\,\epsilon_{abc}\,n^a\,\partial_\mu n^b\,\partial_\nu n^c \quad , \tag{15.154}$$

where $x^0 = ct$ as before. Q_{tx} measures the winding of the field $\hat{\boldsymbol{n}}(x,t)$ over the unit sphere. To see it is an integer, change variables from local coordinates (n^b, n^c) to (ξ_0, ξ_1) in the vicinity of $\hat{\boldsymbol{n}}$. The differential surface area element projected along n^a is

$$d\Sigma_a = \tfrac{1}{2}\,\epsilon^{\mu\nu}\,\epsilon_{abc}\,\frac{\partial n^b}{\partial\xi^\mu}\,\frac{\partial n^c}{\partial\xi^\nu}\,d^2\xi \tag{15.155}$$

[6]Take care not to confuse the coupling g with the g-factor g_0.

and changing variables from (x^0, x^1) to (ξ^0, ξ^1), we obtain

$$Q_{tx} = \frac{1}{4\pi} \int\limits_{\mathcal{S}^2_{\rm int}} d^2\xi \quad , \tag{15.156}$$

which is manifestly an integer.

Put another way, think of $\hat{n}(x,t)$ as a rubber band draped over the surface of a sphere. As time evolves from 0 to T, the configuration of the rubber band changes, but if the configuration itself is periodic, *i.e.* $\hat{\boldsymbol{n}}(x,0) = \hat{\boldsymbol{n}}(x,T)$, The Pontrjagin index measures the number of times the rubber band winds around the sphere. Configurations of $\hat{\boldsymbol{n}}(x,t)$ which yield a nonzero value of Q_{tx} are known as *skyrmions*. An example of a skyrmion configuration on the two-dimensional (x,y) (or (x,t)) plane is obtained by identifying the vector $\hat{\boldsymbol{n}}(x,y)$ with the (inverse) stereographically projected position (x,y). Put another way, we set

$$\frac{v}{u} = \tan(\theta/2)\, e^{i\phi} \equiv (x+iy)/a \quad , \tag{15.157}$$

where a is an arbitrary length scale. This (*exercise!*) is equivalent to

$$n_x = \frac{2ax}{a^2+x^2+y^2} \quad , \quad n_y = \frac{2ay}{a^2+x^2+y^2} \quad , \quad n_z = \frac{a^2-x^2-y^2}{a^2+x^2+y^2} \quad . \tag{15.158}$$

This skyrmion has Pontrjagin index $Q_{xy} = 1$.

Thermodynamic properties are derived from the Euclidean action,

$$\mathcal{A}^{\rm E}_{d=1} = 2\pi i S\, Q_{tx} + \frac{\rho_{\rm s}}{2\hbar c} \int d^2x\, (\vec{\nabla}\hat{\boldsymbol{n}})^2 \quad . \tag{15.159}$$

The effect of the geometric phase term, then, is quite simple and in fact discrete:

$$e^{2\pi i S\, Q_{tx}} = \begin{cases} +1 & \text{if } S \in \boldsymbol{Z} \\ (-1)^{Q_{tx}} & \text{it } S \in \boldsymbol{Z} + \frac{1}{2} \end{cases} \quad . \tag{15.160}$$

Thus, for integer S, the geometric phase term always contributes a factor of unity, and the full quantum field theoretic action is that of the two-dimensional O(3) model, also called the nonlinear sigma model. For half-odd integer S, space-time configurations with even and odd Pontrjagin index destructively interfere with each other.

What have we learned? First of all, we conclude that antiferromagnetic Heisenberg chains generically fall into two classes: those with integer spin and those with half-odd integer spin. The field theory for the first class is simply that of the classical O(3) model in two dimensions. The Hohenberg-Mermin-Wagner theorem precludes any spontaneous breaking of the continuous O(3) symmetry in $d = 2$ at any finite value of $\rho_{\rm s}$. The system has a gap, and correlation functions decay exponentially, up to power law corrections, *viz.*

$$\big\langle\, \Psi_0 \,\big|\, \boldsymbol{S}_0 \cdot \boldsymbol{S}_j \,\big|\, \Psi_0 \,\big\rangle \simeq (-1)^j\, |j|^{-1/2}\, \exp(-|j|/\xi) \quad , \tag{15.161}$$

where the correlation length ξ, in units of the lattice spacing a, is a function of the dimensionless quantity $\rho_{\rm s}/\hbar c$.

For the second class – the half-odd integer antiferromagnetic chains – the field theory includes the so-called 'θ-term',

$$\mathcal{A}_\theta = \frac{\theta}{4\pi}\int dx \int dt\, \hat{\boldsymbol{n}}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial t}\times\frac{\partial\hat{\boldsymbol{n}}}{\partial x} \quad , \tag{15.162}$$

with $\theta = 2\pi S = \pi \mod 2\pi$. While no exact solution to the field theory with the θ-term is yet known, we nonetheless conclude that all half-odd integer antiferromagnetic chains behave equivalently, since they all map onto the same model. Since the $S = \frac{1}{2}$ Heisenberg antiferromagnetic chain is known, from Bethe's *Ansatz*, to possess a disordered ground state with gapless excitations and power law correlations, $\langle \boldsymbol{S}_0\cdot\boldsymbol{S}_j\rangle \sim (-1)^j/|j|$ (up to logarithmic corrections), we conclude that the same is true for the $S = \frac{3}{2}, \frac{5}{2}$, *etc.* spin chains.

15.5.5 The geometric phase in higher dimensions

So long as the Néel field $\hat{\boldsymbol{n}}(\boldsymbol{x},t)$ is a smooth function of space and time, there are no interesting topological terms in the field theory in more than one space dimension. The reason is trivial. Consider a d-dimensional system as a network of parallel one-dimensional chains. Call the longitudinal (chain) coordinate x. For each set of transverse coordinates $\boldsymbol{R}_\perp$, one can define the integer Pontrjagin index $Q_{tx}(\boldsymbol{R}_\perp)$. The geometric phase term in the action is then given by

$$\mathcal{A}_{\rm B} = S\sum_i \eta_i\,\omega[\hat{\boldsymbol{n}}_i] = S\sum_{\boldsymbol{R}_\perp}\eta_{\boldsymbol{R}_\perp}\,Q_{tx}(\boldsymbol{R}_\perp) = 0 \quad , \tag{15.163}$$

where the last equality follows from the assumed smoothness of $\hat{\boldsymbol{n}}(\boldsymbol{x},t)$, which requires that $Q_{tx}(\boldsymbol{R}_\perp)$ be independent of $\boldsymbol{R}_\perp$, since a smooth integer-valued function must be a constant!

When the smoothness constraint is relaxed, however, the geometric phase term can play an important role. For a two-dimensional antiferromagnet, there exist topology-changing instanton for which $\Delta Q_{xy} = \pm 1$. Such field configurations are called 'hedgehogs', because the direction of the field $\hat{\boldsymbol{n}}(t,x,y)$ points radially outward from the center of the hedgehog. For quantum-disordered two-dimensional antiferromagnets (*i.e.* small $\rho_{\rm s}$), Haldane argued that geometrical phase considerations associated with the presence of hedgehogs would distinguish not only between integer and half-odd integer S on the square lattice, but between even and odd integer S as well.

15.6 Large-N Techniques

The basic idea behind large-N approaches is to extend the global symmetry group of some physical model from *e.g.* O(3), SU(2), *etc.* to a larger group, such as O(N), SU(N), or Sp(N). If the

extension is done in a certain way, the resultant model can be solved exactly in the $N \to \infty$ limit. N plays the role of $1/\hbar$, so $N \to \infty$ is a classical limit of sorts, with no quantum fluctuations. Furthermore, one can derive a systematic diagrammatic expansion in powers of $1/N$, which can be used to investigate properties at finite N.

We shall barely scratch the surface of this subject. My aim here is to guide you through a large-N calculation for the nonlinear sigma model.

15.6.1 $1/N$ expansion for an integral

To begin, consider the one-dimensional integral,

$$\mathcal{I} = \int\limits_{-\infty}^{\infty} dx\, e^{-Nf(x)} \quad , \tag{15.164}$$

where $f(x)$ is some function and N is large. Clearly the integral is dominated by values of x near the minimum of $f(x)$. Suppose a unique global minimum exists at $x = x_{\rm c}$. We can then write

$$\begin{aligned}
\mathcal{I} &= e^{-Nf(x_{\rm c})} \int\limits_{-\infty}^{\infty} du\, e^{-\frac{1}{2}Nf''(x_{\rm c})\,u^2}\, e^{-\frac{1}{6}Nf'''(x_{\rm c})\,u^3}\, e^{-\frac{1}{24}Nf''''(x_{\rm c})\,u^4} \cdots \\
&= e^{-Nf(x_{\rm c})} \int\limits_{-\infty}^{\infty} du\, e^{-\frac{1}{2}Nf''(x_{\rm c})\,u^2} \Big\{ 1 - \tfrac{1}{6}Nf'''(x_{\rm c})\,u^3 - \tfrac{1}{24}Nf''''(x_{\rm c})\,u^4 + \ldots \Big\} \\
&= \left(\frac{2\pi}{Nf''(x_{\rm c})} \right)^{1/2} e^{-Nf(x_{\rm c})} \Big\{ 1 - \tfrac{1}{24}Nf''''(x_{\rm c}) \langle u^4 \rangle + \ldots \Big\} \quad .
\end{aligned} \tag{15.165}$$

Thus, we have derived a $1/N$ expansion for the integral:

$$-\ln \mathcal{I} = \overbrace{Nf(x_{\rm c})}^{\mathcal{O}(N^1)} + \overbrace{\tfrac{1}{2}\ln\Big(\frac{Nf''(x_{\rm c})}{2\pi}\Big)}^{\mathcal{O}(N^0)} + \overbrace{\frac{1}{8N}\frac{f''''(x_{\rm c})}{\big[f''(x_{\rm c})\big]^2}}^{\substack{\mathcal{O}(N^{-1}) \\ \text{corrections}}} + \mathcal{O}(N^{-2}) \quad . \tag{15.166}$$

15.6.2 Large-N theory of the nonlinear sigma model

Recall the Euclidean action for the O(3) nonlinear sigma model,

$$\mathcal{A}_{\rm E} = \frac{\rho_{\rm s}}{2\hbar c} \int d^d x \int\limits_0^{L_0} dx^0\, (\partial_\mu n^a)^2 \quad , \tag{15.167}$$

where $\hat{\boldsymbol{n}} = (n^x, n^y, n^z)$ is a three-component unit vector and $L_0 = \hbar c/k_{\rm B}T$. In the case of Haldane's derivation of the sigma model action for quantum antiferromagnets, $\hat{n}(x)$ is physically the Néel field, which varies slowly from site to site even though the local magnetization itself oscillates from one sublattice to the next.[7] It should be emphasized, though, that the D-dimensional nonlinear sigma model also describes the finite temperature phase transition of an isotropic D-dimensional ferromagnet.

Quantum mechanics is irrelevant at finite temperature, since the imaginary time variable is bounded: $0 \leqslant \tau \leqslant \hbar\beta$. At a critical point, the spatial correlation length diverges as $\xi(T) \sim |T - T_{\rm c}|^{-\nu}$, and the temporal correlation length (or correlation time) diverges along with ξ, as $\xi_\tau \sim \xi^z$. Here, ν is the correlation length exponent and z the dynamic critical exponent. With $\hbar\beta$ finite, however, sufficiently close to $T_{\rm c}$ the correlation time exceeds the thickness $\hbar\beta$ of the temporal 'slab', hence the degrees of freedom at a particular location in space are 'locked' as a function of imaginary time. Finite T second order transitions of a d-dimensional quantum system are therefore described by a d-dimensional action.[8] At zero temperature, though, the temporal slab is infinitely thick, and one cannot ignore temporal fluctuations. The action is then for a $(d+1)$-dimensional system.

It is perhaps worth emphasizing that the continuum effective action for the Heisenberg *ferromagnet* is given by

$$\mathcal{A}_{\rm FM} = \int d^dx \int_0^{\hbar\beta} d\tau \left\{ iS\, v_0^{-1}\, \boldsymbol{A}(\hat{\boldsymbol{n}}) \cdot \frac{\partial \hat{\boldsymbol{n}}}{\partial \tau} + \tfrac{1}{2}\,\rho_{\rm s}\,(\vec{\nabla}\hat{\boldsymbol{n}})^2 \right\} \tag{15.168}$$

where v_0 is the unit cell volume, and

$$\rho_{\rm s} = \frac{S^2}{4dv_0} \sum_{\boldsymbol{R}} J(\boldsymbol{R})\, \boldsymbol{R}^2 \quad . \tag{15.169}$$

Note the difference between this and the effective action of the antiferromagnet, in which space and time appear symmetrically. The effective (low-energy) theory for the antiferromagnet possesses a 'Lorentz invariance' where the speed of light is replaced by the spin wave velocity $c = \sqrt{\rho_{\rm s}/\chi}$.

Returning to the nonlinear sigma model, the partition function is given by the functional integral

$$Z = e^{-F/k_{\rm B}T} = \int\limits_{\hat{\boldsymbol{n}}^2=1} \mathcal{D}\hat{\boldsymbol{n}}(x)\, e^{-\mathcal{A}_{\rm E}[\hat{\boldsymbol{n}}]} \quad . \tag{15.170}$$

[7] The notation I adopt here is that $(d+1)$-dimensional vectors are denoted as $x \equiv (x^0, \boldsymbol{x})$.

[8] Note that this *does not* say that quantum mechanics has no effect whatsoever at finite temperature. Indeed, the partition function for the quantum and classical Heisenberg models will be different. What *is* true is that the *critical properties* at a finite temperature *second order* transition are not affected by quantum mechanics.

The extension of the O(3) model to one with an O(N) symmetry is trivial. Simply replace the 3-component unit vector (n_x, n_y, n_z) by an N-component one, $\boldsymbol{n} = (n_1, n_2, \ldots, n_N)$. How do we generalize the unit length constraint to general N? Let us write the constraint as $\boldsymbol{n}^2(x) = qN$, where the parameter q is as yet undetermined. We can envisage two natural extensions to general N:

- Maintain $\boldsymbol{n}^2 = 1$, *i.e.* take $q = N^{-1}$.
- Fix q and let N vary. The length of $\boldsymbol{n}$ then increases with N.

It turns out that it is the second of these schemes which generates a proper $1/N$ expansion, as we shall soon see.

To enforce the length constraint, we insert into the functional integral a δ-function $\delta(\hat{\boldsymbol{n}}^2 - qN)$ at every space-time point. We write the δ-function as

$$\delta(y) = \int\limits_{-i\infty}^{i\infty} d\lambda \, e^{-\lambda y} \quad , \tag{15.171}$$

where the integration contour runs along the imaginary axis, from $-i\infty$ to $+i\infty$. The partition function is then expressed as a double functional integral over the fields $\hat{\boldsymbol{n}}(x)$ and $\lambda(x)$,

$$Z = \int \mathcal{D}[\boldsymbol{n}(x), \lambda(x)] \, e^{-\tilde{\mathcal{A}}_{\rm E}[\boldsymbol{n},\lambda]} \quad , \tag{15.172}$$

where

$$\tilde{\mathcal{A}}_{\rm E} = \int d^{d+1}x \left\{ \frac{1}{2g} (\partial_\mu n^a)^2 + \lambda \left(\boldsymbol{n}^2 - qN \right) \right\} \quad . \tag{15.173}$$

For convenience we have defined the coupling

$$g \equiv \frac{\hbar c}{\rho_{\rm s}} = \frac{\hbar}{\sqrt{\chi \rho_{\rm s}}} \quad . \tag{15.174}$$

The dimensions of g are $[g] = L^{d-1}$.

We now integrate out the $n^a(x)$ fields, which are quadratic in $\tilde{\mathcal{A}}_{\rm E}$. Writing

$$\tilde{\mathcal{A}}_{\rm E} = \tfrac{1}{2} \int d^{d+1}x \int d^{d+1}x' \, n^a(x) \, K(x, x') \, n^a(x') - qN \int d^{d+1}x \, \lambda(x) \tag{15.175}$$

with

$$K(x, x') = -\frac{\rho_{\rm s}}{\hbar c} \frac{\partial}{\partial x^\mu} \, \delta(x - x') \, \frac{\partial}{\partial x'^\mu} + 2\lambda \, \delta(x - x') \quad , \tag{15.176}$$

the partition function can be written in terms of an effective free energy which is a function of the field $\lambda(x)$ alone:

$$Z = \int \mathcal{D}\lambda(x)\, e^{-NF_{\rm eff}[\lambda]/k_{\rm B}T} \tag{15.177}$$

where

$$\begin{aligned} e^{-NF_{\rm eff}[\lambda]/k_{\rm B}T} &= \int \mathcal{D}\boldsymbol{n}(x)\, e^{-\tilde{\mathcal{A}}_{\rm E}[\boldsymbol{n}(x),\lambda(x)]} \\ &= (\det\ K)^{-N}\ \exp\left\{qN\int d^{d+1}x\,\lambda(x)\right\} \quad . \end{aligned} \tag{15.178}$$

Thus, the effective free energy is

$$F_{\rm eff}[\lambda]/k_{\rm B}T = \ln\det\ K - q\int d^{d+1}x\,\lambda(x) \quad , \tag{15.179}$$

where the determinant of the integral operator K is, as always, defined by the product of its eigenvalues,

$$\ln\det\ K = \prod_n \zeta_n \quad . \tag{15.180}$$

The eigenvalue equation is

$$\int d^{d+1}x'\, K(x,x')\,\psi_n(x') = \zeta_n\,\psi_n(x) \quad . \tag{15.181}$$

We can now see why keeping q finite as $N\to\infty$ generates a true $1/N$ expansion. Had we instead taken $\boldsymbol{n}^2 = 1$, we would have $q = 1/N$ and the effective free energy $F_{\rm eff}[\lambda]$ would not be independent of N.

Solution of the $N\to\infty$ theory

When $N\to\infty$ the functional integral is dominated by the saddle point in the action. Before we solve for this saddle point, let us slightly extend our model to include a coupling to a magnetic field. The augmented action is then

$$\tilde{\mathcal{A}}_{\rm E} = \int d^{d+1}x\left\{\frac{1}{2g}(\partial_\mu n^a)^2 + \lambda\,(n^a n^a - qN) - \sqrt{N}\,h^a\,n^a\right\} \quad . \tag{15.182}$$

The $\sqrt{N}$ factor preceding $\boldsymbol{h}\cdot\boldsymbol{n}$ ensures that the action will be proportional to N when $|\boldsymbol{h}|$ is of $\mathcal{O}(N^0)$. In the case of the antiferromagnet, where $\hat{\boldsymbol{n}}$ is the Néel field, $\boldsymbol{h}$ corresponds to the $\boldsymbol{q} = \boldsymbol{\pi}/a$ (zone corner) component of the physical magnetic field, *i.e.*a sublattice-staggered magnetic field. This is of course quite unphysical, however our purpose in introducing $\boldsymbol{h}$ is not

to investigate the effects of an external field *per se*, but rather as an artifice by which we can couple to any condensate, as we shall see presently.

To find the saddle point of $F_{\rm eff}[\lambda]$, we should set its functional variation with respect to $\lambda(x)$ to zero. We will assume that the saddle point occurs for real, constant λ. We will justify this by presenting such a solution to the equation $\delta F_{\rm eff} = 0$. Note that the saddle point lies off the integration contour for $\lambda(x)$, which runs along the imaginary axis.

When λ is constant, the model may be solved by Fourier transform. We write

$$n^a(x) = \frac{1}{\sqrt{L_0 V}} \sum_k \hat{n}^a(k)\, e^{ik\cdot x} \tag{15.183}$$

with

$$L_0 = \beta\hbar c \quad , \quad V = L_1 \cdots L_d \quad , \quad k = \Big(\frac{2\pi j_0}{L_0}, \frac{2\pi j_1}{L_1}, \ldots, \frac{2\pi j_d}{L_d}\Big) \quad . \tag{15.184}$$

Expressed in terms of the Fourier modes,

$$\begin{Bmatrix} n^a(x) \\ h^a(x) \end{Bmatrix} \equiv \frac{1}{\sqrt{V L_0}} \sum_k \begin{Bmatrix} \hat{n}^a(h) \\ \hat{h}^a(k) \end{Bmatrix} e^{ik\cdot x} \quad , \tag{15.185}$$

the Euclidean action is

$$\tilde{\mathcal{A}}_{\rm E} = \sum_{a=1}^{N} \sum_k \left\{ \Big(\lambda + \frac{k^2}{2g}\Big) \big|\hat{n}^a(k)\big|^2 - \sqrt{N}\, \hat{h}^*_a(k)\, \hat{n}^a(k) \right\} - qNVL_0\,\lambda \quad . \tag{15.186}$$

We now integrate out the $\{\hat{n}^a(k)\}$, yielding an effective free energy *function* $F_{\rm eff}(\lambda)$:

$$f(\lambda) \equiv \frac{F_{\rm eff}(\lambda)}{N\hbar c V} = -q\lambda + \frac{1}{2L_0 V} \sum_k \ln\Big(\lambda + \frac{k^2}{2g}\Big) - \frac{1}{2L_0 V} \sum_k \frac{\hat{h}^a(k)\, \hat{h}^a(-k)}{\big(\lambda + \frac{k^2}{2g}\big)} \quad . \tag{15.187}$$

The order parameter $\boldsymbol{m}$, which is the static Néel field in the case of the antiferromagnet and the static magnetization in the case of the ferromagnet, is obtained by differentiating the free energy with respect to the $\boldsymbol{q} = 0$ Fourier component of the field $\hat{h}^a(k)$. We therefore obtain

$$\boldsymbol{m} = \frac{\langle \boldsymbol{n} \rangle}{\sqrt{N}} = -\frac{1}{NL_0 V} \frac{\partial (NF_{\rm eff}/k_{\rm B}T)}{\partial \boldsymbol{h}} = -\frac{\partial f}{\partial \boldsymbol{h}} = \frac{\boldsymbol{h}}{2\lambda} \quad , \tag{15.188}$$

since $\hat{\boldsymbol{h}}(0) = \sqrt{L_0 V}\, \boldsymbol{h}$.

To find the saddle point in λ, we set $\partial f/\partial\lambda = 0$, yielding

$$q = \boldsymbol{m}^2 + \frac{g}{L_0 V} \sum_k \frac{1}{k^2 + 2g\lambda} \quad . \tag{15.189}$$

In the absence of an external field, we also have The second mean field equation,

$$2\lambda\,\boldsymbol{m} = \boldsymbol{h} = 0 \quad . \tag{15.190}$$

This requires either (i) $\lambda = 0$ or (ii) $\boldsymbol{m} = 0$.

We now explore the solution to these equations as we vary dimensionality and temperature.

- $d = 1$, $T = 0$: In this case the integral is infrared divergent when $\lambda = 0$. The mean field equation can always be solved with $\boldsymbol{m} = 0$ for some finite λ:

$$q = \int^{\Lambda} \frac{d^2k}{(2\pi)^2} \frac{1}{2\lambda + \frac{k^2}{g}} = \frac{g}{4\pi}\,\ln\left(1 + \frac{\Lambda^2}{2g\lambda}\right) \quad , \tag{15.191}$$

  yielding

$$\lambda(g) = \frac{\Lambda^2/2g}{\exp(4\pi q/g) - 1} \quad , \tag{15.192}$$

  which monotonically decreases on $g \in [0.\infty]$ from $\lambda(0) = \Lambda^2/8\pi q$ to $\lambda(\infty) = 0$.

- $d > 1$, $T = 0$: In this case there exists a quantum critical point at $g = g_{\rm c}$. The gap λ vanishes for $g \leqslant g_{\rm c}$. To find $g_{\rm c}$, set

$$q = g_{\rm c} \int^{\Lambda} \frac{d^{d+1}k}{(2\pi)^{d+1}}\,\frac{1}{k^2} = \frac{g_{\rm c}}{(2\pi)^{d+1}} \cdot \frac{\Lambda^{d-1}}{d-1} \cdot \begin{cases} \Omega_{d+1} & \text{scheme I} \\ \pi\Omega_d & \text{scheme II} \end{cases} , \tag{15.193}$$

  where Ω_d is the area of the d-dimensional unit sphere:

$$\Omega_d = \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \quad . \tag{15.194}$$

  The cutoff is taken to be isotropic in both frequency and momentum (scheme I) or isotropic in momentum only (scheme II). In scheme II, the integral over the frequency component k_0 extends over the range $(-\infty, \infty)$, which is appropriate since the imaginary time variable is not quantized on a lattice. This gives us an equation for the critical coupling $g_{\rm c}$. The cutoff Λ is proportional to a^{-1}, and it is convenient to write $\Lambda = \zeta\pi/a$, where ζ is a dimensionless constant and a is the lattice spacing.

  Recall that Haldane's mapping for the cubic lattice Heisenberg model resulted in an $\mathrm{O}(3)$ nonlinear sigma model with coupling $g = 2\sqrt{d}\,a^{d-1}/S$. The critical value for the spin quantum number $S_{\rm c}$ is then found to be

$$S_{\rm c} = \frac{\sqrt{d}}{d-1} \cdot \frac{\zeta^{d-1}}{2^d\pi^2 q} \cdot \begin{cases} \Omega_{d+1} & \text{scheme I} \\ \pi\Omega_d & \text{scheme II} \end{cases} , \tag{15.195}$$

For $d = 2$ and $q = N^{-1} = \frac{1}{3}$, one finds $S_{\rm c} = 1.35\,\zeta$ (scheme I) or $S_{\rm c} = 2.12\,\zeta$ (scheme II). Depending on the value of ζ, then, the critical S may be either smaller or greater than the smallest value permitted by quantum mechanics, *i.e.* $S = \frac{1}{2}$. If $S_{\rm c} < \frac{1}{2}$, then we conclude the model is Néel-ordered at zero temperature. Indeed numerical work convincingly shows that the ground state for $S = \frac{1}{2}$ is Néel-ordered, and rigorous proofs exist which show that long-ranged Néel order exists for $S \geqslant 1$.

One might suspect, given eqn. 15.149, that by extending the range of the interactions one can push g above $g_{\rm c}$ and obtain a quantum-disordered 'spin-liquid' ground state for the $S = \frac{1}{2}$ antiferromagnet on a square lattice. For example, if one includes next-nearest neighbor antiferromagnetic coupling J_2 as well as nearest neghibor antiferromagnetic coupling J_1, one has

$$g = \frac{2\sqrt{d}\,a^{d-1}}{S} \cdot \frac{1}{\sqrt{1 - 2J_2/J_1}} \quad , \tag{15.196}$$

which is increased above its value when $J_2 = 0$. In fact, the search for spin liquid states has been an arduous one. On the square lattice, one generally finds that frustrating further-neighbor couplings push the system into another ordered state, for example one with four sublattice antiferromagnetic order. On lattices which are highly geometrically frustrated, such as the Kagomé and pyrochlore lattices, the $S = \frac{1}{2}$ antiferromagnet is generally believed to have a quantum-disordered spin liquid ground state, *i.e.* the ground state has no long-ranged order and breaks no lattice translation or point group symmetries.

- $d > 2$, $T > 0$: In this case there is a finite temperature phase transition. Defining the Matsubara wavevectors $\kappa_n \equiv 2\pi n/L_0$, we use the result

$$\frac{1}{L_0}\sum_{\kappa_n} H(-i\kappa_n) = \frac{H(0)}{L_0} + \int_{-\infty}^{\infty}\frac{d\kappa}{\pi}\,\frac{\mathsf{Im}\;H(\kappa + i0^+)}{\exp(\kappa L_0) - 1} \tag{15.197}$$

to obtain the finite temperature mean field equation,

$$q = \boldsymbol{m}^2 + g\int_{\Lambda}\frac{d^dk}{(2\pi)^d}\left\{\frac{2/L_0}{k^2 + 2g\lambda} + \frac{\mathrm{ctnh}\,(\frac{1}{2}L_0\sqrt{k^2 + 2g\lambda})}{\sqrt{k^2 + 2g\lambda}}\right\} \tag{15.198}$$

The equation for $T_{\rm c}$ is obtained by setting $\lambda = \boldsymbol{m}^2 = 0$:

$$q = g\int_{\Lambda}\frac{d^dk}{(2\pi)^d}\left\{\frac{2}{L_0\,k^2} + \frac{1}{k}\,\mathrm{ctnh}\,(\tfrac{1}{2}kL_{0,{\rm c}})\right\} \quad , \tag{15.199}$$

which is to be solved for $T_{\rm c} = \hbar c/k_{\rm B}L_{0.{\rm c}}$, assuming $g < g_{\rm c}$, *i.e.* that the $T = 0$ (ground) state is ordered.

15.6.3 Correlation functions

The correlation functions are obtained via

$$\langle \hat{n}^a(k)\,\hat{n}^b(-k)\rangle_{\rm c} = -L_0 V\,\frac{\partial^2 f}{\partial\hat{h}^a(-k)\,\partial\hat{h}^b(k)} = \frac{g}{k^2+2g\lambda}\,\delta^{ab} \tag{15.200}$$

hence the full correlator is given by

$$\begin{aligned} C_{ab}(x) \equiv \langle n^a(0)\,n^b(x)\rangle &= m^a m^b + \frac{g}{L_0 V}\sum_k \frac{e^{ik\cdot x}}{k^2+2g\lambda}\;\delta^{ab} \\ &= m^a m^b + g\int\limits_{-\infty}^{\infty}\frac{dk_0}{2\pi}\,e^{ik_0x^0}\int\limits^{\Lambda}\frac{d^dk}{(2\pi)^d}\,\frac{e^{i\boldsymbol{k}\cdot\boldsymbol{x}}}{k_0^2+\boldsymbol{k}^2+2g\lambda} \quad , \end{aligned} \tag{15.201}$$

where in the second line we take the thermodynamic limit, set $T\to 0$, and adopt cutoff scheme II, appropriate for lattice systems. In the latter case, at large distances we obtain the Ornstein-Zernike form,

$$C_{ab}(x) - C_{ab}(\infty) \sim \frac{e^{-|x|/\xi}}{|x|^{d/2}}\,\delta^{ab} \quad , \tag{15.202}$$

with $\xi = (2g\lambda)^{-1/2}$. At the quantum critical point, where λ vanishes, one finds $C(x)\sim|x|^{1-d}$.

Chapter 16

Notes on Line Graphs

16.1 Line graphs: Kagomé and Checkerboard Lattices

A *line graph* is constructed by taking a given graph and associating a site to every link in the graph. Two sites are connected if their corresponding links share a vertex in the original graph. To every closed loop of even perimeter on the original graph there corresponds an eigenstate of the adjacency matrix of the line graph, with eigenvalue $\lambda = 2$, provided those closed loops are all *even-membered*.

16.1.1 Kagomé lattice

The Kagomé lattice, depicted in fig. 16.1, is a triangular Bravais lattice with a three element basis. It is the line graph of the honeycomb lattice. Choosing primitive direct lattice vectors $\boldsymbol{a}_1 = \hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = \frac{1}{2}\hat{\boldsymbol{x}} + \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}$, we then write

$$\begin{pmatrix} a_{\boldsymbol{R}} \\ b_{\boldsymbol{R}} \\ c_{\boldsymbol{R}} \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \begin{pmatrix} a_{\boldsymbol{k}} \\ e^{\frac{i}{2}\boldsymbol{k}\cdot\boldsymbol{a}_1}\, b_{\boldsymbol{k}} \\ e^{\frac{i}{2}\boldsymbol{k}\cdot\boldsymbol{a}_2}\, c_{\boldsymbol{k}} \end{pmatrix} . \tag{16.1}$$

The basis vectors here are 0, $\frac{1}{2}\boldsymbol{a}_1$, and $\frac{1}{2}\boldsymbol{a}_2$.

With $\theta_i = \boldsymbol{k}\cdot\boldsymbol{a}_i$, we then have the Hamiltonian

$$H_{\boldsymbol{k}} = -\begin{pmatrix} 0 & 2tc_1 & 2tc_2 \\ 2tc_1 & 0 & 2tc_{12} \\ 2tc_2 & 2tc_{12} & 0 \end{pmatrix} , \tag{16.2}$$

where as before $c_i = \cos\frac{1}{2}\theta_i$ and $c_{ij} = \cos\frac{1}{2}(\theta_i - \theta_j)$. The eigenvalues are

$$E_{1,2} = -t \pm t\sqrt{3 + 2\cos\theta_1 + 2\cos\theta_2 + 2\cos(\theta_1 - \theta_2)} \tag{16.3}$$

and

$$E_3 = +2t\ . \tag{16.4}$$

Thus, there is a lower band with energies in the interval $E_1 \in [-4t\ ,\ -t]$, a middle band with $E_2 \in [-t\ ,\ +2t]$, and a flat upper band with $E_3 = +2t$.

16.1.2 Checkerboard lattice

The checkerboard lattice, also known as the planar pyrochlore structure, is depicted in fig. 16.2. It is the line graph of the square lattice. It may be represented as an underlying square lattice with primitive direct lattice vectors $\boldsymbol{a}_1 = \hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}$ and $\boldsymbol{a}_2 = \hat{\boldsymbol{x}} + \hat{\boldsymbol{y}}$, with a two element basis 0 and $\hat{\boldsymbol{x}}$.

$$\begin{pmatrix} a_{\boldsymbol{R}} \\ b_{\boldsymbol{R}} \end{pmatrix} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \begin{pmatrix} a_{\boldsymbol{k}} \\ e^{\frac{i}{2}\boldsymbol{k}\cdot\hat{\boldsymbol{x}}}\, b_{\boldsymbol{k}} \end{pmatrix}\ . \tag{16.5}$$

It the hoppings are t along $\pm\hat{\boldsymbol{x}}$ and $\pm\hat{\boldsymbol{y}}$, and t' along $\pm\boldsymbol{a}_1$ and $\pm\boldsymbol{a}_2$, then the Hamiltonian is

$$H_{\boldsymbol{k}} = -\begin{pmatrix} 2t'\cos\theta_2 & 4t\cos(\frac{1}{2}\theta_1)\cos(\frac{1}{2}\theta_2) \\ 4t\cos(\frac{1}{2}\theta_1)\cos(\frac{1}{2}\theta_2) & 2t'\cos\theta_1 \end{pmatrix}\ , \tag{16.6}$$

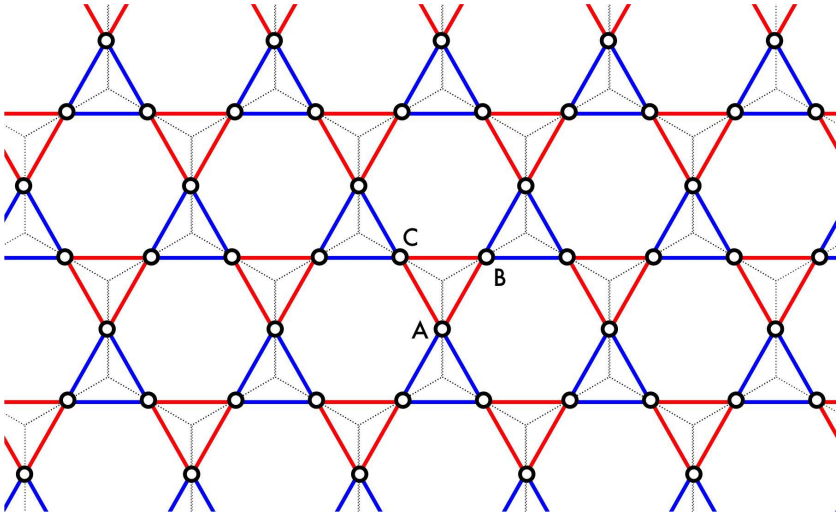


Figure 16.1: The Kagomé lattice, which is a triangular Bravais lattice with a three element basis (A,B,C). The Kagomé lattice is the line graph of a honeycomb lattice.

with eigenvalues

$$E_{1,2} = -t'\big(\cos\theta_1 + \cos\theta_2\big) \pm \sqrt{\big(t'\cos\theta_1 - t'\cos\theta_2\big)^2 + 4t^2\big(1+\cos\theta_1\big)\big(1+\cos\theta_2\big)}\ . \tag{16.7}$$

When $t = t'$ this simplifies to

$$E_1 = -2t\big(1 + \cos\theta_1 + \cos\theta_2\big) \qquad , \qquad E_2 = +2t\ . \tag{16.8}$$

Once again, as in the Kagomé and pyrochlore lattices, the top band is flat.

16.1.3 Square-octagon lattice line graph

The line graph of the square-octagon lattice is shown in Fig. 16.3. The Hamiltonian is

$$H_{\boldsymbol{k}} = -\begin{pmatrix} 0 & t_1 & t_2\,e^{-ik_x} & 0 & t_2\,e^{-ik_y} & t_1 \\ t_1 & 0 & t_2 & t_1 & t_2\,e^{-ik_y} & 0 \\ t_2\,e^{ik_x} & t_2 & 0 & t_2 & 0 & t_2\,e^{ik_x} \\ 0 & t_1 & t_2 & 0 & t_2 & t_1 \\ t_2\,e^{ik_y} & t_2\,e^{ik_y} & 0 & t_2 & 0 & t_2 \\ t_1 & 0 & t_2\,e^{-ik_x} & t_1 & t_2 & 0 \end{pmatrix}\ , \tag{16.9}$$

where t_1 is the hopping along the blue links and t_2 the hopping along the red links. When $t_1 \neq t_2$, the adjacency matrix has a flat band at $\lambda = 2t_1$. When $t_1 = t_2$, the flat band is doubly degenerate.

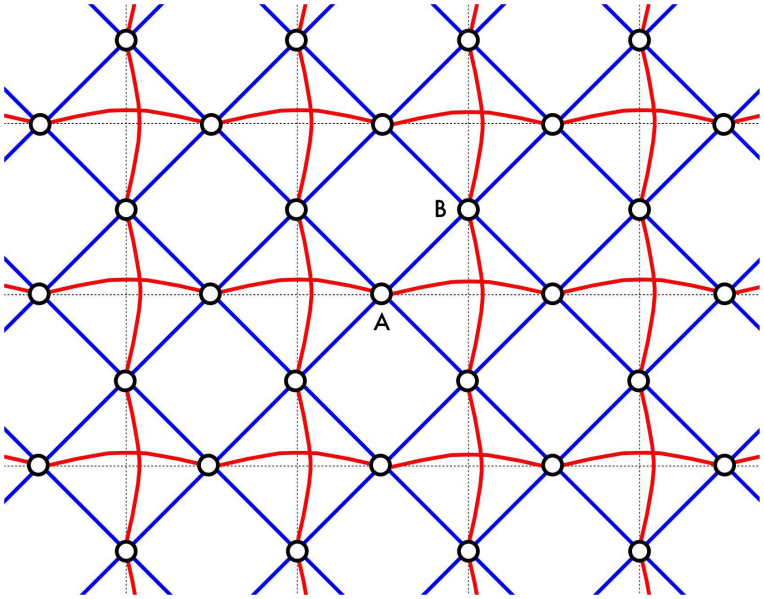


Figure 16.2: The checkerboard (planar pyrochlore) lattice, which is a square lattice with a two element basis (A,B). The checkerboard lattice is the line graph of the square lattice.

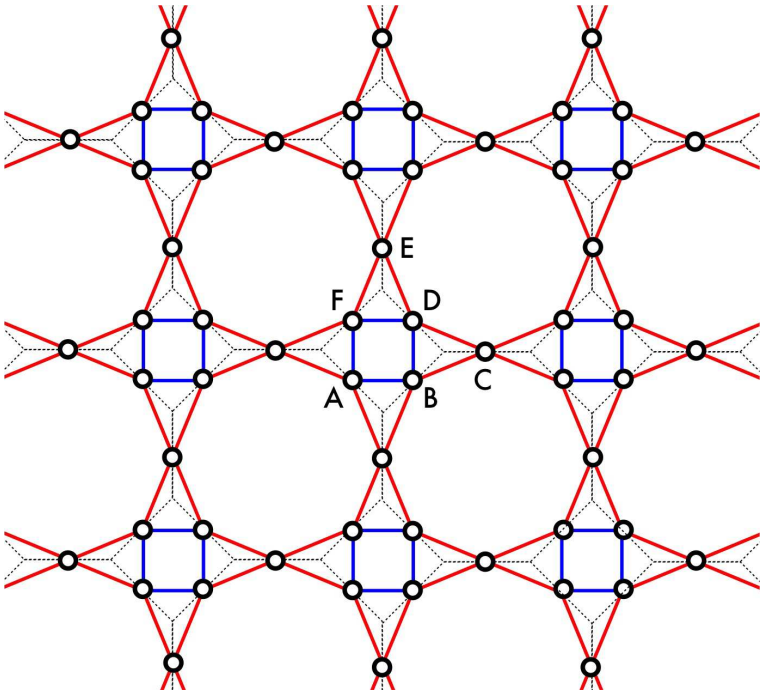


Figure 16.3: This fourfold coordinated lattice is the line graph of the square-octagon lattice. The hopping along the blue links is t_1 and that along the red links is t_2.

16.2 Pyrochlore Lattice

The pyrochlore lattice (fig. 16.5) is a sixfold-coordinated structure whose underlying Bravais lattice is FCC, with a four element basis. It is effectively described as a diamond lattice of 'corner-sharing tetrahedra'. Recall that diamond is FCC with a two element basis. It can be constructed as the line graph of the diamond lattice.

To apprehend the geometry better, consider the sketch in fig. 16.6. Let the center of the cube, indicated by the yellow star, be located at the origin $(0, 0, 0)$. Then the four corners of the tetrahedron are located as

$$\boldsymbol{A} = \tfrac{d}{2}\,(+1, -1, +1) \qquad \boldsymbol{C} = \tfrac{d}{2}\,(+1, +1, -1)$$
$$\boldsymbol{B} = \tfrac{d}{2}(-1, +1, +1) \qquad \boldsymbol{D} = \tfrac{d}{2}\,(-1, -1, -1)\ .$$

There are two species of tetrahedra in the pyrochlore lattice, which we call α and β (see figure 16.7). We identify the tetrahedron depicted in fig. 16.6 as an α-tetrahedron. The centers of the four neighboring β-tetrahedra are then located at $2\boldsymbol{A}$, $2\boldsymbol{B}$, $2\boldsymbol{C}$, and $2\boldsymbol{D}$. That is, to move from the center of an α-tetrahedron to the center of a neighboring β-tetrahedron, we displace by one of these four vectors. This set is not invariant under inversion. To move from the center of a β-tetrahedron to the center of a neighboring α-tetrahedron, we displace by the negative of one of these four vectors.

Thus, the twelve nearest neighbor displacements on the underlying FCC Bravais lattice are given by $2(\boldsymbol{A}-\boldsymbol{B})$, $2(\boldsymbol{A}-\boldsymbol{C})$, *etc.*, since we move from (the center of) one α-tetrahedron to another. These twelve vectors may be written as

$$2\boldsymbol{A}-2\boldsymbol{B}=\frac{a}{\sqrt{2}}\,(+1,-1,0)=-\boldsymbol{a}_1+\boldsymbol{a}_2 \tag{16.10}$$

$$\begin{aligned}
2\boldsymbol{A}-2\boldsymbol{C}&=\frac{a}{\sqrt{2}}\,(0,-1,+1)=\boldsymbol{a}_2-\boldsymbol{a}_3\\
2\boldsymbol{A}-2\boldsymbol{D}&=\frac{a}{\sqrt{2}}\,(+1,0,+1)=\boldsymbol{a}_2\\
2\boldsymbol{B}-2\boldsymbol{C}&=\frac{a}{\sqrt{2}}\,(-1,0,+1)=\boldsymbol{a}_1-\boldsymbol{a}_3\\
2\boldsymbol{B}-2\boldsymbol{D}&=\frac{a}{\sqrt{2}}\,(0,+1,+1)=\boldsymbol{a}_1\\
2\boldsymbol{C}-2\boldsymbol{D}&=\frac{a}{\sqrt{2}}\,(+1,+1,0)=\boldsymbol{a}_3
\end{aligned} \tag{16.11}$$

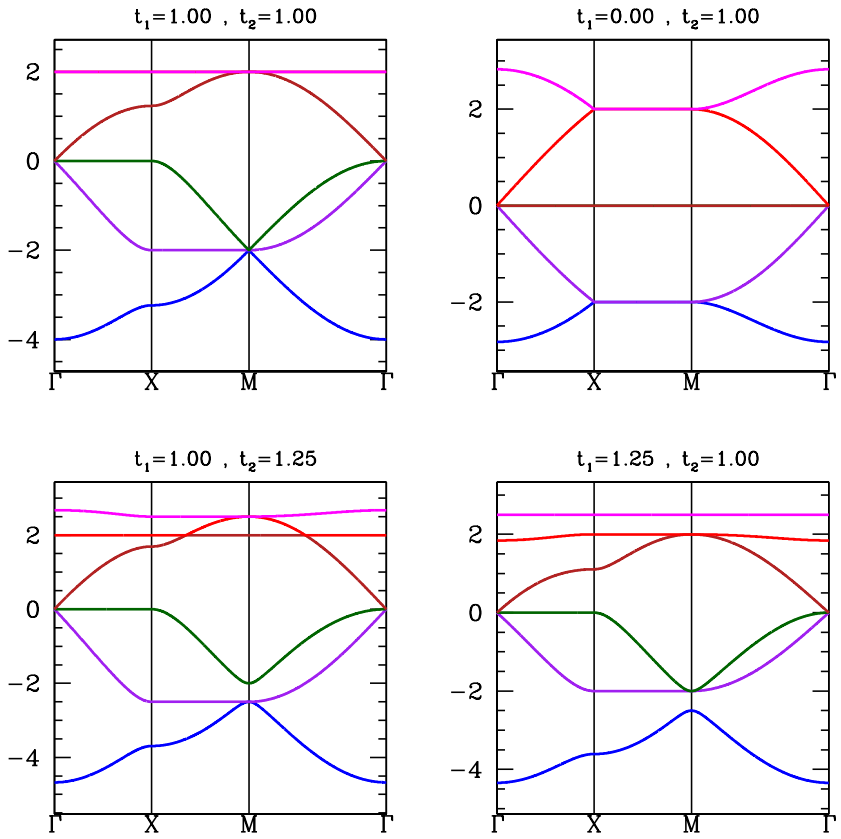


Figure 16.4: Energy bands for the line graph of the square-octagon lattice.

and their negatives, where $a = 2\sqrt{2}\,d$ is the FCC lattice constant. Here

$$\boldsymbol{a}_1 = \frac{a}{\sqrt{2}}\,(0,1,1) \qquad \boldsymbol{a}_2 = \frac{a}{\sqrt{2}}\,(1,0,1) \qquad \boldsymbol{a}_3 = \frac{a}{\sqrt{2}}\,(1,1,0) \tag{16.12}$$

are primitive FCC direct lattice basis vectors. Note $a = 2b$ relates the side length of the tetrahedra to the FCC lattice constant. The center of any α-tetrahedron is then located at

$$\boldsymbol{R} = m_1\,\boldsymbol{a}_1 + m_2\,\boldsymbol{a}_2 + m_3\,\boldsymbol{a}_3\ , \tag{16.13}$$

and the four basis vectors are $\boldsymbol{s}_{1,2,3,4} = \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}\}$, *i.e.*

$$\begin{aligned} \boldsymbol{s}_1 &= \frac{a}{4\sqrt{2}}\,(1,-1,1) & \boldsymbol{s}_3 &= \frac{a}{4\sqrt{2}}\,(1,1,-1) \\ \boldsymbol{s}_2 &= \frac{a}{4\sqrt{2}}\,(-1,1,1) & \boldsymbol{s}_4 &= \frac{a}{4\sqrt{2}}\,(-1,-1,-1)\ . \end{aligned} \tag{16.14}$$

The reciprocal lattice is BCC, with primitive vectors

$$\boldsymbol{b}_1 = \frac{\sqrt{2}\,\pi}{a}\,(-1,1,1) \quad , \quad \boldsymbol{b}_2 = \frac{\sqrt{2}\,\pi}{a}\,(1,-1,1) \quad , \quad \boldsymbol{b}_3 = \frac{\sqrt{2}\,\pi}{a}\,(1,1,-1)\ , \tag{16.15}$$

which satisfy $\boldsymbol{b}_i \cdot \boldsymbol{a}_j = 2\pi\,\delta_{ij}$.

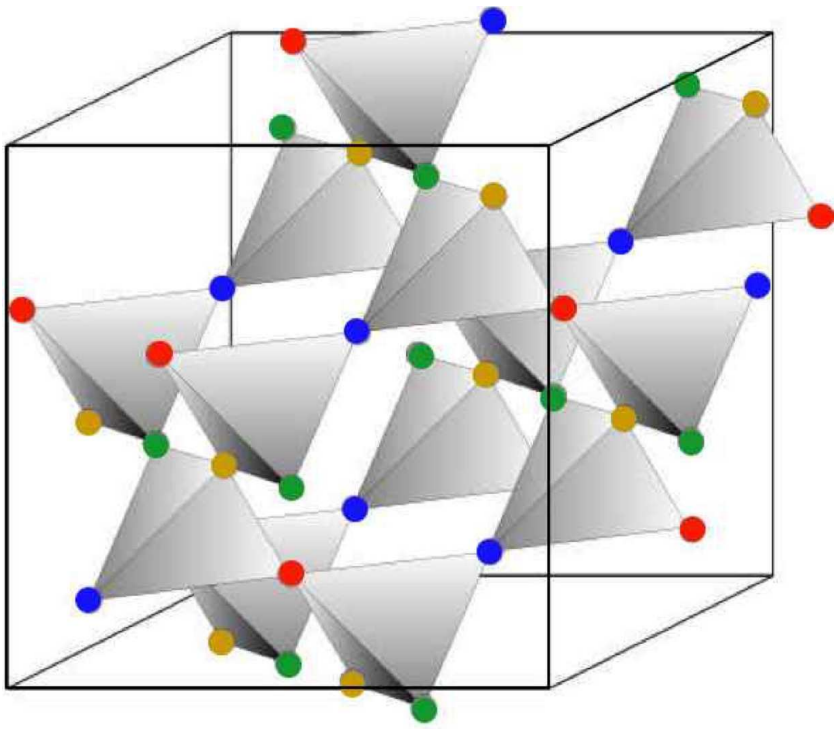


Figure 16.5: The pyrochlore lattice is an FCC Bravais lattice with a four element basis.

16.2.1 Adjacency Matrix

Let us label each vertex on the pyrochlore lattice by a Bravais lattice site $\boldsymbol{R}_{m_1,m_2,m_3}$ (*i.e.* the center of an α-tetrahedron) and by a sublattice index, with $\mathrm{A} \equiv 1$, $\mathrm{B} \equiv 2$, $\mathrm{C} \equiv 3$, $\mathrm{D} \equiv 4$. The adjacency matrix is

$$A_{ij}(\boldsymbol{R}) = \begin{pmatrix} 0 & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_2-\boldsymbol{a}_1} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_2-\boldsymbol{a}_3} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_2} \\ \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_1-\boldsymbol{a}_2} & 0 & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_1-\boldsymbol{a}_3} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_1} \\ \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_3-\boldsymbol{a}_2} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_3-\boldsymbol{a}_1} & 0 & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},\boldsymbol{a}_3} \\ \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},-\boldsymbol{a}_2} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},-\boldsymbol{a}_1} & \delta_{\boldsymbol{R},0} + \delta_{\boldsymbol{R},-\boldsymbol{a}_3} & 0 \end{pmatrix} . \tag{16.16}$$

The definition of the adjacency matrix is that $A_{ij}(\boldsymbol{R}) = 1$ if sites $\boldsymbol{s}_i$ and $\boldsymbol{R} + \boldsymbol{s}_j$ are nearest neighbors.

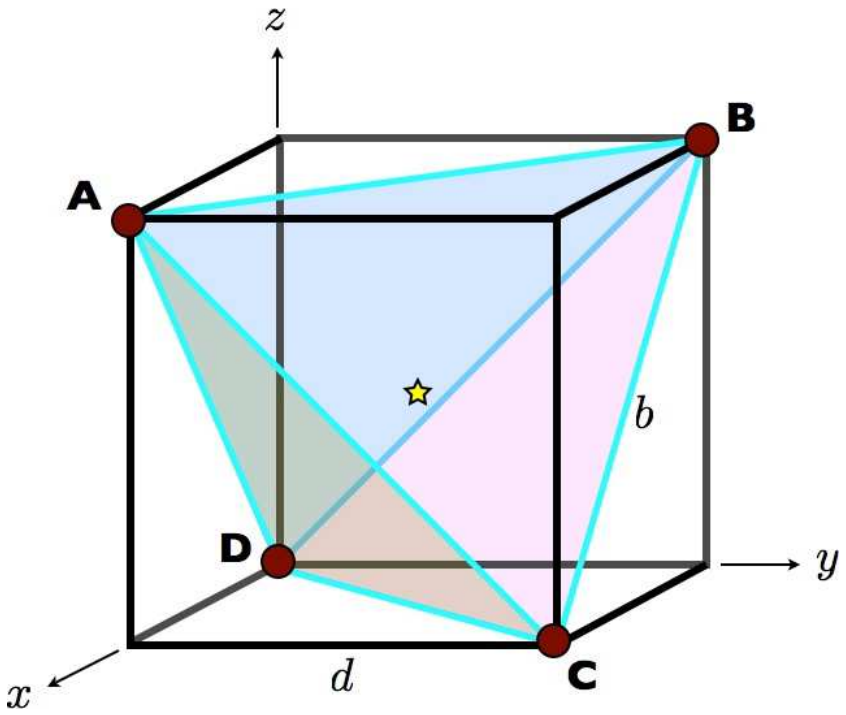


Figure 16.6: A cube of side length d, containing a tetrahedron of side length $b = \sqrt{2}\,d$. The origin of coordinates is located at the yellow star, which lies at the geometric center of the cube.

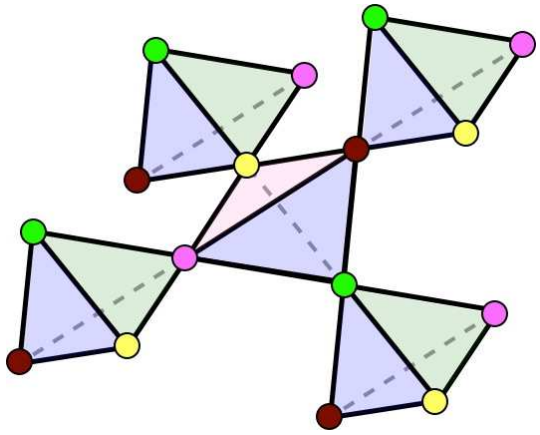


Figure 16.7: The α and β tetrahedra are related by inversion. The colors of the sites show the sublattice structure.

The Fourier transform of the adjacency matrix is

$$\hat{A}_{ij}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} A_{ij}(\boldsymbol{R})\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \tag{16.17}$$

$$= \begin{pmatrix} 0 & 1+e^{i(\theta_2-\theta_1)} & 1+e^{i(\theta_2-\theta_3)} & 1+e^{i\theta_2} \\ 1+e^{i(\theta_1-\theta_2)} & 0 & 1+e^{i(\theta_1-\theta_3)} & 1+e^{i\theta_1} \\ 1+e^{i(\theta_3-\theta_2)} & 1+e^{i(\theta_3-\theta_1)} & 0 & 1+e^{i\theta_3} \\ 1+e^{-i\theta_2} & 1+e^{-i\theta_1} & 1+e^{-i\theta_3} & 0 \end{pmatrix} , \tag{16.18}$$

where we define

$$\boldsymbol{k} \equiv \frac{\theta_1\,\boldsymbol{b}_1}{2\pi} + \frac{\theta_2\,\boldsymbol{b}_2}{2\pi} + \frac{\theta_3\,\boldsymbol{b}_3}{2\pi} , \tag{16.19}$$

so that

$$\boldsymbol{k}\cdot\boldsymbol{R}_{m_1,m_2,m_3} = m_1\,\theta_1 + m_2\,\theta_2 + m_3\,\theta_3 . \tag{16.20}$$

Note that the unitary transformation,

$$U = \begin{pmatrix} e^{i\theta_2/2} & 0 & 0 & 0 \\ 0 & e^{i\theta_1/2} & 0 & 0 \\ 0 & 0 & e^{i\theta_3/2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16.21}$$

has the effect

$$U^\dagger\,\hat{A}_{ij}(\boldsymbol{k})\,U = 2\begin{pmatrix} 0 & c_{12} & c_{23} & c_2 \\ c_{12} & 0 & c_{13} & c_1 \\ c_{23} & c_{13} & 0 & c_3 \\ c_2 & c_1 & c_3 & 0 \end{pmatrix} , \tag{16.22}$$

where $c_i = \cos\frac{1}{2}\theta_i$ and $c_{ij} = \cos\frac{1}{2}(\theta_i - \theta_j)$.

The characteristic polynomial is found to be

$$\begin{aligned} P(\lambda) &= \det\left(\lambda - \hat{A}\right) \\ &= \lambda^4 - 4\alpha\lambda^2 - 16\beta\lambda + 16\gamma \ , \end{aligned} \tag{16.23}$$

where

$$\begin{aligned} \alpha &= c_1^2 + c_2^2 + c_3^2 + c_{12}^2 + c_{13}^2 + c_{23}^2 \\ \beta &= c_1\,c_2\,c_{12} + c_1\,c_3\,c_{13} + c_2\,c_3\,c_{23} + c_{12}\,c_{13}\,c_{23} \\ \gamma &= c_1^2\,c_{23}^2 + c_2^2\,c_{13}^2 + c_3^2\,c_{12}^2 - 2\,c_1\,c_2\,c_{13}\,c_{23} - 2\,c_1\,c_3\,c_{12}\,c_{23} - 2\,c_2\,c_3\,c_{12}\,c_{13} \ . \end{aligned} \tag{16.24}$$

Simplifying, using results such as

$$c_1^2 + c_2^2 + c_{12}^2 = 1 + 2\,c_1\,c_2\,c_{12} \ , \tag{16.25}$$

we find

$$\beta = 1 + \tfrac{1}{2}\Big(\cos\theta_1 + \cos\theta_2 + \cos\theta_3 + \cos(\theta_1 - \theta_2) + \cos(\theta_1 - \theta_3) + \cos(\theta_2 - \theta_3)\Big) \tag{16.26}$$

and

$$\alpha = 2 + \beta \qquad , \qquad \gamma = 1 - \beta \ . \tag{16.27}$$

The function $\beta(\boldsymbol{\theta})$ takes its minimum value of $\beta_{\rm min} = 0$ at $\boldsymbol{\theta} = \frac{\pi}{2}\,(1,1,0)$ (and at symmetry-related points in the Brillouin zone). The maximum occurs at the zone center $\boldsymbol{\theta} = 0$, where $\beta_{\rm max} = 4$.

The characteristic polynomial may be factored, using the results of eqn. 16.27. If we define $\lambda \equiv \varepsilon - 2$, then

$$\begin{aligned} P(\varepsilon - 2) &= \varepsilon^4 - 8\varepsilon^3 + 4(6 - \alpha)\,\varepsilon^2 + 16(2 + \beta - \alpha)\,\varepsilon + 16(1 - \alpha + 2\beta + \gamma) \\ &= \varepsilon^2\left(\varepsilon^2 - 8\,\varepsilon + 4(4 - \beta)\right) \ , \end{aligned} \tag{16.28}$$

and thus with $\lambda = \varepsilon - 2$ we have the four bands

$$\lambda_1 = -2 \quad , \quad \lambda_2 = -2 \quad , \quad \lambda_3 = 2 - 2\sqrt{\beta} \quad , \quad \lambda_4 = 2 + 2\sqrt{\beta} \ . \tag{16.29}$$

Note that there are two flat bands at $\lambda_{1,2} = -2$. Note also that the largest eigenvalue is $\lambda_{4,\rm max} = 2 + 2\beta_{\rm max}^{1/2} = 6$, which is (correctly) the lattice coordination number. To make contact with some results of ref. 1, the adjacency matrix in eqn. 4 of ref. 1 is equivalent to $\hat{A} + 2$, hence the eigenvalues are our $\varepsilon_{1,2,3,4}$, which is to say $\varepsilon_{1,2} = 0$ and $\varepsilon_{3,4} = 4 \pm 2\sqrt{\beta}$. I presume that there is a typo and the eigenvalues $\nu_{1,2,3,4}$ in ref. 1 are for half the adjacency matrix.

For the electronic hopping Hamiltonian $\mathcal{H} = -t\sum_{\langle ij\rangle}\left(c_i^\dagger c_j + c_j^\dagger c_i\right)$ on the pyrochlore lattice, the energy eigenvalues are

$$E_1 = -2t - 2t\sqrt{\beta(\boldsymbol{\theta})} \quad , \quad E_2 = -2t + 2t\sqrt{\beta(\boldsymbol{\theta})} \quad , \quad E_{3,4} = +2t \ , \tag{16.30}$$

where $\beta(\boldsymbol{\theta}) \in [0, 4]$, as discussed above.

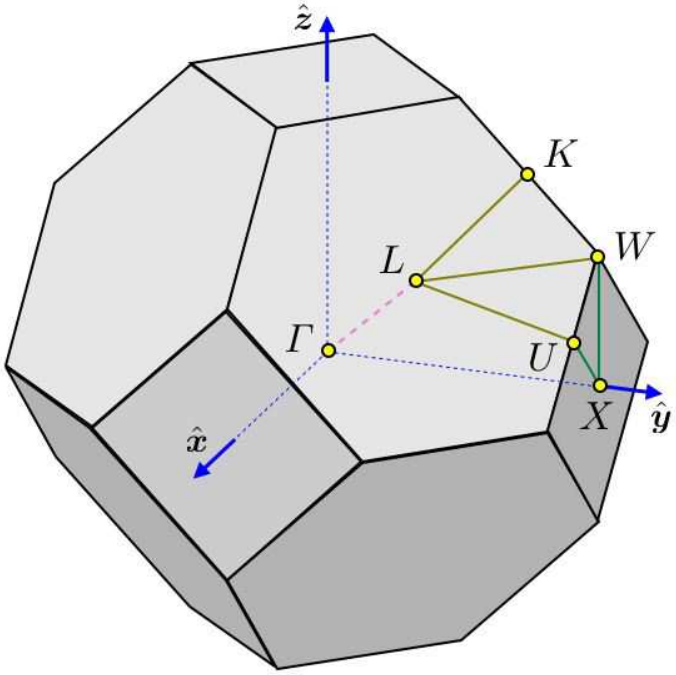


Figure 16.8: First Brillouin zone for an FCC structure, with high symmetry points labeled.

16.2.2 The FCC lattice Brillouin zone

Fig. 16.8 shows the first Brillouin zone for a (real space) FCC structure. Note that

$$\Gamma = (0,0,0) \tag{16.31}$$

$$L = \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3\big) = \frac{\sqrt{2}\,\pi}{a}\,\big(1,1,1\big) \tag{16.32}$$

$$X = \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_3\big) = \frac{\sqrt{2}\,\pi}{a}\,\big(0,1,0\big)\ . \tag{16.33}$$

Consider the point L' residing in the center of the hexagonal face adjacent to that containing L but with negative x coordinate. (This face is hidden in the figure.) One has

$$L' = \tfrac{1}{2}\boldsymbol{b}_1 = \frac{\sqrt{2}\,\pi}{a}\,\big(-1,1,1\big)\ . \tag{16.34}$$

To find W we note that it lies at the confluence of the two hexagonal faces containing L and L', respectively, and the square face containing X. We may therefore write

$$\begin{aligned} W &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3\big) + r\cdot\tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_2\big) + s\cdot\tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_3\big) \\ W &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_3\big) + u\cdot\tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2\big) + v\cdot\tfrac{1}{2}\big(\boldsymbol{b}_2 + \boldsymbol{b}_3\big) \\ W &= \tfrac{1}{2}\boldsymbol{b}_1 + x\cdot\tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + 2\,\boldsymbol{b}_3\big) + y\cdot\tfrac{1}{2}\big(\boldsymbol{b}_2 - \boldsymbol{b}_3\big)\ , \end{aligned} \tag{16.35}$$

where we must solve for (r, s, u, v, x, y). What we have done here is to write the location of W as the location of each of the face centers plus an unknown vector lying in the plane of that face. Since the elementary reciprocal lattice vectors $\boldsymbol{b}_{1,2,3}$ are a linearly independent set, the coefficients of each elementary reciprocal lattice vector must independently sum to zero in the two equations which result from equating the first and second, and the second and third lines in eqn. 16.35. This yields six equations in our six unknowns, and solving for the unknowns we obtain

$$r = \tfrac{1}{2} \quad , \quad s = 0 \quad , \quad u = \tfrac{1}{2} \quad , \quad v = 0 \quad , \quad x = \tfrac{1}{2} \quad , \quad y = 0 \, , \tag{16.36}$$

and therefore

$$W = \tfrac{3}{4}\,\boldsymbol{b}_1 + \tfrac{1}{4}\,\boldsymbol{b}_2 + \tfrac{1}{2}\,\boldsymbol{b}_3 = \frac{\sqrt{2}\,\pi}{a}\left(0, 1, \tfrac{1}{2}\right) . \tag{16.37}$$

Consider next the point W' lying on the other end of the edge containing U and K. We have

$$\begin{aligned} W' &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3\big) + r \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_2\big) + s \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_3\big) \\ W' &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2\big) + u \cdot \tfrac{1}{2}\big(\boldsymbol{b}_2 + \boldsymbol{b}_3\big) + v \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_3\big) \\ W' &= \tfrac{1}{2}\boldsymbol{b}_1 + x \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + 2\,\boldsymbol{b}_3\big) + y \cdot \tfrac{1}{2}\big(\boldsymbol{b}_2 - \boldsymbol{b}_3\big) \, . \end{aligned} \tag{16.38}$$

Solving for the unknowns, we obtain

$$r = \tfrac{1}{2} \quad , \quad s = \tfrac{1}{2} \quad , \quad u = 0 \quad , \quad v = \tfrac{1}{2} \quad , \quad x = \tfrac{1}{2} \quad , \quad y = \tfrac{1}{2} \, , \tag{16.39}$$

and therefore

$$W' = \tfrac{3}{4}\,\boldsymbol{b}_1 + \tfrac{1}{2}\,\boldsymbol{b}_2 + \tfrac{1}{4}\,\boldsymbol{b}_3 = \frac{\sqrt{2}\,\pi}{a}\left(0, \tfrac{1}{2}, 1\right) . \tag{16.40}$$

Finally, consider the point W'' lying on the other end of the edge containing W and U. One then has

$$\begin{aligned} W &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2 + \boldsymbol{b}_3\big) + r \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_2\big) + s \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_3\big) \\ W &= \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_3\big) + u \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 + \boldsymbol{b}_2\big) + v \cdot \tfrac{1}{2}\big(\boldsymbol{b}_2 + \boldsymbol{b}_3\big) \\ W &= \tfrac{1}{2}\boldsymbol{b}_3 + x \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 - \boldsymbol{b}_2\big) + y \cdot \tfrac{1}{2}\big(\boldsymbol{b}_1 + 2\,\boldsymbol{b}_2 + \boldsymbol{b}_3\big) \, . \end{aligned} \tag{16.41}$$

Solving for the unknowns, we obtain

$$r = 0 \quad , \quad s = -\tfrac{1}{2} \quad , \quad u = 0 \quad , \quad v = \tfrac{1}{2} \quad , \quad x = \tfrac{1}{2} \quad , \quad y = \tfrac{1}{2} \, , \tag{16.42}$$

and therefore

$$W'' = \tfrac{1}{2}\,\boldsymbol{b}_1 + \tfrac{1}{4}\,\boldsymbol{b}_2 + \tfrac{3}{4}\,\boldsymbol{b}_3 = \frac{\sqrt{2}\,\pi}{a}\left(\tfrac{1}{2}, 1, 0\right) . \tag{16.43}$$

Since K and U lie a the midpoints of WW' and WW'' respectively, we may now write

$$K = \tfrac{3}{4}\,\boldsymbol{b}_1 + \tfrac{3}{8}\,\boldsymbol{b}_3 + \tfrac{3}{8}\,\boldsymbol{b}_3 = \left(0, \tfrac{3}{4}, \tfrac{3}{4}\right) \tag{16.44}$$

$$U = \tfrac{5}{8}\,\boldsymbol{b}_1 + \tfrac{1}{4}\,\boldsymbol{b}_2 + \tfrac{5}{8}\,\boldsymbol{b}_3 = \left(\tfrac{1}{4}, 1, \tfrac{1}{4}\right) . \tag{16.45}$$

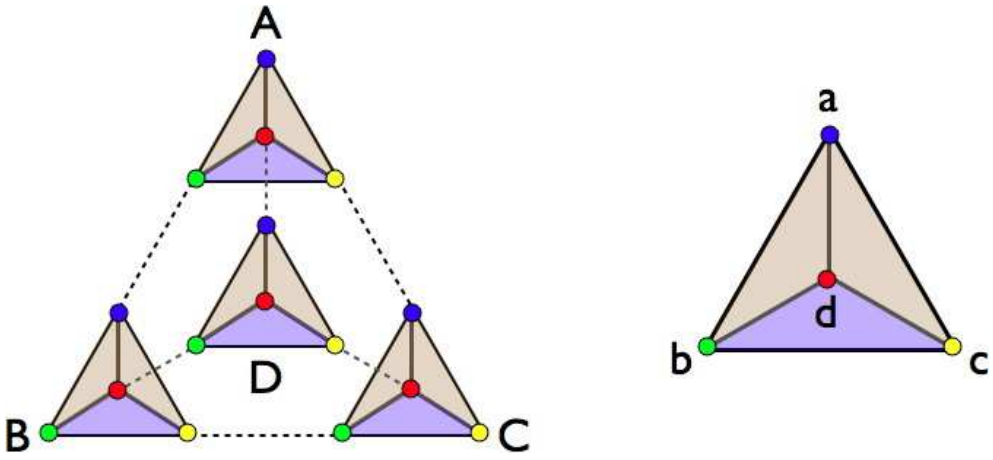


Figure 16.9: Supertetrahedron basis for the pyrochlore lattice.

16.3 Depleted Pyrochlores

16.3.1 Pyrochlore with 16 element basis

We change notation slightly. We define

$$\boldsymbol{a}_1 = \tfrac{1}{\sqrt{2}}\,a\big(0,1,1\big) \quad , \quad \boldsymbol{a}_2 = \tfrac{1}{\sqrt{2}}\,a\big(1,0,1\big) \quad , \quad \boldsymbol{a}_3 = \tfrac{1}{\sqrt{2}}\,a\big(1,1,0\big)\ . \tag{16.46}$$

Consider the 16 site basis for the pyrochlore lattice defined in Fig. 16.9, composed of a 'supertetrahedron' formed from four primary tetrahedra. The sites on each tetrahedron are labeled $\{\mathrm{a},\mathrm{b},\mathrm{c},\mathrm{d}\}$ and the tetrahedra are labeled $\{\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D}\}$. The origin is taken to be site Dd. The supertetrahedra form a simple cubic lattice with primary direct lattice vectors

$$\begin{aligned}
\boldsymbol{c}_1 &= -\boldsymbol{a}_1 + \boldsymbol{a}_2 + \boldsymbol{a}_3 = \sqrt{2}\,a\,(1,0,0) \\
\boldsymbol{c}_2 &= +\boldsymbol{a}_1 - \boldsymbol{a}_2 + \boldsymbol{a}_3 = \sqrt{2}\,a\,(0,1,0) \\
\boldsymbol{c}_3 &= +\boldsymbol{a}_1 + \boldsymbol{a}_2 - \boldsymbol{a}_3 = \sqrt{2}\,a\,(0,0,1)\ .
\end{aligned} \tag{16.47}$$

To see why the supertetrahedra fill all space with no missing tetrahedra, write the location of Dd on supertetrahedron (n_1,n_2,n_3) as

$$\begin{aligned}
\boldsymbol{R}_{n_1,n_2,n_3} &= n_1\,\boldsymbol{c}_1 + n_2\,\boldsymbol{c}_2 + n_3\,\boldsymbol{c}_3 \\
&= (-n_1+n_2+n_3)\,\boldsymbol{a}_1 + (n_1-n_2+n_3)\,\boldsymbol{a}_2 + (n_1+n_2-n_3)\,\boldsymbol{a}_3 \\
&\equiv \ell_1\,\boldsymbol{a}_1 + \ell_2\,\boldsymbol{a}_2 + \ell_3\,\boldsymbol{a}_3\ .
\end{aligned} \tag{16.48}$$

Thus,

$$\ell_1 = -n_1+n_2+n_3 \quad , \quad \ell_2 = n_1-n_2+n_3 \quad , \quad \ell_3 = n_1+n_2-n_3\ . \tag{16.49}$$

Note that $\ell_1 - \ell_2$, $\ell_2 - \ell_3$, and $\ell_3 - \ell_1$ are all even integers. $\boldsymbol{R}_{n_1,n_2,n_3}$ is the location of the Dd site in unit cell (n_1, n_2, n_3). The location of the Ad site is then $\boldsymbol{R}_{n_1,n_2,n_3} + \boldsymbol{a}_1$. The location of the Bd site is $\boldsymbol{R}_{n_1,n_2,n_3} + \boldsymbol{a}_2$. The location of the Cd site is $\boldsymbol{R}_{n_1,n_2,n_3} + \boldsymbol{a}_3$. Thus, we have

$$\begin{aligned}(\ell_1 - \ell_2\,,\,\ell_2 - \ell_3\,,\,\ell_3 - \ell_1) &= (\text{odd}\,,\,\text{odd}\,,\,\text{even}) \quad \text{(A)} \\ &= (\text{odd}\,,\,\text{even}\,,\,\text{odd}) \quad \text{(B)} \\ &= (\text{even}\,,\,\text{odd}\,,\,\text{odd}) \quad \text{(C)} \\ &= (\text{even}\,,\,\text{even}\,,\,\text{even}) \quad \text{(D)}\,.\end{aligned} \tag{16.50}$$

Since $(\ell_1 - \ell_2) + (\ell_2 - \ell_3) + (\ell_3 - \ell_1) = 0$, there number of odd differences must itself be even (*i.e.* 0 or 2). This exhausts all the possibilities for (ℓ_1, ℓ_2, ℓ_3), so we have identified every tetrahedron.

The locations of the 16 sites in the $(0, 0, 0)$ supertetrahedron are then

$$\begin{aligned}
&\text{Aa} = \tfrac{3}{2}\boldsymbol{a}_1 && \text{Ab} = \boldsymbol{a}_1 + \tfrac{1}{2}\boldsymbol{a}_2 && \text{Ac} = \boldsymbol{a}_1 + \tfrac{1}{2}\boldsymbol{a}_3 && \text{Ad} = \boldsymbol{a}_1 && (16.51)\\
&\text{Ba} = \boldsymbol{a}_2 + \tfrac{1}{2}\boldsymbol{a}_1 && \text{Bb} = \tfrac{3}{2}\boldsymbol{a}_2 && \text{Bc} = \boldsymbol{a}_2 + \tfrac{1}{2}\boldsymbol{a}_3 && \text{Bd} = \boldsymbol{a}_2 && (16.52)\\
&\text{Ca} = \boldsymbol{a}_3 + \tfrac{1}{2}\boldsymbol{a}_1 && \text{Cb} = \boldsymbol{a}_3 + \tfrac{1}{2}\boldsymbol{a}_2 && \text{Cc} = \tfrac{3}{2}\boldsymbol{a}_3 && \text{Cd} = \boldsymbol{a}_3 && (16.53)\\
&\text{Da} = \tfrac{1}{2}\boldsymbol{a}_1 && \text{Db} = \tfrac{1}{2}\boldsymbol{a}_2 && \text{Dc} = \tfrac{1}{2}\boldsymbol{a}_3 && \text{Dd} = 0\,. && (16.54)
\end{aligned}$$

The Fourier transform of the adjacency matrix is given by

$M(\boldsymbol{\theta}) =$

	Aa	Ab	Ac	Ad	Ba	Bb	Bc	Bd	Ca	Cb	Cc	Cd	Da	Db	Dc	Dd
Aa	0	1	1	1	0	$\bar{z}_1 z_2$	0	0	0	0	$\bar{z}_1 z_3$	0	0	0	0	$z_2 z_3$
Ab	1	0	1	1	1	0	0	0	0	0	0	z_3	0	0	z_3	0
Ac	1	1	0	1	0	0	0	z_2	1	0	0	0	0	z_2	0	0
Ad	1	1	1	0	0	0	$\bar{z}_1$	0	0	$\bar{z}_1$	0	0	1	0	0	0
Ba	0	1	0	0	0	1	1	1	0	0	0	z_3	0	0	z_3	0
Bb	$z_1 \bar{z}_2$	0	0	0	1	0	1	1	0	0	$\bar{z}_2 z_3$	0	0	0	0	$z_1 z_3$
Bc	0	0	0	z_1	1	1	0	1	0	1	0	0	z_1	0	0	0
Bd	0	0	$\bar{z}_2$	0	1	1	1	0	$\bar{z}_2$	0	0	0	0	1	0	0
Ca	0	0	1	0	0	0	0	z_2	0	1	1	1	0	z_2	0	0
Cb	0	0	0	z_1	0	0	1	0	1	0	1	1	z_1	0	0	0
Cc	$z_1 \bar{z}_3$	0	0	0	0	$z_2 \bar{z}_3$	0	0	1	1	0	1	0	0	0	$z_1 z_2$
Cd	0	$\bar{z}_3$	0	0	$\bar{z}_3$	0	0	0	1	1	1	0	0	0	1	0
Da	0	0	0	1	0	0	$\bar{z}_1$	0	0	$\bar{z}_1$	0	0	0	1	1	1
Db	0	0	$\bar{z}_2$	0	0	0	0	1	$\bar{z}_2$	0	0	0	1	0	1	1
Dc	0	$\bar{z}_3$	0	0	$\bar{z}_3$	0	0	0	0	0	0	1	1	1	0	1
Dd	$\bar{z}_2 \bar{z}_3$	0	0	0	0	$\bar{z}_1 \bar{z}_3$	0	0	0	0	$\bar{z}_1 \bar{z}_2$	0	1	1	1	0

where $z_j = e^{i\theta_j}$ with $j \in \{1, 2, 3\}$. The way to read this matrix is as follows. There are six nonzero entries in each row and in each column, corresponding to coordination number six.

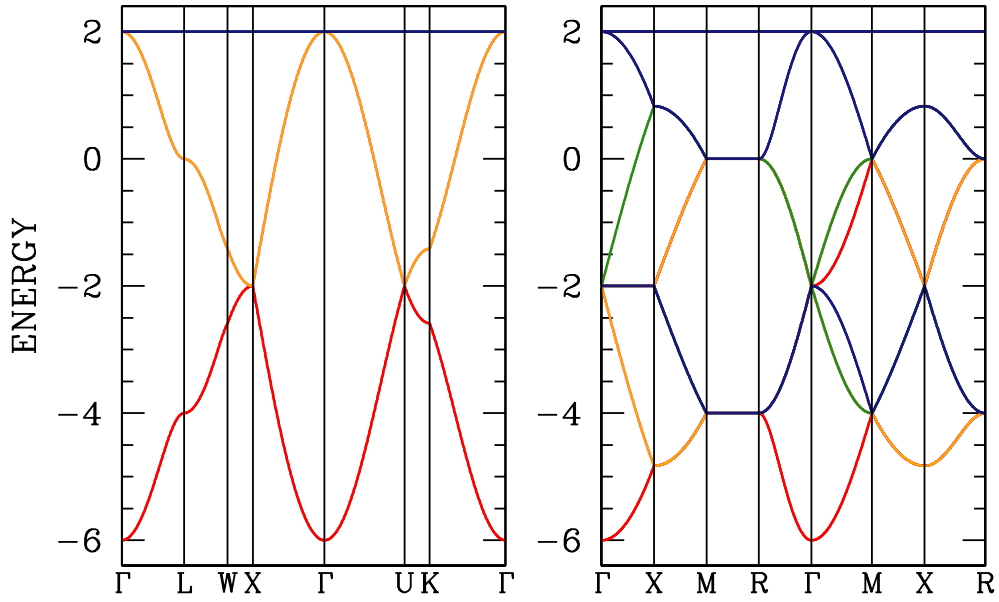


Figure 16.10: Comparison of pyrochlore band structures computed in 4-element FCC basis (left) and 16-element SC basis (right).

The rows correspond to sites Aa (1) through Dd (16). Reading along the first row, we see that Aa has neighbors Ab, Ac, and Ad in the unit cell $(0,0,0)$, as well as a neighbor Bb in unit cell $\boldsymbol{c}_2 - \boldsymbol{c}_1$ (from $M_{16} = \bar{z}_1 z_2$), a neighbor Cc in unit cell $\boldsymbol{c}_3 - \boldsymbol{c}_1$, and a neighbor Dd in unit cell $\boldsymbol{c}_2 + \boldsymbol{c}_3$. The Hamiltonian is then $H(\boldsymbol{\theta}) = -t\, M(\boldsymbol{\theta})$.

The high-symmetry points in the cubic lattice Brillouin zone are as follows:

$$\begin{aligned} \Gamma &: (\theta_1, \theta_2, \theta_3) = (0,0,0) \qquad \text{(zone center)} \\ \mathrm{X} &: (\theta_1, \theta_2, \theta_3) = (\pi,0,0) \qquad \text{(face center)} \\ \mathrm{M} &: (\theta_1, \theta_2, \theta_3) = (\pi,\pi,0) \qquad \text{(edge center)} \\ \mathrm{R} &: (\theta_1, \theta_2, \theta_3) = (\pi,\pi,\pi) \qquad \text{(zone corner)}\ . \end{aligned} \tag{16.55}$$

In Fig. 16.10, we plot the dispersion for the tight binding Hamiltonian on the pyrochlore lattice with $t = 1$ using both the 4-element FCC basis as well as the 16-element SC basis.

For the Kondo Hamiltonian with fixed local moments, we write

$$H(\boldsymbol{\theta}) = -t \sum_{a,b=1}^{16} \sum_{\mu=1}^{2} M_{ab}(\boldsymbol{\theta})\, c^\dagger_{a,\mu}(\boldsymbol{\theta})\, c_{b,\mu}(\boldsymbol{\theta}) - J \sum_{a=1}^{16} \sum_{\mu,\nu=1}^{2} \hat{n}_a \cdot c^\dagger_{a,\mu}(\boldsymbol{\theta})\, \boldsymbol{\sigma}_{\mu\nu}\, c_{a,\nu}(\boldsymbol{\theta})\ , \tag{16.56}$$

where J is the Kondo coupling and $\hat{n}_a$ is the direction of the local moment, with all local moments assumed to be of unit magnitude. At each $\boldsymbol{\theta}$ point in the Brillouin zone, the Hamiltonian is a 32×32 matrix.

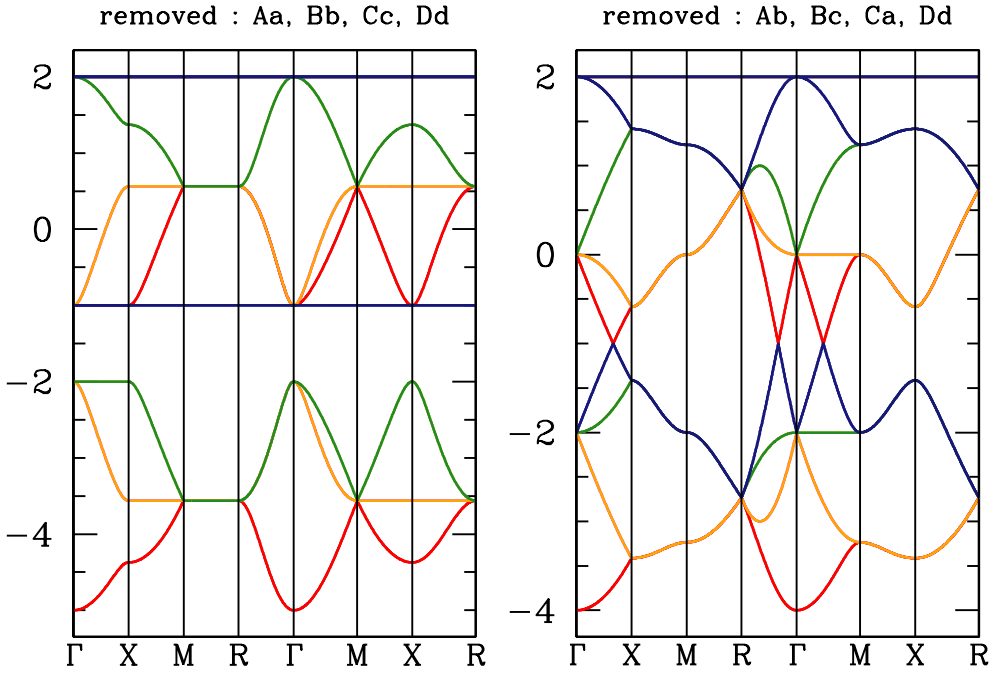


Figure 16.11: Comparison of depleted pyrochlore band structures ($t = 1$, $J = 0$). Left panel: sites Aa, Bb, Cc, Dd removed. Right panel: sites Ab, Bc, Ca, Dd removed.

16.3.2 Depleted pyrochlore

To deplete the pyrochlore, we knock out rows and columns of M. Equivalently, we can add an infinite on-site energy to the diagonal elements $H_{aa}(\boldsymbol{\theta})$. For the Kondo Hamiltonian, we add the infinite on-site energy to $H_{a\uparrow,a\uparrow}(\boldsymbol{\theta})$ and $H_{a\downarrow,a\downarrow}(\boldsymbol{\theta})$. In Fig. 16.11 we plot the band structures for two depleted pyrochlore structures, one in which sites Aa, Bb, Cc, and Dd have been removed (I), and another in which sites Ab, Bc, Ca, and Dd have been removed (II). The resulting structures may not have the full cubic lattice symmetry.

Eliminating the necessary rows and columns from $M(\boldsymbol{\theta})$, we arrive at the adjacency matrices

for the two depleted structures (I) and (II). We find for (I),

$$M^{(\mathrm{I})}(\boldsymbol{\theta}) = \begin{pmatrix}
 & \mathrm{Ab} & \mathrm{Ac} & \mathrm{Ad} & \mathrm{Ba} & \mathrm{Bc} & \mathrm{Bd} & \mathrm{Ca} & \mathrm{Cb} & \mathrm{Cd} & \mathrm{Da} & \mathrm{Db} & \mathrm{Dc} \\
\mathrm{Ab} & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & z_3 & 0 & 0 & z_3 \\
\mathrm{Ac} & 1 & 0 & 1 & 0 & 0 & z_2 & 1 & 0 & 0 & 0 & z_2 & 0 \\
\mathrm{Ad} & 1 & 1 & 0 & 0 & \bar{z}_1 & 0 & 0 & \bar{z}_1 & 0 & 1 & 0 & 0 \\
\mathrm{Ba} & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & z_3 & 0 & 0 & z_3 \\
\mathrm{Bc} & 0 & 0 & z_1 & 1 & 0 & 1 & 0 & 1 & 0 & z_1 & 0 & 0 \\
\mathrm{Bd} & 0 & \bar{z}_2 & 0 & 1 & 1 & 0 & \bar{z}_2 & 0 & 0 & 0 & 1 & 0 \\
\mathrm{Ca} & 0 & 1 & 0 & 0 & 0 & z_2 & 0 & 1 & 1 & 0 & z_2 & 0 \\
\mathrm{Cb} & 0 & 0 & z_1 & 0 & 1 & 0 & 1 & 0 & 1 & z_1 & 0 & 0 \\
\mathrm{Cd} & \bar{z}_3 & 0 & 0 & \bar{z}_3 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
\mathrm{Da} & 0 & 0 & 1 & 0 & \bar{z}_1 & 0 & 0 & \bar{z}_1 & 0 & 0 & 1 & 1 \\
\mathrm{Db} & 0 & \bar{z}_2 & 0 & 0 & 0 & 1 & \bar{z}_2 & 0 & 0 & 1 & 0 & 1 \\
\mathrm{Dc} & \bar{z}_3 & 0 & 0 & \bar{z}_3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{pmatrix}.$$

This corresponds to a five-fold coordinated structure, since there are five nonzero elements in each row and each column. For the (II) structure, we have

$$M^{(\mathrm{II})}(\boldsymbol{\theta}) = \begin{pmatrix}
 & \mathrm{Aa} & \mathrm{Ac} & \mathrm{Ad} & \mathrm{Ba} & \mathrm{Bb} & \mathrm{Bd} & \mathrm{Cb} & \mathrm{Cc} & \mathrm{Cd} & \mathrm{Da} & \mathrm{Db} & \mathrm{Dc} \\
\mathrm{Aa} & 0 & 1 & 1 & 0 & \bar{z}_1 z_2 & 0 & 0 & \bar{z}_1 z_3 & 0 & 0 & 0 & 0 \\
\mathrm{Ac} & 1 & 0 & 1 & 0 & 0 & z_2 & 0 & 0 & 0 & 0 & z_2 & 0 \\
\mathrm{Ad} & 1 & 1 & 0 & 0 & 0 & 0 & \bar{z}_1 & 0 & 0 & 1 & 0 & 0 \\
\mathrm{Ba} & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & z_3 & 0 & 0 & z_3 \\
\mathrm{Bb} & z_1 \bar{z}_2 & 0 & 0 & 1 & 0 & 1 & 0 & \bar{z}_2 z_3 & 0 & 0 & 0 & 0 \\
\mathrm{Bd} & 0 & \bar{z}_2 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\mathrm{Cb} & 0 & 0 & z_1 & 0 & 0 & 0 & 0 & 1 & 1 & z_1 & 0 & 0 \\
\mathrm{Cc} & z_1 \bar{z}_3 & 0 & 0 & 0 & z_2 \bar{z}_3 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
\mathrm{Cd} & 0 & 0 & 0 & \bar{z}_3 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
\mathrm{Da} & 0 & 0 & 1 & 0 & 0 & 0 & \bar{z}_1 & 0 & 0 & 0 & 1 & 1 \\
\mathrm{Db} & 0 & \bar{z}_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
\mathrm{Dc} & 0 & 0 & 0 & \bar{z}_3 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{pmatrix},$$

corresponding to a four-fold coordinated structure, known as the hyperkagome lattice.

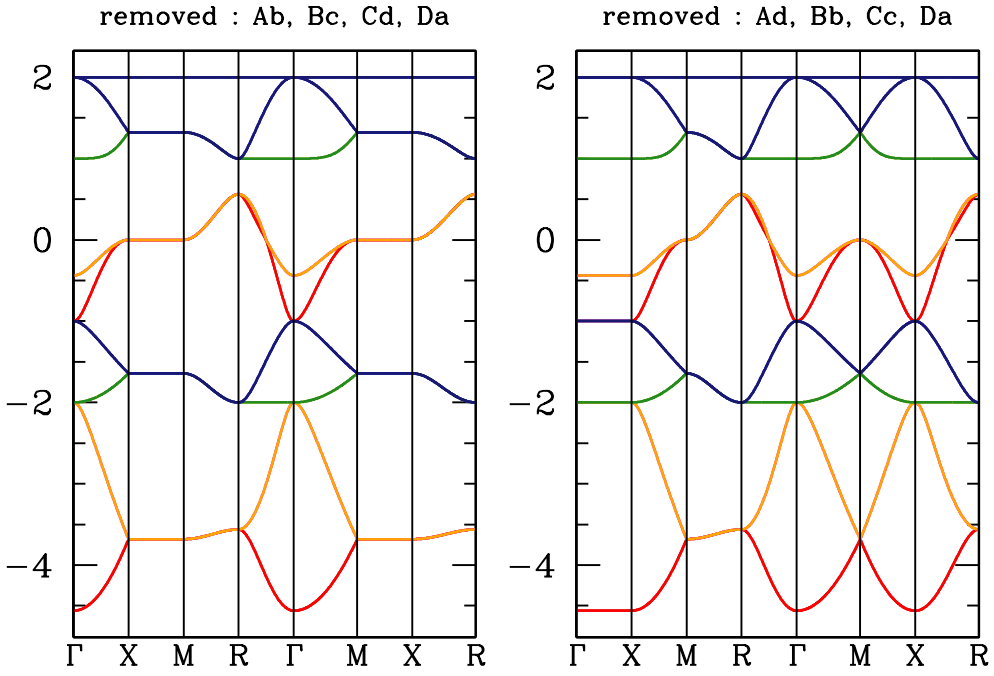


Figure 16.12: Comparison of two additional depleted pyrochlore band structures III and IV ($t = 1$, $J = 0$).

The energy bands for two additional structures are shown in Fig. 16.12. We find

$$M^{(\mathrm{III})}(\boldsymbol{\theta}) = \begin{pmatrix} & \mathrm{Aa} & \mathrm{Ac} & \mathrm{Ad} & \mathrm{Ba} & \mathrm{Bb} & \mathrm{Bd} & \mathrm{Ca} & \mathrm{Cb} & \mathrm{Cc} & \mathrm{Db} & \mathrm{Dc} & \mathrm{Dd} \\ \mathrm{Aa} & 0 & 1 & 1 & 0 & \bar{z}_1 z_2 & 0 & 0 & 0 & \bar{z}_1 z_3 & 0 & 0 & z_2 z_3 \\ \mathrm{Ac} & 1 & 0 & 1 & 0 & 0 & z_2 & 1 & 0 & 0 & z_2 & 0 & 0 \\ \mathrm{Ad} & 1 & 1 & 0 & 0 & 0 & 0 & 0 & \bar{z}_1 & 0 & 0 & 0 & 0 \\ \mathrm{Ba} & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & z_3 & 0 \\ \mathrm{Bb} & z_1 \bar{z}_2 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & \bar{z}_2 z_3 & 0 & 0 & z_1 z_3 \\ \mathrm{Bd} & 0 & \bar{z}_2 & 0 & 1 & 1 & 0 & \bar{z}_2 & 0 & 0 & 1 & 0 & 0 \\ \mathrm{Ca} & 0 & 1 & 0 & 0 & 0 & z_2 & 0 & 1 & 1 & z_2 & 0 & 0 \\ \mathrm{Cb} & 0 & 0 & z_1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ \mathrm{Cc} & z_1 \bar{z}_3 & 0 & 0 & 0 & z_2 \bar{z}_3 & 0 & 1 & 1 & 0 & 0 & 0 & z_1 z_2 \\ \mathrm{Db} & 0 & \bar{z}_2 & 0 & 0 & 0 & 1 & \bar{z}_2 & 0 & 0 & 0 & 1 & 1 \\ \mathrm{Dc} & 0 & 0 & 0 & \bar{z}_3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ \mathrm{Dd} & \bar{z}_2 \bar{z}_3 & 0 & 0 & 0 & \bar{z}_1 \bar{z}_3 & 0 & 0 & 0 & \bar{z}_1 \bar{z}_2 & 1 & 1 & 0 \end{pmatrix}$$

In this structure, nine of the 12 sites are 5-fold coordinated, and the remaining three are 3-fold

coordinated. What seems to be an equivalent structure is (IV):

$$M^{(\mathrm{IV})}(\boldsymbol{\theta}) = \begin{pmatrix}
 & \mathrm{Aa} & \mathrm{Ab} & \mathrm{Ac} & \mathrm{Ba} & \mathrm{Bc} & \mathrm{Bd} & \mathrm{Ca} & \mathrm{Cb} & \mathrm{Cd} & \mathrm{Db} & \mathrm{Dc} & \mathrm{Dd} \\
\mathrm{Aa} & 0 & 1 & 1 & 0 & \bar{z}_1 z_2 & 0 & 0 & 0 & 0 & 0 & 0 & z_2 z_3 \\
\mathrm{Ab} & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & z_3 & 0 & z_3 & 0 \\
\mathrm{Ac} & 1 & 0 & 1 & 0 & 0 & z_2 & 1 & 0 & 0 & z_2 & 0 & 0 \\
\mathrm{Ba} & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & z_3 & 0 & z_3 & 0 \\
\mathrm{Bc} & 0 & 0 & z_1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\mathrm{Bd} & 0 & \bar{z}_2 & 0 & 1 & 1 & 0 & \bar{z}_2 & 0 & 0 & 1 & 0 & 0 \\
\mathrm{Ca} & 0 & 1 & 0 & 0 & 0 & z_2 & 0 & 1 & 1 & z_2 & 0 & 0 \\
\mathrm{Cb} & 0 & 0 & z_1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
\mathrm{Cd} & 0 & 0 & 0 & \bar{z}_3 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
\mathrm{Db} & 0 & \bar{z}_2 & 0 & 0 & 0 & 1 & \bar{z}_2 & 0 & 0 & 0 & 1 & 1 \\
\mathrm{Dc} & 0 & 0 & 0 & \bar{z}_3 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
\mathrm{Dd} & \bar{z}_2 \bar{z}_3 & 0 & 0 & 0 & \bar{z}_1 \bar{z}_3 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix}$$

16.4 References

[1] S. Isakov, K. Gregor, R. Moessner, and S. L. Sondhi, *Phys. Rev. Lett.* **93**, 106702 (2004).

Chapter 17

Quadratic Hamiltonians

17.1 Bosonic Models

The general noninteracting bosonic Hamiltonian is written

$$\hat{H} = \tfrac{1}{2}\Psi_r^\dagger \mathcal{H}_{rs}\Psi_s \quad , \tag{17.1}$$

where Ψ is a rank-$2N$ column vector whose Hermitian conjugate is the row vector

$$\Psi^\dagger = \left(\psi_1^\dagger\,,\,\cdots\,,\,\psi_N^\dagger\,,\,\psi_1\,,\,\cdots\,,\,\psi_N\right) \quad . \tag{17.2}$$

Since $\left[\psi_i\,,\,\psi_j^\dagger\right] = \delta_{ij}$, we have

$$\left[\Psi_r\,,\,\Psi_s^\dagger\right] = \Sigma_{rs} \quad , \quad \Sigma = \begin{pmatrix} \mathbb{I}_{N\times N} & 0 \\ 0 & -\mathbb{I}_{N\times N} \end{pmatrix} \quad , \tag{17.3}$$

with $\mathbb{I}$ the identity matrix. Note that the indices r and s run from 1 to $2N$, while i and j run from 1 to N. The matrix $\mathcal{H}$ is of the form

$$\mathcal{H} = \begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \tag{17.4}$$

where $A = A^\dagger$ is Hermitian and $B = B^{\mathsf{t}}$ is symmetric.

The Hamiltonian is brought to diagonal form by a canonical transformation:

$$\begin{pmatrix} \psi \\ \psi^\dagger \end{pmatrix} = \overbrace{\begin{pmatrix} U & V^* \\ V & U^* \end{pmatrix}}^{\mathcal{S}} \begin{pmatrix} \phi \\ \phi^\dagger \end{pmatrix} \quad , \tag{17.5}$$

which is to say $\Psi = \mathcal{S}\,\Phi$, or in component form

$$\begin{aligned} \psi_i &= U_{ia}\,\phi_a + V^*_{ia}\,\phi^\dagger_a \\ \psi^\dagger_i &= V_{ia}\,\phi_a + U^*_{ia}\,\phi^\dagger_a \quad , \end{aligned} \tag{17.6}$$

where a, like i, runs from 1 to N. In order that the transformation be canonical, we must preserve the commutation relations, meaning $\left[\phi_a\,,\,\phi^\dagger_b\right] = \delta_{ab}$, *i.e.*

$$\left[\Phi_r\,,\,\Phi^\dagger_s\right] = \Sigma_{rs} \quad . \tag{17.7}$$

This then requires

$$\mathcal{S}\,\Sigma\,\mathcal{S}^\dagger = \mathcal{S}^\dagger\,\Sigma\,\mathcal{S} = \Sigma \quad , \tag{17.8}$$

which entails

$$U^\dagger U - V^\dagger V = \mathbb{I} \qquad\qquad U^{\rm t} V - V^{\rm t} U = 0 \tag{17.9}$$

$$U U^\dagger - V^* V^{\rm t} = \mathbb{I} \qquad\qquad U^* V^{\rm t} - V U^\dagger = 0 \quad . \tag{17.10}$$

Note that $\Sigma^2 = \mathcal{I}$, where $\mathcal{I} = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix}$, hence

$$\mathcal{S}^{-1} = \Sigma\,\mathcal{S}^\dagger\,\Sigma = \begin{pmatrix} U^\dagger & -V^\dagger \\ -V^{\rm t} & U^{\rm t} \end{pmatrix} \quad . \tag{17.11}$$

Thus, the inverse relation between the Ψ and Φ operators is $\Phi = \mathcal{S}^{-1}\Psi = \Sigma\,\mathcal{S}^\dagger\,\Sigma\,\Psi$, or

$$\begin{aligned} \phi_a &= U^*_{ia}\,\psi_i - V^*_{ia}\,\psi^\dagger_i \\ \phi^\dagger_a &= -V_{ia}\,\psi_i + U_{ia}\,\psi^\dagger_i \quad , \end{aligned} \tag{17.12}$$

17.1.1 Bogoliubov equations

We are now in the position to demand

$$\mathcal{S}^\dagger \mathcal{H} \mathcal{S} = \mathcal{E} = \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \quad , \tag{17.13}$$

where E is a diagonal $N \times N$ matrix. Thus,

$$\mathcal{H}\,\mathcal{S} = \mathcal{S}^{\dagger\,-1}\mathcal{E} = \Sigma\,\mathcal{S}\,\Sigma\,\mathcal{E} \quad , \tag{17.14}$$

which is to say

$$\begin{pmatrix} A & B \\ B^* & A \end{pmatrix} \begin{pmatrix} U & V^* \\ V & U^* \end{pmatrix} = \begin{pmatrix} U & -V^* \\ -V & U^* \end{pmatrix} \begin{pmatrix} E & 0 \\ 0 & E \end{pmatrix} \quad . \tag{17.15}$$

If the bosonic system is stable, each of the eigenvalues E_a is nonnegative. In component form, this yields the Bogoliubov equations,

$$\begin{aligned} A_{ij}\,U_{ja} + B_{ij}\,V_{ja} &= +U_{ia}\,E_a \\ B^*_{ij}\,U_{ja} + A^*_{ij}\,V_{ja} &= -V_{ia}\,E_a \quad , \end{aligned} \tag{17.16}$$

with no implied sum on a on either RHS. The Hamiltonian is then

$$\hat{H} = \sum_a E_a \left(\phi^\dagger_a \phi_a + \tfrac{1}{2}\right) \quad . \tag{17.17}$$

At temperature T, we have

$$\left\langle \phi^\dagger_a\,\phi_b \right\rangle = n(E_a)\,\delta_{ab} \quad , \tag{17.18}$$

where

$$n(E) = \frac{1}{\exp(E/k_{\rm B}T) - 1} \tag{17.19}$$

is the Bose distribution. The anomalous correlators all vanish, *e.g.* $\langle \phi_a \phi_b \rangle = 0$. The finite temperature two-point correlation functions are then

$$\langle \psi^\dagger_i \psi_j \rangle = \sum_a \Big\{ n_a\,U^*_{ia}\,U_{ja} + (1+n_a)\,V_{ia}\,V^*_{ja} \Big\} \tag{17.20}$$

$$\langle \psi_i \psi_j \rangle = \sum_a \Big\{ n_a\,V^*_{ia}\,U_{ja} + (1+n_a)\,U_{ia}\,V^*_{ja} \Big\} \quad , \tag{17.21}$$

where $n_a \equiv n(E_a)$.

17.1.2 Ground state

We have found

$$\Phi = \mathcal{S}^{-1}\Psi = \Sigma\,\mathcal{S}^\dagger\,\Sigma\,\Psi \quad , \tag{17.22}$$

hence

$$\begin{aligned} \phi_a &= U^\dagger_{ai}\,\psi_i - V^\dagger_{ai}\,\psi^\dagger_i \\ &= \psi_i\,U^*_{ia} - \psi^\dagger_i\,V^*_{ia} \quad . \end{aligned} \tag{17.23}$$

We assume the following Bogoliubov form for the ground state of $\hat{H}$:

$$|\,{\rm G}\,\rangle = C\,\exp\left(\tfrac{1}{2} Q_{ij}\,\psi^\dagger_i \psi^\dagger_j\right) |\,0\,\rangle \quad , \tag{17.24}$$

where C is a normalization constant, Q is a symmetric matrix, and $|\,0\,\rangle$ is the vacuum for the ψ bosons: $\psi_i|\,0\,\rangle = 0$. We now demand that $|\,{\rm G}\,\rangle$ be the vacuum for the ϕ bosons: $\phi_a|\,{\rm G}\,\rangle \equiv 0$. This means

$$\phi_a\,e^{\hat{Q}}\,|\,0\,\rangle = e^{\hat{Q}} \left(e^{-\hat{Q}}\,\phi_a\,e^{\hat{Q}}\right) |\,0\,\rangle \quad , \tag{17.25}$$

where

$$\hat{Q} \equiv \tfrac{1}{2} Q_{ij}\, \psi_i^\dagger \psi_j^\dagger \quad . \tag{17.26}$$

We now define

$$\psi_i(x) \equiv e^{-x\hat{Q}}\, \psi_i\, e^{x\hat{Q}} \tag{17.27}$$

and we find

$$\frac{d\psi_i(x)}{dx} = e^{-x\hat{Q}} \big[\psi_i\,,\, \hat{Q}\big] e^{x\hat{Q}} = Q_{ij}\, \psi_j^\dagger \quad , \tag{17.28}$$

and integrating[1] we obtain

$$\psi_i(x) \equiv e^{-x\hat{Q}}\, \psi_i\, e^{x\hat{Q}} = \psi_i(x) + x\, Q_{ij}\, \psi_j^\dagger \quad . \tag{17.29}$$

We may now write

$$e^{-\hat{Q}}\, \phi_a\, e^{\hat{Q}} = U_{ai}^\dagger\, \psi_i + \big(U_{ai}^\dagger\, Q_{ij} - V_{aj}^\dagger\big)\, \psi_j^\dagger \quad , \tag{17.30}$$

and we demand that the coefficient of $\psi_j^\dagger$ vanish for all a, which yields

$$Q = \big(U^\dagger\big)^{-1} V^\dagger \quad , \tag{17.31}$$

or, equivalently, $Q^\dagger = VU^{-1}$. Note that $Q^{\rm t} = V^*(U^*)^{-1} = Q$ since $U^\dagger V^* = V^\dagger U^*$.

17.1.3 A final note on the boson problem

Note that $\mathcal{S}^\dagger \mathcal{H} \mathcal{S}$ has the same eigenvalues as $\mathcal{H}$ only if $\mathcal{S}^\dagger = \mathcal{S}^{-1}$, *i.e.* only if $\mathcal{S}$ is Hermitian. We have $\mathcal{S}^\dagger = \Sigma \mathcal{S}^{-1} \Sigma$ and therefore

$$\mathcal{S}^\dagger \mathcal{H}\, \mathcal{S} = \Sigma\, \mathcal{S}^{-1}\, \Sigma\, \mathcal{H}\, \mathcal{S} \quad . \tag{17.32}$$

Now

$$\Sigma\, \mathcal{H} = \begin{pmatrix} A & B \\ -B^* & -A^* \end{pmatrix} \quad . \tag{17.33}$$

Consider the characteristic polynomial $P(E) = \det(E - \Sigma\mathcal{H})$. Since $\det(M) = \det(M^{\rm t})$ for any matrix M, we consider

$$(\Sigma\, \mathcal{H})^{\rm t} = \begin{pmatrix} A^{\rm t} & -B^\dagger \\ B^{\rm t} & -A^\dagger \end{pmatrix} = \begin{pmatrix} A^* & -B^* \\ B & -A \end{pmatrix} = -\mathcal{J}^{-1} (\Sigma\, \mathcal{H})\, \mathcal{J} \quad , \tag{17.34}$$

where

$$\mathcal{J} = \begin{pmatrix} 0 & \mathbb{I} \\ -\mathbb{I} & 0 \end{pmatrix} \tag{17.35}$$

[1] Note that $e^{-x\hat{Q}}\, \psi_i^\dagger\, e^{x\hat{Q}} = \psi_i^\dagger$ since $\big[\psi_i^\dagger, \hat{Q}\big] = 0$.

and $\mathcal{J}^{-1} = -\mathcal{J}$, *i.e.* $\mathcal{J}^2 = -\mathcal{I}$. But then we have

$$P(E) = \det(E - \Sigma\,\mathcal{H}) = \det(E + \mathcal{J}^{-1}\,\Sigma\,\mathcal{H}\,\mathcal{J}) = \det(E + \Sigma\,\mathcal{H}) = P(-E) \quad . \tag{17.36}$$

We conclude that the eigenvalues of $\Sigma\,\mathcal{H}$ come in $(+E, -E)$ pairs. To obtain the eigenenergies for the bosonic Hamiltonian $\hat{H}$, however, as per eqn. 17.32, we must multiply $\mathcal{S}^{-1}\Sigma\,\mathcal{H}\,\mathcal{S}$ on the left by Σ, which reverses the sign of the negative eigenvalues, resulting in a nonnegative definite spectrum of bosonic eigenoperators (for stable bosonic systems).

17.2 Fermionic Models

The general noninteracting fermionic Hamiltonian is written

$$\hat{H} = \tfrac{1}{2}\Psi_r^\dagger\,\mathcal{H}_{rs}\Psi_s \quad , \tag{17.37}$$

where once again Ψ is a rank-$2N$ column vector whose Hermitian conjugate is the row vector

$$\Psi^\dagger = \big(\psi_1^\dagger\,,\,\cdots\,,\,\psi_N^\dagger\,,\,\psi_1\,,\,\cdots\,,\,\psi_N\big) \quad . \tag{17.38}$$

In contrast to the bosonic case, we now have $\big\{\,\psi_i\,,\,\psi_j^\dagger\,\big\} = \delta_{ij}$ with the anticommutator, hence

$$\big\{\Psi_r\,,\,\Psi_s^\dagger\big\} = \delta_{rs} \quad . \tag{17.39}$$

The matrix $\mathcal{H}$ is of the form

$$\mathcal{H} = \begin{pmatrix} A & B \\ -B^* & -A^* \end{pmatrix} \quad , \tag{17.40}$$

where $A = A^\dagger$ is Hermitian and $B = -B^{\mathsf{t}}$ is antisymmetric. Since this is of the same form as eqn. 17.33, we conclude that the eigenvalues of $\mathcal{H}$ come in $(+E, -E)$ pairs[2].

As with the bosonic case, the Hamiltonian is brought to diagonal form by a canonical transformation:

$$\begin{pmatrix} \psi \\ \psi^\dagger \end{pmatrix} = \overbrace{\begin{pmatrix} U & V^* \\ V & U^* \end{pmatrix}}^{\mathcal{S}} \begin{pmatrix} \phi \\ \phi^\dagger \end{pmatrix} \quad , \tag{17.41}$$

which is to say $\Psi = \mathcal{S}\,\Phi$, or in component form

$$\begin{aligned} \psi_i &= U_{ia}\,\phi_a + V_{ia}^*\,\phi_a^\dagger \\ \psi_i^\dagger &= V_{ia}\,\phi_a + U_{ia}^*\,\phi_a^\dagger \quad . \end{aligned} \tag{17.42}$$

[2]This is true even though B in eqn. 17.33 is symmetric rather than antisymmetric. In proving the evenness of the characteristic polynomial $P(E) = P(-E)$, we did not appeal to the symmetry or antisymmetry of B.

In order that the transformation be canonical, we must preserve the anticommutation relations, *i.e.* $\{\phi_a\,,\,\phi_b^\dagger\} = \delta_{ab}$, meaning

$$\{\Phi_r\,,\,\Phi_s^\dagger\} = \delta_{rs} \quad , \tag{17.43}$$

which requires that $\mathcal{S}$ is unitary:

$$\mathcal{S}^\dagger\mathcal{S} = \mathcal{S}\mathcal{S}^\dagger = \mathcal{I} \quad , \tag{17.44}$$

where $\mathcal{I}$ is again the identity matrix of rank $2N$. Thus,

$$U^\dagger U + V^\dagger V = \mathbb{I} \qquad\qquad U^{\mathsf{t}} V + V^{\mathsf{t}} U = 0 \tag{17.45}$$

$$U U^\dagger + V^* V^{\mathsf{t}} = \mathbb{I} \qquad\qquad U^* V^{\mathsf{t}} + V U^\dagger = 0 \quad . \tag{17.46}$$

The inverse relation between the operators follows from $\Phi = \mathcal{S}^{-1}\Psi = \mathcal{S}^\dagger\Psi$:

$$\begin{aligned} \phi_a &= U_{ia}^*\,\psi_i + V_{ia}^*\,\psi_i^\dagger \\ \phi_a^\dagger &= V_{ia}\,\psi_i + U_{ia}\,\psi_i^\dagger \quad , \end{aligned} \tag{17.47}$$

The transformed Hamiltonian matrix is

$$\mathcal{S}^\dagger\mathcal{H}\,\mathcal{S} = \mathcal{E} \equiv \begin{pmatrix} E & 0 \\ 0 & -E \end{pmatrix} \quad . \tag{17.48}$$

Without loss of generality, we may take E to be a diagonal matrix with nonnegative entries. In component notation, the eigenvalue equations are

$$\begin{aligned} A_{ij}\,U_{ja} + B_{ij}\,V_{ja} &= U_{ia}\,E_a \\ -B_{ij}^*\,U_{ja} - A_{ij}^*\,V_{ja} &= V_{ia}\,E_a \quad . \end{aligned} \tag{17.49}$$

The Hamiltonian then takes the form

$$\hat{H} = \sum_a E_a\left(\phi_a^\dagger\phi_a - \tfrac{1}{2}\right) \quad . \tag{17.50}$$

At temperature T, we have

$$\left\langle \phi_a^\dagger\phi_b \right\rangle = f(E_a)\,\delta_{ab} \quad , \tag{17.51}$$

where

$$f(E) = \frac{1}{\exp(E/k_{\rm B}T) + 1} \tag{17.52}$$

is the Fermi distribution. As for bosons, the anomalous correlators all vanish: $\langle\phi_a\phi_b\rangle = 0$. The finite temperature two-point correlation functions are then

$$\begin{aligned} \langle\psi_i^\dagger\psi_j\rangle &= \sum_a \Big\{ f_a\,U_{ia}^*\,U_{ja} + (1-f_a)\,V_{ia}\,V_{ja}^* \Big\} \\ \langle\psi_i\psi_j\rangle &= \sum_a \Big\{ f_a\,V_{ia}^*\,U_{ja} + (1-f_a)\,U_{ia}\,V_{ja}^* \Big\} \quad , \end{aligned} \tag{17.53}$$

where $f_a = f(E_a)$.

17.2.1 Ground state

We write

$$|\,\mathrm{G}\,\rangle = C\exp\left(\tfrac{1}{2}Q_{ij}\,\psi_i^\dagger\psi_j^\dagger\right)|\,0\,\rangle \quad , \tag{17.54}$$

with $Q = -Q^{\rm t}$, and we demand, as in the bosonic case, that $\phi_a\,|\,\mathrm{G}\,\rangle \equiv 0$. Again we define $\hat{Q} = \frac{1}{2}Q_{ij}\,\psi_i^\dagger\psi_j^\dagger$, and

$$\psi_i(x) = e^{-x\hat{Q}}\,\psi_i\,e^{x\hat{Q}} \quad . \tag{17.55}$$

We then have

$$\frac{d\psi_i(x)}{dx} = e^{-x\hat{Q}}\left[\psi_i,\hat{Q}\right]e^{x\hat{Q}} = Q_{ij}\,\psi_j^\dagger \quad \Rightarrow \quad \psi_i(x) = \psi_i + x\,Q_{ij}\,\psi_j^\dagger \quad . \tag{17.56}$$

Thus,

$$e^{-\hat{Q}}\,\phi_a\,e^{\hat{Q}} = U_{ai}^\dagger\,\psi_i + \left(V_{aj}^\dagger + U_{ai}^\dagger\,Q_{ij}\right)\psi_j^\dagger \quad , \tag{17.57}$$

from which we obtain

$$Q = -\big(U^\dagger\big)^{-1}V^\dagger \quad . \tag{17.58}$$

Since $U^\dagger V^* + V^\dagger U^* = 0$, we recover $Q = -Q^{\rm t}$.

17.3 Majorana Fermion Models

Majorana fermions satisfy the anticommutation relations $\left\{\theta_i\,,\,\theta_j\right\} = 2\delta_{ij}$. Thus, $(\theta_i)^2 = 1$ for every i. We also have $\theta_i^\dagger = \theta_i$ and for this reason they are sometimes called 'real' fermions. If c is the annihilator for a Dirac particle, with $\left\{c\,,\,c^\dagger\right\} = 1$, we may define Majorana fermions η and $\widetilde{\eta}$ as follows:

$$\eta = c + c^\dagger \qquad\qquad c = \tfrac{1}{2}(\eta - i\eta') \tag{17.59}$$

$$\widetilde{\eta} = i(c - c^\dagger) \qquad\qquad c^\dagger = \tfrac{1}{2}(\eta + i\widetilde{\eta}) \quad . \tag{17.60}$$

The most general noninteracting Majorana Hamiltonian is of the form

$$\hat{H} = \tfrac{i}{4}M_{ij}\,\theta_i\,\theta_j \quad , \tag{17.61}$$

where $M = -M^{\rm t} = M^*$ is a real antisymmetric matrix of even dimension $2N$. This is brought to canonical form by a real orthogonal transformation,

$$\theta_i = \mathcal{R}_{ia}\,\xi_a \quad , \tag{17.62}$$

where $\mathcal{R}^{\rm t}\mathcal{R} = \mathcal{I}$, and where $\big\{\xi_a\,,\,\xi_b\big\} = 2\delta_{ab}$. We have

$$\mathcal{R}^{\rm t}\mathcal{M}\mathcal{R} = E\otimes i\sigma^y = \begin{pmatrix} 0 & -E_1 & 0 & 0 & \cdots \\ E_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & -E_2 & \cdots \\ 0 & 0 & E_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad . \tag{17.63}$$

Thus,

$$\hat{H} = -\tfrac{i}{2}\sum_{a=1}^{N} E_a\,\xi_{2a-1}\,\xi_{2a} = \sum_a E_a\big(c_a^\dagger c_a - \tfrac{1}{2}\big) \quad , \tag{17.64}$$

where

$$c_a \equiv \tfrac{1}{2}\big(\xi_{2a-1} - i\,\xi_{2a}\big) \quad , \quad c_a^\dagger \equiv \tfrac{1}{2}\big(\xi_{2a-1} + i\,\xi_{2a}\big) \quad . \tag{17.65}$$

17.3.1 Majorana chain

Consider the Hamiltonian

$$\hat{H} = -i\sum_{n=1}^{N}\sigma_n\,\alpha_n\,\alpha_{n+1} \tag{17.66}$$

where $\sigma_n = \pm 1$ is a $\mathbb{Z}_2$ gauge field and $\{\alpha_m,\alpha_n\} = 2\,\delta_{mn}$ is the Majorana fermion anticommutator. Periodic boundary conditions are assumed, *i.e.* $\alpha_{N+1} = \alpha_1$. We now make a gauge transformation to a new set of Majorana fermions,

$$\theta_1 \equiv \alpha_1 \quad , \quad \theta_2 \equiv \sigma_1\alpha_2 \quad , \quad \theta_3 \equiv \sigma_1\sigma_2\,\alpha_3 \quad , \quad \ldots \quad , \quad \theta_N \equiv \sigma_1\sigma_2\cdots\sigma_{N-1}\,\alpha_N \quad . \tag{17.67}$$

The Hamiltonian may now be written as

$$\hat{H} = -i\sum_{n=1}^{N}\theta_n\,\theta_{n+1} \quad , \tag{17.68}$$

where $\theta_{N+1} = \sigma\,\theta_1$, with $\sigma = \prod_{j=1}^{N}\sigma_j$. So the boundary conditions on the θ Majoranas are either periodic ($\sigma = +1$) or antiperiodic ($\sigma = -1$). We now switch to crystal momentum space, defining

$$\hat{\theta}_k = \frac{1}{\sqrt{N}}\sum_{n=1}^{N} e^{-ikn}\,\theta_n \qquad , \qquad \theta_n = \frac{1}{\sqrt{N}}\sum_k e^{ikn}\,\hat{\theta}_k \quad . \tag{17.69}$$

The k-values are quantized according to $e^{ikN} = \sigma$. The anticommutators are

$$\big\{\theta_m\,,\,\theta_n\big\} = 2\,\delta_{m-n\,,\,0\,\mathrm{mod}\,N} \qquad , \qquad \big\{\hat{\theta}_k\,,\,\hat{\theta}_p\big\} = 2\,\delta_{k+p\,,\,0\,\mathrm{mod}\,2\pi} \quad . \tag{17.70}$$

There are four cases to consider:

Case I : $\sigma = +1$, N even. We have $e^{ikN} = +1$, and the N allowed k values are

$$k \in \pm\frac{2\pi}{N} \times \Big\{1\,,\,\ldots\,,\,\tfrac{1}{2}N-1\Big\} \quad , \quad k=0 \quad , \quad k=\pi \quad . \tag{17.71}$$

Note that the allowed crystal momenta all occur in $\{+k,-k\}$ pairs, with the exception of $k=0$ and $k=\pi$, which are unpaired.

Case II : $\sigma = +1$, N odd. We have $e^{ikN} = +1$, and the N allowed k values are

$$k \in \pm\frac{2\pi}{N} \times \Big\{1\,,\,\ldots\,,\,\tfrac{1}{2}(N-1)\Big\} \quad , \quad k=0 \quad . \tag{17.72}$$

Only $k=0$ is unpaired.

Case III : $\sigma = 1$, N even. We have $e^{ikN} = -1$, and the N allowed k values are

$$k \in \pm\frac{2\pi}{N} \times \Big\{\tfrac{1}{2}\,,\,\ldots\,,\,\tfrac{1}{2}(N-1)\Big\} \quad . \tag{17.73}$$

All the crystal momenta are paired.

Case IV : $\sigma = 1$, N odd. We have $e^{ikN} = -1$, and the N allowed k values are

$$k \in \pm\frac{2\pi}{N} \times \Big\{\tfrac{1}{2}\,,\,\ldots\,,\,\tfrac{1}{2}N-1\Big\} \quad , \quad k=\pi \quad . \tag{17.74}$$

Only $k=\pi$ is unpaired.

We may now write

$$\begin{aligned}\hat{H} &= -i\sum_k e^{-ik}\,\hat\theta_k\,\hat\theta_{-k} \\ &= -i\sum_{k\in(0,\pi)}\Big(e^{ik}\,\hat\theta_{-k}\,\hat\theta_k + e^{-ik}\,\hat\theta_k\,\hat\theta_{-k}\Big) - i\sum_{k\in\mathrm{U}} e^{-ik}\,\hat\theta_k^2 \\ &= \sum_{k\in(0,\pi)} 2\sin k\;\hat\theta_{-k}\,\hat\theta_k - 2i\sum_{k\in(0,\pi)} e^{-ik} - i\sum_{k\in\mathrm{U}} e^{-ik} \quad .\end{aligned} \tag{17.75}$$

where U denotes the set of unpaired (or self-paired) crystal momenta, *i.e.* the set of k for which $e^{ik} = e^{-ik}$. Note that $\big\{\hat\theta_{-k}\,,\,\hat\theta_{k'}\big\} = 2\,\delta_{k,k'}$ and $\hat\theta_{-k} = \hat\theta_k^\dagger$, so we may define $\hat\theta_{-k} \equiv \sqrt{2}\,c_k^\dagger$ and $\hat\theta_k \equiv \sqrt{2}\,c_k$, where c_k is a complex fermion. Thus, we have

$$\hat{H} = \sum_{k\in(0,\pi)} 4\sin k\;c_k^\dagger\,c_k + E_0 \quad , \tag{17.76}$$

where

$$E_0 = -2i\sum_{k\in(0,\pi)} e^{-ik} - i\sum_{k\in\mathrm{U}} e^{-ik} \quad . \tag{17.77}$$

We now proceed to evaluate E_0 for our four cases.

Case I : Since $\mathrm{U} = \{0, \pi\}$, we have $\sum_{k\in \mathrm{U}} e^{-ik} = 0$. For $k \in (0, \pi)$ we may write $k = 2\pi\ell/N$ with $\ell \in \big\{1\,,\,\ldots\,,\,\frac{1}{2}N - 1\big\}$. We then have

$$E_0^{(\mathrm{I})} = -2i \sum_{\ell=1}^{\frac{N}{2}-1} e^{-2\pi i\ell/N} = -2\,\mathrm{ctn}\Big(\frac{\pi}{N}\Big) \quad . \tag{17.78}$$

Note that we have used the identity

$$\sum_{\ell=1}^{J-1} x^\ell = \frac{x - x^J}{1 - x} \quad . \tag{17.79}$$

Case II : We have $\mathrm{U} = \{0\}$. For the main set $k \in (0, \pi)$ we may write $k = 2\pi\ell/N$ with $\ell \in \big\{1\,,\,\ldots\,,\,\frac{1}{2}(N-1)\big\}$. We then have

$$E_0^{(\mathrm{II})} = -2i \sum_{\ell=1}^{\frac{N+1}{2}-1} e^{-2\pi i\ell/N} - i = -2i\left(\frac{e^{-2\pi i/N} + e^{-i\pi/N}}{1 - e^{-2\pi i/N}}\right) - i = -\,\mathrm{ctn}\left(\frac{\pi}{2N}\right) \quad . \tag{17.80}$$

Case III : We have $\mathrm{U} = \{\emptyset\}$. For $k \in (0, \pi)$ we may write $k = 2\pi\ell/N + \pi/N$ with $\ell \in \big\{0\,,\,\ldots\,,\,\frac{1}{2}N - 1\big\}$. Then

$$E_0^{(\mathrm{III})} = -2i\, e^{-i\pi/N} \sum_{\ell=0}^{\frac{N}{2}-1} e^{-2\pi\ell/N} = -2\,\mathrm{csc}\left(\frac{\pi}{N}\right) \quad . \tag{17.81}$$

Case IV : We have $\mathrm{U} = \{\pi\}$. For $k \in (0, \pi)$ we may write $k = 2\pi\ell/N - \pi/N$ with $\ell \in \big\{1\,,\,\ldots\,,\,\frac{1}{2}(N-1)\big\}$. Thus,

$$E_0^{(\mathrm{IV})} = -2i\, e^{i\pi/N} \sum_{\ell=1}^{\frac{N+1}{2}-1} e^{-2\pi i\ell/N} + i = -2i\left(\frac{e^{-i\pi/N} + 1}{1 - e^{-2\pi i/N}}\right) + i = -\,\mathrm{ctn}\left(\frac{\pi}{2N}\right) \quad . \tag{17.82}$$

Note that in the $N \to \infty$ limit, in all four cases we have $E_0 = 2N/\pi + \mathcal{O}(1)$.

17.4 Jordan-Wigner Transformation

The Jordan-Wigner transformation is an equivalence, in one-dimensional lattice systems, between the $S = \frac{1}{2}$ SU(2) algebra and the algebra of spinless fermions. Explicitly, we have

$$\begin{aligned} S_n^+ &= \exp\left(i\pi \sum_{j=1}^{n-1} c_j^\dagger c_j\right) c_n^\dagger \\ S_n^- &= \exp\left(i\pi \sum_{j=1}^{n-1} c_j^\dagger c_j\right) c_n \\ S_n^z &= c_n^\dagger c_n - \tfrac{1}{2} \quad . \end{aligned} \tag{17.83}$$

The inverse is then

$$\begin{aligned} c_n^\dagger &= \exp\left(i\pi \sum_{j=1}^{n-1} \big(S_j^z + \tfrac{1}{2}\big)\right) S_n^+ \\ c_n &= \exp\left(i\pi \sum_{j=1}^{n-1} \big(S_j^z + \tfrac{1}{2}\big)\right) S_n^- \quad . \end{aligned} \tag{17.84}$$

Note that $e^{i\pi c^\dagger c}$ has eigenvalues ± 1, and that

$$c\, e^{i\pi c^\dagger c} = -c \quad , \quad c^\dagger e^{i\pi c^\dagger c} = c^\dagger \quad . \tag{17.85}$$

Taking the Hermitian conjugate,

$$e^{i\pi c^\dagger c}\, c^\dagger = -c^\dagger \quad , \quad e^{i\pi c^\dagger c}\, c = c \quad . \tag{17.86}$$

The expression

$$\exp\left(i\pi \sum_{j=1}^{n-1} \big(S_j^z + \tfrac{1}{2}\big)\right) = \prod_{j=1}^{n-1} \exp\Big(i\pi\big(S_j^z + \tfrac{1}{2}\big)\Big) \tag{17.87}$$

is known as a *Jordan-Wigner string*.

The nearest-neighbor bilinear transverse spin interaction terms are

$$\begin{aligned} S_n^+ S_{n+1}^- &= c_n^\dagger\, e^{i\pi c_n^\dagger c_n}\, c_{n+1} = c_n^\dagger\, c_{n+1} \\ S_n^- S_{n+1}^+ &= c_n\, e^{i\pi c_n^\dagger c_n}\, c_{n+1}^\dagger = c_{n+1}^\dagger\, c_n \\ S_n^+ S_{n+1}^+ &= c_n^\dagger\, e^{i\pi c_n^\dagger c_n}\, c_{n+1}^\dagger = c_n^\dagger\, c_{n+1}^\dagger \\ S_n^- S_{n+1}^+ &= c_n\, e^{i\pi c_n^\dagger c_n}\, c_{n+1} = c_{n+1}\, c_n \quad . \end{aligned} \tag{17.88}$$

On an N-site ring, however, on the 'last' link, which connects site N back to site 1, yields

$$\begin{aligned} S_N^+ \, S_1^- &= -e^{i\pi\hat{M}} \, c_N^\dagger c_1 \\ S_N^- \, S_1^+ &= -e^{i\pi\hat{M}} \, c_1^\dagger c_N \\ S_N^+ \, S_1^+ &= -e^{i\pi\hat{M}} \, c_N^\dagger c_1^\dagger \\ S_N^- \, S_1^+ &= -e^{i\pi\hat{M}} \, c_1 c_N \quad . \end{aligned} \tag{17.89}$$

where

$$\hat{M} = \sum_{j=1}^{N} c_j^\dagger c_j \quad . \tag{17.90}$$

Note that $e^{i\pi\hat{M}} = (-1)^{\hat{M}}$ must commute with every possible term we could write, since fermion number parity must be conserved.

17.4.1 Anisotropic XY model

Consider the anisotropic XY model in a perpendicular field on an N-site chain[3], with

$$\begin{aligned} \hat{H}_{\text{chain}} &= \sum_{n=1}^{N-1} \Big\{ J_x \, S_n^x \, S_{n+1}^x + J_y \, S_n^y \, S_{n+1}^y \Big\} + h \sum_{n=1}^{N} S_n^z \\ &= \tfrac{1}{2} \sum_{n=1}^{N-1} \Big\{ J_+ \big(c_n^\dagger \, c_{n+1} + c_{n+1}^\dagger \, c_n \big) + J_- \big(c_n^\dagger \, c_{n+1}^\dagger + c_{n+1} \, c_n \big) \Big\} + h \sum_{n=1}^{N} \big(c_n^\dagger \, c_n - \tfrac{1}{2} \big) \quad , \end{aligned} \tag{17.91}$$

where $J_\pm = \frac{1}{2}(J_x \pm J_y)$. On an N-site ring, we add the term

$$\begin{aligned} \Delta H &= J_x \, S_N^x \, S_1^x + J_y \, S_N^y \, S_1^y \\ &= -\tfrac{1}{2} \, e^{i\pi\hat{M}} \Big\{ J_+ \big(c_N^\dagger \, c_1 + c_1^\dagger \, c_N \big) + J_- \big(c_N^\dagger \, c_1^\dagger + c_1 \, c_N \big) \Big\} \quad . \end{aligned} \tag{17.92}$$

Since $e^{i\pi\hat{M}}$ commutes with $\hat{H}_{\text{chain}}$ and with all fermion bilinears (hence with ΔH as well), we can specify the eigenvalues as $\eta \equiv e^{i\pi\hat{M}} = \pm 1$, which are the even and odd fermion number sectors, respectively. We then define

$$c_1 \equiv \begin{cases} -c_{N+1} & \text{if } \eta = +1 \\ +c_{N+1} & \text{if } \eta = -1 \end{cases} \quad . \tag{17.93}$$

If we write

$$c_n = \frac{1}{\sqrt{N}} \sum_k e^{ikn} \, c_k \quad , \tag{17.94}$$

[3]See E. Lieb, T. Schultz, and D. Mattis, *Ann. Phys.* **16**, 407 (1961).

where the index n refers to real space and k to momentum space, we have the wave vector quantization rule $e^{ikN} = -\eta$, *i.e.* for even and odd sectors

$$k_j = \begin{cases} 2\pi(j+\frac{1}{2})/N & \text{if } \eta = +1 \\ 2\pi j/N & \text{if } \eta = -1 \end{cases} \quad . \tag{17.95}$$

Thus, the Hamiltonian becomes

$$\begin{aligned} \hat{H}_{\rm ring} &= \sum_k \Big\{ (J_+ \cos k + h)\, c_k^\dagger\, c_k + \tfrac{1}{2} J_-\, e^{ik}\, c_k^\dagger\, c_{-k}^\dagger + \tfrac{1}{2} J_-\, e^{-ik}\, c_{-k}\, c_k \Big\} + \tfrac{1}{2} Nh \\ &= \sum_{k>0} \begin{pmatrix} c_k^\dagger & c_{-k} \end{pmatrix} \overbrace{\begin{pmatrix} \omega_k & \Delta_k \\ \Delta_k^* & -\omega_k \end{pmatrix}}^{H_k} \begin{pmatrix} c_k \\ c_{-k}^\dagger \end{pmatrix} \quad , \end{aligned} \tag{17.96}$$

where

$$\omega_k = J_+ \cos k + h \qquad . \qquad \Delta_k = i\, J_- \sin k \quad . \tag{17.97}$$

Diagonalizing via a unitary transformation, we obtain

$$\hat{H}_{\rm ring} = \sum_k E_k \left(\gamma_k^\dagger \gamma_k - \tfrac{1}{2} \right) \quad , \tag{17.98}$$

where the dispersion relation is

$$E_k = \sqrt{\omega_k^2 + |\Delta_k|^2} = \sqrt{(J_+ \cos k + h)^2 + J_-^2 \sin^2 k} \quad . \tag{17.99}$$

Note that $S_k^\dagger\, H_k\, S_k = \mathsf{diag}(E_k, -E_k)$, where

$$S_k = \begin{pmatrix} u_k & -v_k^* \\ v_k & u_k \end{pmatrix} \tag{17.100}$$

where

$$u_k = \frac{E_k + \omega_k}{\sqrt{2E_k(E_k+\omega_k)}} \qquad , \qquad v_k = \frac{\Delta_k^*}{\sqrt{2E_k(E_k+\omega_k)}} \quad . \tag{17.101}$$

Thus,

$$\begin{aligned} \gamma_k &= u_k\, c_k - v_k^*\, c_{-k}^\dagger \\ \gamma_k^\dagger &= -v_k\, c_{-k} + u_k\, c_k^\dagger \quad . \end{aligned} \tag{17.102}$$

Note that $u_{-k} = u_k = u_k^*$ while $v_{-k} = -v_k = v_k^*$, and that

$$\begin{aligned} c_k &= u_k\, \gamma_k + v_k^*\, \gamma_{-k}^\dagger \\ c_k^\dagger &= v_k\, \gamma_{-k} + u_k\, \gamma_k^\dagger \quad . \end{aligned} \tag{17.103}$$

When we compute correlation functions, we use the fact that

$$e^{i\pi c^\dagger c} = (c^\dagger + c)(c^\dagger - c) = -(c^\dagger - c)(c^\dagger + c) \quad , \tag{17.104}$$

and, defining $A_j \equiv c_j^\dagger + c_j$ and $B_j \equiv c_j^\dagger - c_j$, Then the correlation functions are

$$\begin{aligned}
\rho_x(\ell) &= \big\langle S_n^x \, S_{n+\ell}^x \big\rangle = \tfrac{1}{4} \big\langle B_n \, A_{n+1} \, B_{n+1} \cdots A_{n+\ell-1} \, B_{n+\ell-1} \, A_{n+\ell} \big\rangle \\
\rho_y(\ell) &= \big\langle S_n^y \, S_{n+\ell}^y \big\rangle = \tfrac{1}{4} (-1)^\ell \big\langle A_n \, B_{n+1} \, A_{n+1} \cdots B_{n+\ell-1} \, A_{n+\ell-1} \, B_{n+\ell} \big\rangle \\
\rho_z(\ell) &= \big\langle S_n^z \, S_{n+\ell}^z \big\rangle = \tfrac{1}{4} \big\langle A_n \, B_n \, A_{n+\ell} \, B_{n+\ell} \big\rangle \quad ,
\end{aligned} \tag{17.105}$$

where, without loss of generality, we presume $\ell > 0$. These expressions may be evaluated using Wick's theorem,

$$\big\langle \mathcal{O}_1 \, \mathcal{O}_2 \, \cdots \, \mathcal{O}_{2m} \big\rangle = \sum_{\sigma \in \mathcal{C}_{2r}} (-1)^\sigma \big\langle \mathcal{O}_{\sigma(1)} \, \mathcal{O}_{\sigma(2)} \big\rangle \cdots \big\langle \mathcal{O}_{\sigma(2r-1)} \, \mathcal{O}_{\sigma(2r)} \big\rangle \quad , \tag{17.106}$$

where σ is one of a special set of permutations $\mathcal{C}_{2r}$ of the set $\{1, \ldots, 2r\}$ called *contractions*, which are arrangements of the $2r$ indices into r pairs. Exchanging any two pairs, or exchanging the indices within a pair results in the same contraction, so the number of such contractions is $|\mathcal{C}_{2r}| = (2r)!/(2^r \cdot r!)$. Here $(-1)^\sigma$ is the sign of the permutation σ. As an example, for $r = 2$ there are $4!/(4 \cdot 2) = 3$ contractions. We then have

$$\rho_z(\ell) = \tfrac{1}{4} \big\langle A_n B_n \big\rangle \big\langle A_{n+\ell} B_{n+\ell} \big\rangle - \tfrac{1}{4} \big\langle A_n A_{n+\ell} \big\rangle \big\langle B_n B_{n+\ell} \big\rangle + \tfrac{1}{4} \big\langle A_n B_{n+\ell} \big\rangle \big\langle B_n A_{n+\ell} \big\rangle \quad . \tag{17.107}$$

Now we need the following:

$$\big\langle A_n A_{n'} \big\rangle = \delta_{nn'} \qquad , \qquad \big\langle B_n B_{n'} \big\rangle = -\delta_{nn'} \qquad , \qquad \big\langle A_n B_{n'} \big\rangle \equiv G(n' - n) \tag{17.108}$$

The first two of these relations follow by inversion symmetry, *i.e.*

$$\big\langle A_n A_{n'} \big\rangle = \big\langle A_{n'} A_n \big\rangle \quad \Rightarrow \quad \big\langle A_n A_{n'} \big\rangle = \tfrac{1}{2} \big\langle \{A_n, A_{n'}\} \big\rangle = \delta_{nn'} \quad , \tag{17.109}$$

with a corresponding argument showing $\big\langle B_n B_{n'} \big\rangle = -\delta_{nn'}$. We then have

$$\begin{aligned}
G(n' - n) &= \big\langle (c_n^\dagger + c_n) \, (c_{n'}^\dagger - c_{n'}) \big\rangle \\
&= \frac{1}{N} \sum_{k,k'} \Big(\big\langle c_k^\dagger \, c_{k'}^\dagger \big\rangle - \big\langle c_{-k} c_{k'} \big\rangle + \big\langle c_{-k} c_{-k}^\dagger \big\rangle - \big\langle c_k^\dagger c_k \big\rangle \Big) \, e^{ik(n'-n)} \\
&= \frac{1}{N} \sum_k \Big(u_k^2 - |v_k|^2 + 2 u_k v_k \Big) \, e^{-ikn} \, e^{ik'n'} = \frac{1}{N} \sum_k \left(\frac{\omega_k + \Delta_k}{E_k} \right) e^{ik(n'-n)}
\end{aligned} \tag{17.110}$$

for $n \neq n'$, and at $T = 0$. Note that $\big\langle B_{n'} A_n \big\rangle = -G(n - n')$ for $n \neq n'$ and that $G(0) = 1 - 2\nu$ where $\nu = \langle c_j^\dagger c_j \rangle$ is the fermion occupation per site, which is translationally invariant. Thus, we have

$$\rho_z(\ell) = \tfrac{1}{4} \, G^2(0) - \tfrac{1}{4} \, G(\ell) \, G(-\ell) \tag{17.111}$$

The transverse spin correlations may be expressed as determinants, *viz.*

$$\rho_x(\ell) = \det \begin{pmatrix} G(1) & G(2) & \cdots & G(\ell) \\ G(0) & G(1) & \cdots & G(\ell-1) \\ \vdots & \vdots & \ddots & \vdots \\ G(2-\ell) & G(3-\ell) & \cdots & G(1) \end{pmatrix} \tag{17.112}$$

and

$$\rho_y(\ell) = \det \begin{pmatrix} G(-1) & G(0) & \cdots & G(\ell-2) \\ G(-2) & G(-1) & \cdots & G(\ell-3) \\ \vdots & \vdots & \ddots & \vdots \\ G(-\ell) & G(1-\ell) & \cdots & G(-1) \end{pmatrix} \quad . \tag{17.113}$$

Matrices like these which are constant along the diagonals are called *Toeplitz matrices*. A matrix M is Toeplitz if $M_{i,j} = M_{i+1,j+1} = m(i-j)$.

17.4.2 Majorana representation of the JW transformation

With Eqn. 17.65, which describes how one can write a single Dirac fermion with operators c and $c^\dagger$ in terms of two Majorana fermions α and β, *i.e.* $\alpha = c + c^\dagger$ and $\beta = i(c - c^\dagger)$, we can write the JW transformation as follows:

$$\begin{aligned} X_n &= (i\,\alpha_1\,\beta_1)\,(i\,\alpha_2\,\beta_2)\cdots(i\,\alpha_{n-1}\,\beta_{n-1})\,\alpha_n \\ Y_n &= (i\,\alpha_1\,\beta_1)\,(i\,\alpha_2\,\beta_2)\cdots(i\,\alpha_{n-1}\,\beta_{n-1})\,\beta_n \\ Z_n &= -i\,\alpha_n\,\beta_n \quad . \end{aligned} \tag{17.114}$$

Here we write (X_n, Y_n, Z_n) for the Pauli matrices $(\sigma^x_n, \sigma^y_n, \sigma^z_n) = (2S^x_n, 2S^y_n, 2S^z_n)$. Note that $X_n\,Y_n = i\,Z_n$. Thus, we have written the N spin operators along the chain in terms of $2N$ Majorana fermions $\{\alpha_1, \beta_1, \ldots, \alpha_N, \beta_N\}$, and, through the relations $\alpha_n = c_n + c_n^\dagger$ and $\beta_n = i(c_n - c_n^\dagger)$, in terms of N Dirac fermions $\big\{(c_1, c_1^\dagger), \ldots, (c_N, c_N^\dagger)\big\}$. Note that

$$i\,\alpha_n\,\beta_n = -Z_n = \exp(i\pi c_n^\dagger c_n) = 1 - 2\,c_n^\dagger c_n \quad , \tag{17.115}$$

and we thereby recover Eqn. 17.84.